Boundless Statistics for Organizations

BOUNDLESS STATISTICS FOR ORGANIZATIONS

Data Analysis & Interpretation

BRAD GRIFFITH AND LISA FRIESEN

Online Consortium of Oklahoma
Oklahoma City

Boundless Statistics for Organizations by Brad Griffith and Lisa Friesen is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](), except where otherwise noted.

CONTENTS

Section IX. 5.XLSX – Excel Challenge - Formulas, Functions, Logical and Lookup Functions

ABOUT THIS BOOK

Boundless Statistics for Organizations is a 2021 adaptation of *Boundless Statistics* from Lumen Learning, customized for the Reach Higher Organizational Leadership program in Oklahoma. For more information, visit the Reach Higher website (click image below):

If you have suggestions for improvement or need to report an issue, please contact Brad Griffith(bgriffith@osrhe.edu).

TABLE OF CONTENTS

Appendix

SECTION I

1. INTRODUCTION TO STATISTICS AND STATISTICAL THINKING

1.1 OVERVIEW

1.1: Overview

1.1.1: Collecting and Measuring Data

There are four main levels of measurement: nominal, ordinal, interval, and ratio.

Learning Objective

Distinguish between the nominal, ordinal, interval and ratio methods of data measurement.

Key Takeaways

Key Points

- Ratio measurements provide the greatest flexibility in statistical methods that can be used for analyzing the data.
- Interval data allows for the degree of difference between items, but not the ratio between them.
- Ordinal measurements have imprecise differences between consecutive values, but have a meaningful order to those values.
- Variables conforming only to nominal or ordinal measurements cannot be reasonably

measured numerically, they are often grouped together as categorical variables.

- Ratio and interval measurements are grouped together as quantitative variables.
- Nominal measurements have no meaningful rank order among values.

Key Terms

sampling
: the process or technique of obtaining a representative sample

population
: a group of units (persons, objects, or other items) enumerated in a census or from which a sample is drawn

Example

An example of an observational study is one that explores the correlation between smoking and lung cancer. This type of study typically uses a survey to collect observations about the area of interest and then performs statistical analysis. In this case, the researchers would collect observations of both smokers and non-smokers, perhaps through a case-control study, and then look for the number of cases of lung cancer in each group.

There are four main levels of measurement used in statistics: nominal, ordinal, interval, and ratio. Each of these have different degrees of usefulness in statistical research. Data is collected about a population by random sampling .

Nominal measurements have no meaningful rank order among values. Nominal data differentiates between items or subjects based only on qualitative classifications they belong to. Examples include gender, nationality, ethnicity, language, genre, style, biological species, visual pattern, etc.

Defining a population

In applying statistics to a scientific, industrial, or societal problem, it is necessary to begin with a population or process to be studied. Populations can be diverse topics such as "all persons living in a country" or "all stamps produced in the year 1943".

Ordinal measurements have imprecise differences between consecutive values, but have a meaningful order to those values. Ordinal data allows for rank order (1st, 2nd, 3rd, etc) by which data can be sorted, but it still does not allow for relative degree of difference between them. Examples of ordinal data include dichotomous values such as "sick" versus "healthy" when measuring health, "guilty" versus "innocent" when making judgments in courts, "false" versus "true", when measuring truth value. Examples also include non-dichotomous data consisting of a spectrum of values, such as "completely agree", "mostly agree", "mostly disagree", or "completely disagree" when measuring opinion.

Interval measurements have meaningful distances between measurements defined, but the zero value is arbitrary (as in the case with longitude and temperature measurements in Celsius or Fahrenheit). Interval data allows for the degree of difference between items, but not the ratio between them. Ratios are not allowed with interval data since 20°C cannot be said to be "twice as hot" as 10°C, nor can multiplication/division be carried out between any two dates directly. However, ratios of differences can be expressed; for example, one difference can be twice another. Interval type variables are sometimes also called "scaled variables".

Ratio measurements have both a meaningful zero value and the distances between different measurements are defined; they provide the greatest flexibility in statistical methods that can be used for analyzing the data.

Because variables conforming only to nominal or ordinal measurements cannot be reasonably measured numerically, sometimes they are grouped together as categorical variables, whereas ratio and interval measurements are grouped together as quantitative variables, which can be either discrete or continuous, due to their numerical nature.

Measurement processes that generate statistical data are also subject to error. Many of these errors are classified as random (noise) or systematic (bias), but other important types of errors (e.g., blunder, such as when an analyst reports incorrect units) can also be important.

1.1.2: What Is Statistics?

Statistics is the study of the collection, organization, analysis, interpretation, and presentation of data.

Learning Objective

Define the field of Statistics in terms of its definition, application and history.

Key Takeaways

Key Points

- Statistics combines mathematical and non-mathematical procedures into one discipline.
- Statistics is generally broken down into two categories: descriptive statistics and inferential statistics.
- Statistics is an applied science and is used in many fields, including the natural and social sciences, government, and business.
- The use of statistical methods dates back to at least the 5th century BC.

Key Terms

statistics
: a mathematical science concerned with data collection, presentation, analysis, and interpretation

empirical
: verifiable by means of scientific experimentation

Example

Say you want to conduct a poll on whether your school should use its funding to build a new athletic complex or a new library. Appropriate questions to ask would include: How many people do you have to poll? How do you ensure that your poll is free of bias? How do you interpret your results?

Statistics Overview

Statistics is the study of the collection, organization, analysis, interpretation, and presentation of data. It deals with all aspects of data, including the planning of its collection in terms of the design of surveys and experiments. Some consider statistics a mathematical body of science that pertains to the collection, analysis, interpretation or explanation, and presentation of data, while others consider it a branch of mathematics concerned with collecting and interpreting data. Because of its empirical roots and its focus on applications, statistics is usually considered a distinct mathematical science rather than a branch of mathematics. As one would expect, statistics is largely grounded in mathematics, and the study of statistics has lent itself to many major concepts in mathematics, such as:

- probability,
- distributions ,
- samples and populations,
- estimation, and
- data analysis.

However, much of statistics is also non-mathematical. This includes:

- ensuring that data collection is undertaken in a way that produces valid conclusions,
- coding and archiving data so that information is retained and made useful for international comparisons of official statistics,
- reporting of results and summarized data (tables and graphs) in ways comprehensible to those who must use them, and
- implementing procedures that ensure the privacy of census information.

In short, statistics is the study of data. It includes descriptive statistics (the study of methods and tools for collecting data, and mathematical models to describe and interpret data) and inferential statistics (the systems and techniques for making probability-based decisions and accurate predictions based on incomplete data).

How Do We Use Statistics?

A statistician is someone who is particularly well-versed in the ways of thinking necessary to successfully apply statistical analysis. Such people often gain experience through working in any of a wide number of fields. Statisticians improve data quality by developing specific experimental designs and survey samples. Statistics itself also provides tools for predicting and forecasting the use of data and statistical models. Statistics is applicable to a wide variety of academic disciplines, including natural and social sciences, government, and business. Statistical consultants can help organizations and companies that don't have in-house expertise relevant to their particular questions.

History of Statistics

Statistical methods date back at least to the 5^{th} century BC. The earliest known writing on statistics appears in a 9^{th} century book entitled *Manuscript on Deciphering Cryptographic Messages*, written by Al-Kindi. In this book, Al-Kindi provides a detailed description of how to use statistics and frequency analysis to decipher encrypted messages. This was the birth of both statistics and cryptanalysis, according to the Saudi engineer Ibrahim Al-Kadi.

The *Nuova Cronica*, a 14^{th} century history of Florence by the Florentine banker and official Giovanni

Villani, includes much statistical information on population, ordinances, commerce, education, and religious facilities, and has been described as the first introduction of statistics as a positive element in history.

Some scholars pinpoint the origin of statistics to 1663, with the publication of *Natural and Political Observations upon the Bills of Mortality* by John Graunt. Early applications of statistical thinking revolved around the needs of states to base policy on demographic and economic data, hence its "stat-" etymology. The scope of the discipline of statistics broadened in the early 19th century to include the collection and analysis of data in general.

1.1.3: The Purpose of Statistics

Statistics teaches people to use a limited sample to make intelligent and accurate conclusions about a greater population.

Learning Objective

Describe how Statistics helps us to make inferences about a population, understand and interpret variation, and make more informed everyday decisions.

Key Takeaways

Key Points

- Statistics is an extremely powerful tool available for assessing the significance of experimental data and for drawing the right conclusions from it.
- Statistics helps scientists, engineers, and many other professionals draw the right conclusions from experimental data.

- Variation is ubiquitous in nature, and probability and statistics are the fields that allow us to study, understand, model, embrace and interpret this variation.

Key Terms

sample
: a subset of a population selected for measurement, observation, or questioning to provide statistical information about the population

population
: a group of units (persons, objects, or other items) enumerated in a census or from which a sample is drawn

Example

A company selling the cat food brand “Cato” (a fictitious name here), may claim quite truthfully in their advertisements that eight out of ten cat owners said that their cats preferred Cato brand cat food to “the other leading brand” cat food. What they may not mention is that the cat owners questioned were those they found in a supermarket buying Cato, which doesn’t represent an unbiased sample of cat owners.

Imagine reading a book for the first few chapters and then being able to get a sense of what the ending will be like. This ability is provided by the field of inferential statistics. With the appropriate tools and solid grounding in the field, one can use a limited sample (e.g., reading the first five chapters of Pride & Prejudice) to make intelligent and accurate statements about the population (e.g., predicting the ending of Pride & Prejudice).

Those proceeding to higher education will learn that statistics is an extremely powerful tool available for assessing the significance of experimental data and for drawing the right conclusions from the vast amounts of data encountered by engineers, scientists, sociologists, and other professionals in most spheres of learning. There is no study with scientific, clinical, social, health, environmental or political goals that does not rely on statistical methodologies. The most essential reason for this fact is that variation is ubiquitous in nature, and probability and statistics are the fields that allow us to study, understand, model, embrace and interpret this variation.

In today's information-overloaded age, statistics is one of the most useful subjects anyone can learn. Newspapers are filled with statistical data, and anyone who is ignorant of statistics is at risk of being seriously misled about important real-life decisions such as what to eat, who is leading the polls, how dangerous smoking is, et cetera. Statistics are often used by politicians, advertisers, and others to twist the truth for their own gain. Knowing at least a little about the field of statistics will help one to make more informed decisions about these and other important questions.

The Purpose of Statistics

Statistics teaches people to use a limited sample to make intelligent and accurate conclusions about a greater population. The use of tables, graphs, and charts play a vital role in presenting the data being used to draw these conclusions.

1.1.4: Inferential Statistics

The mathematical procedure in which we make intelligent guesses about a population based on a sample is called inferential statistics.

Learning Objective

Discuss how inferential statistics allows us to draw conclusions about a population from a random sample and corresponding tests of significance.

Key Takeaways

Key Points

- Inferential statistics is used to describe systems of procedures that can be used to draw conclusions from data sets arising from systems affected by random variation, such as observational errors, random sampling, or random experimentation.
- Samples must be representative of the entire population in order to induce a conclusion about that population.
- Statisticians use tests of significance to determine the probability that the results were found by chance.

Key Term

inferential statistics
: A branch of mathematics that involves drawing conclusions about a population based on sample data drawn from it.

In statistics, statistical inference is the process of drawing conclusions from data that is subject to random variation–for example, observational errors or sampling variation. More substantially, the terms statistical inference, statistical induction, and inferential statistics are used to describe systems of procedures that can be used to draw conclusions from data sets arising from systems affected by random variation, such as observational errors, random sampling, or random experimentation. Initial requirements of such a system of procedures for inference and induction are that the system should produce reasonable answers when applied to well-defined situations and that it should be general enough to be applied across a range of situations.

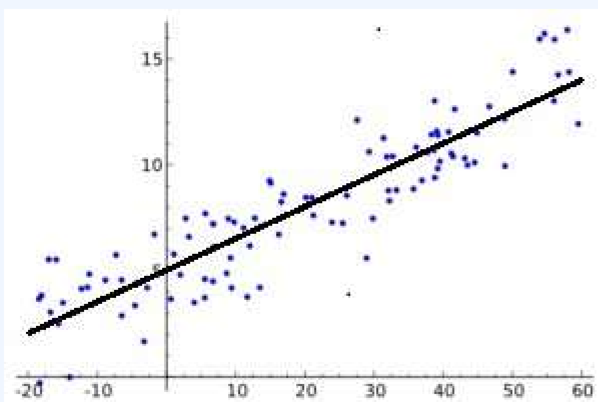
The outcome of statistical inference may be an answer to the question "what should be done next? " where this might be a decision about making further experiments or surveys, or about drawing a conclusion before implementing some organizational or governmental policy.

Suppose you have been hired by the National Election Commission to examine how the American people feel about the fairness of the voting procedures in the U.S. How will you do it? Who will you ask?

It is not practical to ask every single American how he or she feels about the fairness of the voting procedures. Instead, we query a relatively small number of Americans, and draw inferences about the entire country from their responses. The Americans actually queried constitute our sample of the larger population of all Americans. The mathematical procedures whereby we convert information about the sample into intelligent guesses about the population fall under the rubric of inferential statistics.

In the case of voting attitudes, we would sample a few thousand Americans, drawn from the hundreds of millions that make up the country. In choosing a sample, it is therefore crucial that it be representative. It must not over-represent one kind of citizen at the expense of others. For example, something would be wrong with our sample if it happened to be made up entirely of Florida residents. If the sample held only Floridians, it could not be used to infer the attitudes of other Americans. The same problem would arise if the sample were comprised only of Republicans. Inferential statistics are based on the assumption that sampling is random. We trust a random sample to represent different segments of society in close to the appropriate proportions (provided the sample is large enough).

Furthermore, when generalizing a trend found in a sample to the larger population, statisticians uses tests of significance (such as the Chi-Square test or the T-test). These tests determine the probability that the results found were by chance, and therefore not representative of the entire population.



Linear Regression in Inferential Statistics

This graph shows a linear regression model, which is a tool used to make inferences in statistics.

1.1.5: Types of Data

Data can be categorized as either primary or secondary and as either qualitative or quantitative.

Learning Objective

Differentiate between primary and secondary data and qualitative and quantitative data.

Key Takeaways

Key Points

- Primary data is data collected first-hand. Secondary data is data reused from another source.
- Qualitative data is a categorical measurement expressed not in terms of numbers, but rather by means of a natural language description.
- Quantitative data is a numerical measurement expressed not by means of a natural language description, but rather in terms of numbers.

Key Terms

primary data
: data that has been compiled for a specific purpose, and has not been collated or merged with others

qualitative data
: data centered around descriptions or distinctions based on some quality or characteristic rather than on some quantity or measured value

quantitative
: of a measurement based on some quantity or number rather than on some quality

Example**s**

Qualitative data: race, religion, gender, etc. Quantitative data: height in inches, time in seconds, temperature in degrees, etc.

Primary and Secondary Data

Data can be classified as either primary or secondary. Primary data is original data that has been collected specially for the purpose in mind. This type of data is collected first hand. Those who gather primary data may be an authorized organization, investigator, enumerator or just someone with a clipboard. These people are acting as a witness, so primary data is only considered as reliable as the people who gather it. Research where one gathers this kind of data is referred to as field research. An example of primary data is conducting your own questionnaire.

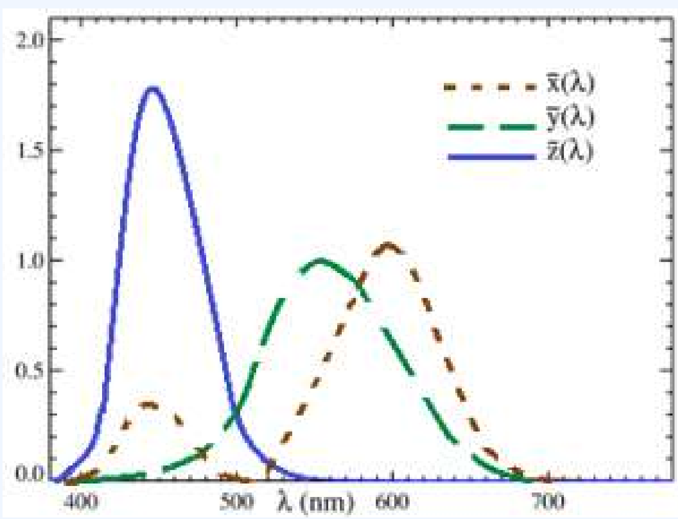
Secondary data is data that has been collected for another purpose. This type of data is reused, usually in a different context from its first use. You are not the original source of the data–rather, you are collecting it from elsewhere. An example of secondary data is using numbers and information found inside a textbook.

Knowing how the data was collected allows critics of a study to search for bias in how it was conducted. A good study will welcome such scrutiny. Each type has its own weaknesses and strengths. Primary data is gathered by people who can focus directly on the purpose in mind. This helps ensure that questions are meaningful to the purpose, but this can introduce bias in those same questions. Secondary data doesn't have the privilege of this focus, but is only susceptible to bias introduced in the choice of what data to reuse. Stated another way, those who gather secondary data get to pick the questions. Those who gather primary data get to write the questions. There may be bias either way.

Qualitative and Quantitative Data

Qualitative data is a categorical measurement expressed not in terms of numbers, but rather by means of a natural language description. In statistics, it is often used interchangeably with "categorical" data. Collecting information about a favorite color is an example of collecting qualitative data. Although we may have categories, the categories may have a structure to them. When there is not a natural ordering of the categories, we call these nominal categories. Examples might be gender, race, religion, or sport. When the categories may be ordered, these are called ordinal categories. Categorical data that judge size (small, medium, large, etc.) are ordinal categories. Attitudes (strongly disagree, disagree, neutral, agree, strongly agree) are also ordinal categories; however, we may not know which value is the best or worst of these issues. Note that the distance between these categories is not something we can measure.

Quantitative data is a numerical measurement expressed not by means of a natural language description, but rather in terms of numbers. Quantitative data always are associated with a scale measure. Probably the most common scale type is the ratio-scale. Observations of this type are on a scale that has a meaningful zero value but also have an equidistant measure (i.e. the difference between 10 and 20 is the same as the difference between 100 and 110). For example, a 10 year-old girl is twice as old as a 5 year-old girl. Since you can measure zero years, time is a ratio-scale variable. Money is another common ratio-scale quantitative measure. Observations that you count are usually ratio-scale (e.g. number of widgets). A more general quantitative measure is the interval scale. Interval scales also have an equidistant measure. However, the doubling principle breaks down in this scale. A temperature of 50 degrees Celsius is not "half as hot" as a temperature of 100, but a difference of 10 degrees indicates the same difference in temperature anywhere along the scale.



Quantitative Data: The graph shows a display of quantitative data.

1.1.6: Applications of Statistics

Statistics deals with all aspects of the collection, organization, analysis, interpretation, and presentation of data.

Learning Objective

Describe how statistics is applied to scientific, industrial, and societal problems.

Key Takeaways

Key Points

- Statistics can be used to improve data quality by developing specific experimental designs and survey samples.
- Statistics includes the planning of data collection in terms of the design of surveys and experiments.
- Statistics provides tools for prediction and forecasting and is applicable to a wide variety of academic disciplines, including natural and social sciences, as well as government, and business.

Key Terms

population
: a group of units (persons, objects, or other items) enumerated in a census or from which a sample is drawn

statistics
: The study of the collection, organization, analysis, interpretation, and presentation of data.

sample
: a subset of a population selected for measurement, observation, or questioning to provide statistical information about the population

Example

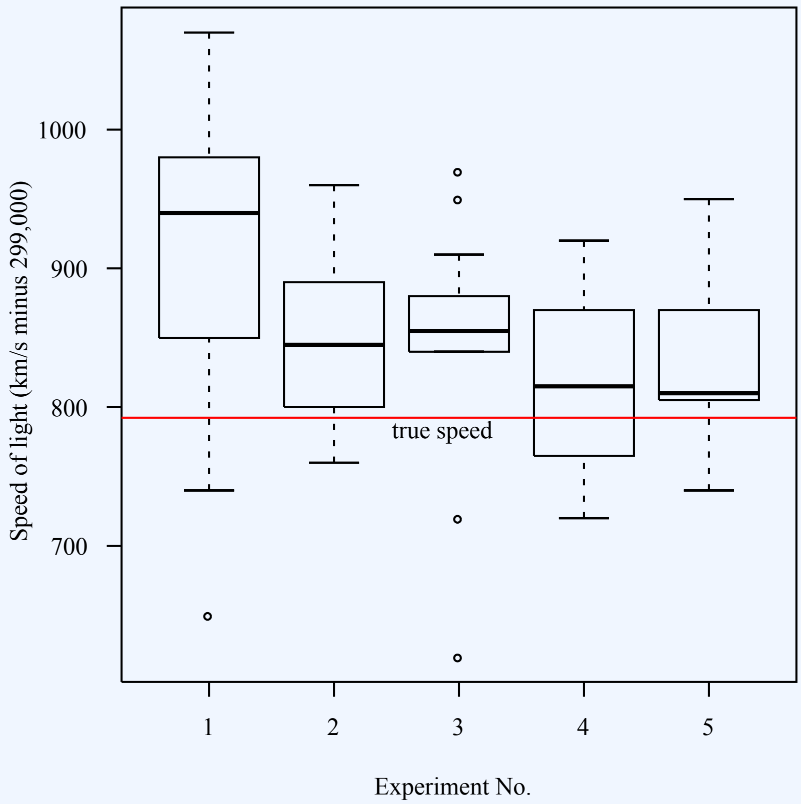
In calculating the arithmetic mean of a sample, for example, the algorithm works by summing all the data values observed in the sample and then dividing this sum by the number of data items. This single measure, the mean of the sample, is called a statistic; its value is frequently used as an estimate of the mean value of all items comprising the population from which the sample is drawn. The population mean is also a single measure; however, it is not called a statistic; instead it is called a population parameter.

Statistics deals with all aspects of the collection, organization, analysis, interpretation, and presentation of data. It includes the planning of data collection in terms of the design of surveys and experiments.

Statistics can be used to improve data quality by developing specific experimental designs and survey samples. Statistics also provides tools for prediction and forecasting. Statistics is applicable to a wide variety of academic disciplines, including natural and social sciences as well as government and business. Statistical consultants can help organizations and companies that don’t have in-house expertise relevant to their particular questions.

Descriptive and Inferential Statistics

Statistical methods can summarize or describe a collection of data. This is called descriptive statistics . This is particularly useful in communicating the results of experiments and research. Statistical models can also be used to draw statistical inferences about the process or population under study—a practice called inferential statistics. Inference is a vital element of scientific advancement, since it provides a way to draw conclusions from data that are subject to random variation. Conclusions are tested in order to prove the propositions being investigated further, as part of the scientific method. Descriptive statistics and analysis of the new data tend to provide more information as to the truth of the proposition.



Summary statistics: In descriptive statistics, summary statistics are used to summarize a set of observations, in order to communicate the largest amount as simply as possible. This Boxplot represents Michelson and Morley's data on the speed of light. It consists of five experiments, each made of 20 consecutive runs.

The Statistical Process

When applying statistics to a scientific, industrial, or societal problems, it is necessary to begin with a population or process to be studied. Populations can be diverse topics such as "all persons living in a country" or "every atom composing a crystal". A population can also be composed of observations of a process at various times, with the data from each observation serving as a different member of the overall group. Data collected about this kind of "population" constitutes what is called a time series. For practical reasons, a chosen subset of the population called a sample is studied—as opposed to compiling data about the entire group (an operation called census). Once a sample that is representative of the population is determined, data is collected for the sample members in an observational or experimental setting. This data can then be subjected to statistical analysis, serving two related purposes: description and inference.

Descriptive statistics summarize the population data by describing what was observed in the sample numerically or graphically. Numerical descriptors include mean and standard deviation for continuous data types (like heights or weights), while frequency and percentage are more useful in terms of describing categorical data (like race). Inferential statistics uses patterns in the sample data to draw inferences about the population represented, accounting for randomness. These inferences may take the form of: answering yes/no questions about the data (*hypothesis testing*), estimating numerical characteristics of the data (*estimation*), describing associations within the data (correlation) and modeling relationships within the data (for example, using regression *analysis*). Inference can extend to forecasting, prediction and estimation of unobserved values either in or associated with the population being studied. It can include extrapolation and interpolation of time series or spatial data and can also include data mining.

Statistical Analysis

Statistical analysis of a data set often reveals that two variables of the population under consideration tend to vary together, as if they were connected. For example, a study of annual income that also looks at age of death might find that poor people tend to have shorter lives than affluent people. The two variables are said to be correlated; however, they may or may not be the cause of one another. The correlation could be caused by a third, previously unconsidered phenomenon, called a confounding variable. For this reason, there is no way to immediately infer the existence of a causal relationship between the two variables.

To use a sample as a guide to an entire population, it is important that it truly represent the overall population. Representative sampling assures that inferences and conclusions can safely extend from the sample to the population as a whole. A major problem lies in determining the extent that the sample chosen is actually representative. Statistics offers methods to estimate and correct for any random trending within the sample and data collection procedures. There are also methods of experimental design for experiments that can lessen these issues at the outset of a study, strengthening its capability to discern truths about the population. Randomness is studied using the mathematical discipline of probability theory. Probability is used in “mathematical statistics” (alternatively, “statistical theory”) to study the sampling distributions of sample statistics and, more generally, the properties of statistical procedures. The use of any statistical method is valid when the system or population under consideration satisfies the assumptions of the method.

1.1.7: Fundamentals of Statistics

In applying statistics to a scientific, industrial, or societal problem, it is necessary to begin with a population or process to be studied.

Learning Objective

Recall that the field of Statistics involves using samples to make inferences about populations and describing how variables relate to each other.

Key Takeaways

Key Points

- For practical reasons, a chosen subset of the population called a sample is studied—as opposed to compiling data about the entire group (an operation called census).
- Descriptive statistics summarizes the population data by describing what was observed in the sample numerically or graphically.
- Inferential statistics uses patterns in the sample data to draw inferences about the population represented, accounting for randomness.
- Statistical analysis of a data set often reveals that two variables (properties) of the population under consideration tend to vary together, as if they were connected.
- To use a sample as a guide to an entire population, it is important that it truly represent the overall population.

Key Terms

sample
: a subset of a population selected for measurement, observation, or questioning to provide statistical information about the population

variable
: a quantity that may assume any one of a set of values

population
: a group of units (persons, objects, or other items) enumerated in a census or from which a sample is drawn

Example

A population can be composed of observations of a process at various times, with the data from each observation serving as a different member of the overall group. Data collected about this kind of "population" constitutes what is called a time series.

In applying statistics to a scientific, industrial, or societal problem, it is necessary to begin with a population or process to be studied. Populations can be diverse topics such as "all persons living in a country" or "every atom composing a crystal.". A population can also be composed of observations of a process at various times, with the data from each observation serving as a different member of the overall group. Data collected about this kind of "population" constitutes what is called a time series.

For practical reasons, a chosen subset of the population called a sample is studied—as opposed to compiling data about the entire group (an operation called census). Once a sample that is representative of the population is determined, data is collected for the sample members in an observational or experimental setting. This data can then be subjected to statistical analysis, serving two related purposes: description and inference.

- *Descriptive statistics* summarizes the population data by describing what was observed in the sample numerically or graphically. Numerical descriptors include mean and standard deviation for continuous data types (like heights or weights), while frequency and percentages are more useful in terms of describing categorical data (like race).
- *Inferential statistics* uses patterns in the sample data to draw inferences about the population represented, accounting for randomness. These inferences may take the form of: answering yes/no questions about the data (hypothesis testing), estimating numerical characteristics of the data (estimation), describing associations within the data (correlation) and modeling relationships within the data (for example, using regression analysis). Inference can extend to forecasting, prediction and estimation of unobserved values either in or associated with the population being studied. It can include

extrapolation and interpolation of time series or spatial data, and can also include data mining.

The concept of correlation is particularly noteworthy for the potential confusion it can cause. Statistical analysis of a data set often reveals that two variables (properties) of the population under consideration tend to vary together, as if they were connected. For example, a study of annual income that also looks at age of death might find that poor people tend to have shorter lives than affluent people. The two variables are said to be correlated; however, they may or may not be the cause of one another. The correlation phenomena could be caused by a third, previously unconsidered phenomenon, called a confounding variable. For this reason, there is no way to immediately infer the existence of a causal relationship between the two variables.

To use a sample as a guide to an entire population, it is important that it truly represent the overall population. Representative sampling assures that inferences and conclusions can safely extend from the sample to the population as a whole. A major problem lies in determining the extent that the sample chosen is actually representative. Statistics offers methods to estimate and correct for any random trending within the sample and data collection procedures. There are also methods of experimental design for experiments that can lessen these issues at the outset of a study, strengthening its capability to discern truths about the population.

Randomness is studied using the mathematical discipline of probability theory. Probability is used in "mathematical statistics" (alternatively, "statistical theory") to study the sampling distributions of sample statistics and, more generally, the properties of statistical procedures. The use of any statistical method is valid when the system or population under consideration satisfies the assumptions of the method.

1.1.8: Critical Thinking

The essential skill of critical thinking will go a long way in helping one to develop statistical literacy.

Learning Objective

Interpret the role that the process of critical thinking plays in statistical literacy.

Key Takeaways

Key Points

- Statistics can be made to produce misrepresentations of data that may seem valid.
- Statistical literacy is necessary to understand what makes a poll trustworthy and to properly weigh the value of poll results and conclusions.
- Critical thinking is a way of deciding whether a claim is always true, sometimes true, partly true, or false.
- The list of core critical thinking skills includes observation, interpretation, analysis, inference, evaluation, explanation, and meta-cognition.

Key Terms

statistical literacy
: the ability to understand statistics, necessary for citizens to understand mateiral presented in publications such as newspapers, television, and the Internet

critical thinking
: the application of logical principles, rigorous standards of evidence, and careful reasoning to the analysis and discussion of claims, beliefs, and issues

Each day people are inundated with statistical information from advertisements ("4 out of 5 dentists recommend"), news reports ("opinion polls show the incumbent leading by four points"), and even general conversation ("half the time I don't know what you're talking about"). Experts and advocates often use numerical claims to bolster their arguments, and statistical literacy is a necessary skill to help one decide what experts mean and which advocates to believe. This is important because statistics can be made to produce misrepresentations of data that may seem valid. The aim of statistical literacy is to improve the public understanding of numbers and figures.

For example, results of opinion polling are often cited by news organizations, but the quality of such polls varies considerably. Some understanding of the statistical technique of sampling is necessary in order to be able to correctly interpret polling results. Sample sizes may be too small to draw meaningful conclusions, and samples may be biased. The wording of a poll question may introduce a bias, and thus can even be used intentionally to produce a biased result. Good polls use unbiased techniques, with much time and effort being spent in the design of the questions and polling strategy. Statistical literacy is necessary to understand what makes a poll trustworthy and to properly weigh the value of poll results and conclusions.

Critical Thinking

The essential skill of critical thinking will go a long way in helping one to develop statistical literacy. Critical thinking is a way of deciding whether a claim is always true, sometimes true, partly true, or false. The list of core critical thinking skills includes observation, interpretation, analysis, inference, evaluation, explanation, and meta-cognition. There is a reasonable level of consensus that an individual or group engaged in strong critical thinking gives due consideration to establish:

- Evidence through observation,
- Context skills,
- Relevant criteria for making the judgment well,
- Applicable methods or techniques for forming the judgment, and
- Applicable theoretical constructs for understanding the problem and the question at hand.

Critical thinking calls for the ability to:

- Recognize problems, to find workable means for meeting those problems,
- Understand the importance of prioritization and order of precedence in problem solving,
- Gather and marshal pertinent (relevant) information,
- Recognize unstated assumptions and values,
- Comprehend and use language with accuracy, clarity, and discernment,
- Interpret data, to appraise evidence and evaluate arguments,
- Recognize the existence (or non-existence) of logical relationships between propositions,
- Draw warranted conclusions and generalizations,
- Put to test the conclusions and generalizations at which one arrives,
- Reconstruct one’s patterns of beliefs on the basis of wider experience, and
- Render accurate judgments about specific things and qualities in everyday life.

Critical Thinking

Critical thinking is an inherent part of data analysis and statistical literacy.

1.1.9: Experimental Design

Experimental design is the design of studies where variation, which may or may not be under full control of the experimenter, is present.

Learning Objective

Outline the methodology for designing experiments in terms of comparison, randomization, replication, blocking, orthogonality, and factorial experiments

Key Takeaways

Key Points

- The experimenter is often interested in the effect of some process or intervention (the "treatment") on some objects (the "experimental units"), which may be people, parts of people, groups of people, plants, animals, etc.
- A methodology for designing experiments involves comparison, randomization, replication, blocking, orthogonality, and factorial considerations.
- It is best that a process be in reasonable statistical control prior to conducting designed experiments.
- One of the most important requirements of experimental research designs is the necessity of eliminating the effects of spurious, intervening, and antecedent variables.

Key Terms

dependent variable
: in an equation, the variable whose value depends on one or more variables in the equation

independent variable
: in an equation, any variable whose value is not dependent on any other in the equation

experiment
: A test under controlled conditions made to either demonstrate a known truth, examine the validity of a hypothesis, or determine the efficacy of something previously untried.

Example

For example, if a researcher feeds an experimental artificial sweetener to sixty laboratory rats and observes that ten of them subsequently become sick, the underlying cause could be the sweetener itself or something unrelated. Other variables, which may not be readily obvious, may interfere with the experimental design. For instance, perhaps the rats were simply not supplied with enough food or water, or the water was contaminated and undrinkable, or the rats were under some psychological or physiological stress, etc. Eliminating each of these possible explanations individually would be time-consuming and difficult. However, if a control group is used that does not receive the sweetener but is otherwise treated identically, any difference between the two groups can be ascribed to the sweetener itself with much greater confidence.

In general usage, design of experiments or experimental design is the design of any information-gathering exercises where variation is present, whether under the full control of the experimenter or not. Formal planned experimentation is often used in evaluating physical objects, chemical formulations, structures, components, and materials. In the design of experiments, the experimenter is often interested in the effect of some process or intervention (the "treatment") on some objects (the "experimental units"), which may be people, parts of people, groups of people, plants, animals, etc. Design of experiments is thus a discipline that has very broad application across all the natural and social sciences and engineering.

A methodology for designing experiments was proposed by Ronald A. Fisher in his innovative books *The Arrangement of Field Experiments* (1926) and *The Design of Experiments* (1935). These methods have been broadly adapted in the physical and social sciences.

Old-fashioned scale

A scale is emblematic of the methodology of experimental design which includes comparison, replication, and factorial considerations.

- Comparison: In some fields of study it is not possible to have independent measurements to a traceable standard. Comparisons between treatments are much more valuable and are usually preferable. Often one compares against a scientific control or traditional treatment that acts as baseline.
- Randomization: Random assignment is the process of assigning individuals at random to groups or to different groups in an experiment. The random assignment of individuals to groups (or conditions within a group) distinguishes a rigorous, "true" experiment from an adequate, but less-than-rigorous, "quasi-experiment". Random does not mean haphazard, and great care must be taken that appropriate random methods are used.
- Replication: Measurements are usually subject to variation and uncertainty. Measurements are repeated and full experiments are replicated to help identify the sources of variation, to better estimate the true effects of treatments, to further strengthen the experiment's reliability and validity, and to add to the existing knowledge of the topic.
- Blocking: Blocking is the arrangement of experimental units into groups (blocks) consisting of units that are similar to one another. Blocking reduces known but irrelevant sources of variation between units and thus allows greater precision in the estimation of the source of variation under study.
- Orthogonality: Orthogonality concerns the forms of comparison (contrasts) that can be legitimately and efficiently carried out. Contrasts can be represented by vectors and sets of orthogonal contrasts are uncorrelated and independently distributed if the data are normal. Because of this independence, each orthogonal treatment provides different information to the others. If there are T">T treatments and T−1">T−1 orthogonal contrasts, all the information that can be captured from the experiment is obtainable from the set of contrasts.
- Factorial experiments: Use of factorial experiments instead of the one-factor-at-a-time method. These

> are efficient at evaluating the effects and possible interactions of several factors (independent variables). Analysis of experiment design is built on the foundation of the analysis of variance, a collection of models that partition the observed variance into components, according to what factors the experiment must estimate or test.

It is best that a process be in reasonable statistical control prior to conducting designed experiments. When this is not possible, proper blocking, replication, and randomization allow for the careful conduct of designed experiments. To control for nuisance variables, researchers institute control checks as additional measures. Investigators should ensure that uncontrolled influences (e.g., source credibility perception) are measured do not skew the findings of the study.

One of the most important requirements of experimental research designs is the necessity of eliminating the effects of spurious, intervening, and antecedent variables. In the most basic model, cause (X) leads to effect (Y">Y). But there could be a third variable (Z">Z) that influences (Y">Y), and X">X might not be the true cause at all. Z">Z is said to be a spurious variable and must be controlled for. The same is true for intervening variables (a variable in between the supposed cause (X">X) and the effect (Y">Y)), and anteceding variables (a variable prior to the supposed cause (X">X) that is the true cause). In most designs, only one of these causes is manipulated at a time.

1.1.10: Random Samples

An unbiased random selection of individuals is important so that in the long run, the sample represents the population.

Learning Objective

Explain how simple random sampling leads to every object having the same possibility of being chosen.

Key Takeaways

Key Points

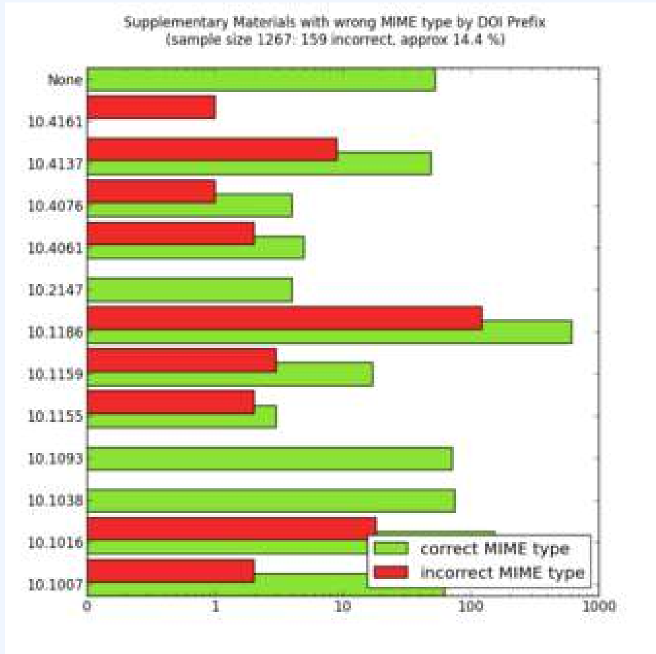
- Simple random sampling merely allows one to draw externally valid conclusions about the entire population based on the sample.
- Advantages of random sampling are that it is free of classification error, and it requires minimum advance knowledge of the population other than the frame.
- Simple random sampling best suits situations where not much information is available about the population and data collection can be efficiently conducted on randomly distributed items, or where the cost of sampling is small enough to make efficiency less important than simplicity.

Key Terms

population
: a group of units (persons, objects, or other items) enumerated in a census or from which a sample is drawn

random sample
: a sample randomly taken from an investigated population

Sampling is concerned with the selection of a subset of individuals from within a statistical population to estimate characteristics of the whole population . Two advantages of sampling are that the cost is lower and data collection is faster than measuring the entire population.



Random Sampling

MIME types of a random sample of supplementary materials from the Open Access subset in PubMed Central as of October 23, 2012. The colour code means that the MIME type of the supplementary files is indicated correctly (green) or incorrectly (red) in the XML at PubMed Central.

Each observation measures one or more properties (such as weight, location, color) of observable bodies distinguished as independent objects or individuals. In survey sampling, weights can be applied to the data to adjust for the sample design, particularly stratified sampling (blocking). Results from probability theory and statistical theory are employed to guide practice. In business and medical research, sampling is widely used for gathering information about a population.

A simple random sample is a subset of individuals chosen from a larger set (a population). Each individual is chosen randomly and entirely by chance, such that each individual has the same probability of being chosen at any stage during the sampling process and each subset of **k** individuals has the same probability of being chosen for the sample as any other subset of **k** individuals. A simple random sample is an unbiased surveying technique.

Simple random sampling is a basic type of sampling, since it can be a component of other more complex sampling methods. The principle of simple random sampling is that every object has the same possibility to be chosen. For example, **N** college students want to get a ticket for a basketball game, but there are not enough tickets (**X**) for them, so they decide to have a fair way to see who gets to go. Then, everybody is given a number (**0 to N-1**), and random numbers are generated. The first **X** numbers would be the lucky ticket winners.

In small populations and often in large ones, such sampling is typically done "without replacement" (i.e., one deliberately avoids choosing any member of the population more than once). Although simple random sampling can be conducted with replacement instead, this is less common and would normally be described more fully as simple random sampling with replacement. Sampling done without replacement is no longer independent, but still satisfies exchangeability. Hence, many results still hold. Further, for a small sample from a large population, sampling without replacement is approximately the same as sampling with replacement, since the odds of choosing the same individual twice is low.

An unbiased random selection of individuals is important so that, in the long run, the sample represents the population. However, this does not guarantee that a particular sample is a perfect representation of the population. Simple random sampling merely allows one to draw externally valid conclusions about the entire population based on the sample.

Conceptually, simple random sampling is the simplest of the probability sampling techniques. It requires a complete sampling frame, which may not be available or feasible to construct for large populations. Even if a complete frame is available, more efficient approaches may be possible if other useful information is available about the units in the population.

Advantages are that it is free of classification error, and it requires minimum advance knowledge of the population other than the frame. Its simplicity also makes it relatively easy to interpret data collected via SRS. For these reasons, simple random sampling best suits situations where not much information is available about the population and data collection can be efficiently conducted on randomly distributed items, or where the cost of sampling is small enough to make efficiency less important than simplicity. If these conditions are not true, stratified sampling or cluster sampling may be a better choice.

Attributions

- Collecting and Measuring Data
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Wikipedia, the free encyclopedia 无标题页 ."
    http://121.101.213.122/wiki/writeTxt.aspx?bookinfoid=1812482&school=testip&key=Statistics.
    http://121.101.213.122/wiki/writeTxt.aspx?bookinfoid=1812482&school=testip&key=Statistics
    CC BY-SA.
  - "Levels of measurement."
    https://en.wikipedia.org/wiki/Levels_of_measurement.

Wikipedia
CC BY-SA 3.0.

- "sampling."
http://en.wiktionary.org/wiki/sampling.
Wiktionary
CC BY-SA 3.0.
- "population."
http://en.wiktionary.org/wiki/population.
Wiktionary
CC BY-SA 3.0.
- "Centenarul artileriei române."
http://commons.wikimedia.org/wiki/File:Centenarul_artileriei_rom%C3%A2ne.jpg.
Wikimedia
CC BY-SA.

- What Is Statistics?
  - "Boundless."
  http://www.boundless.com/.
  Boundless Learning
  CC BY-SA 3.0.
  - "Statistics."
  http://en.wikipedia.org/wiki/Statistics.
  Wikipedia
  CC BY-SA 3.0.
  - "empirical."
  http://en.wiktionary.org/wiki/empirical.
  Wiktionary
  CC BY-SA 3.0.
  - "statistics."
  http://en.wiktionary.org/wiki/statistics.
  Wiktionary
  CC BY-SA 3.0.
  - "Statistics/Introduction/What is Statistics."
  http://en.wikibooks.org/wiki/Statistics/Introduction/What_is_Statistics.
  Wikibooks
  CC BY-SA 3.0.
- The Purpose of Statistics
  - "Boundless."

http://www.boundless.com/.
Boundless Learning
CC BY-SA 3.0.
  - "Misuse of statistics."
    https://en.wikipedia.org/wiki/Misuse_of_statistics.
    Wikipedia
    CC BY-SA 3.0.
  - "population."
    http://en.wiktionary.org/wiki/population.
    Wiktionary
    CC BY-SA 3.0.
  - "sample."
    http://en.wiktionary.org/wiki/sample.
    Wiktionary
    CC BY-SA 3.0.
  - "Creative Commons — CC0 1.0 Universal."
    http://creativecommons.org/publicdomain/zero/1.0/.
    Creative Commons
    CC BY.
- Inferential Statistics
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "inferential statistics."
    http://en.wiktionary.org/wiki/inferential_statistics.
    Wiktionary
    CC BY-SA 3.0.
  - "Statistical inference."
    http://en.wikipedia.org/wiki/Statistical_inference.
    Wikipedia
    CC BY-SA 3.0.
  - "David Lane, Inferential Statistics. September 17, 2013."
    http://cnx.org/content/m10185/latest/.
    OpenStax CNX
    CC BY 3.0.
  - "Linear regression."

http://en.wikipedia.org/wiki/File:Linear_regression.svg.
Wikipedia
CC BY-SA.

- Types of Data
    - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
    - "primary data."
    http://en.wiktionary.org/wiki/primary_data.
    Wiktionary
    CC BY-SA 3.0.
    - "qualitative data."
    http://en.wikipedia.org/wiki/qualitative%20data.
    Wikipedia
    CC BY-SA 3.0.
    - "Statistics/Different Types of Data/PS."
    http://en.wikibooks.org/wiki/Statistics/Different_Types_of_Data/PS.
    Wikibooks
    CC BY-SA 3.0.
    - "quantitative."
    http://en.wiktionary.org/wiki/quantitative.
    Wiktionary
    CC BY-SA 3.0.
    - "Terms of Use – Creative Commons."
    http://creativecommons.org/terms.
    Creative Commons
    CC BY.
- Applications of Statistics
    - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
    - "Statistic."
    http://en.wikipedia.org/wiki/Statistic.
    Wikipedia
    CC BY-SA 3.0.

  - "Statistics."
    https://en.wikipedia.org/wiki/Statistics.
    Wikipedia
    CC BY-SA 3.0.
  - "Statistics."
    http://en.wikipedia.org/wiki/Statistics.
    Wikipedia
    CC BY-SA 3.0.
  - "statistics."
    http://en.wiktionary.org/wiki/statistics.
    Wiktionary
    CC BY-SA 3.0.
  - "population."
    http://en.wiktionary.org/wiki/population.
    Wiktionary
    CC BY-SA 3.0.
  - "sample."
    http://en.wiktionary.org/wiki/sample.
    Wiktionary
    CC BY-SA 3.0.
  - "Summary statistics."
    http://en.wikipedia.org/wiki/Summary_statistics.
    Wikipedia
    Public domain.
- Fundamentals of Statistics
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Wikipedia, the free encyclopedia 无."
    http://121.101.213.122/wiki/
    writeTxt.aspx?bookinfoid=1812482&school=testip&key=Statistics.
    http://121.101.213.122/wiki/
    writeTxt.aspx?bookinfoid=1812482&school=testip&key=Statistics
    CC BY-SA.
  - "Outline of statistics."
    http://en.wikipedia.org/wiki/Outline_of_statistics.

Wikipedia
CC BY-SA 3.0.

- "Statistics."
  http://en.wikipedia.org/wiki/Statistics.
  Wikipedia
  CC BY-SA 3.0.
- "variable."
  http://en.wiktionary.org/wiki/variable.
  Wiktionary
  CC BY-SA 3.0.
- "population."
  http://en.wiktionary.org/wiki/population.
  Wiktionary
  CC BY-SA 3.0.
- "sample."
  http://en.wiktionary.org/wiki/sample.
  Wiktionary
  CC BY-SA 3.0.

- Critical Thinking
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "critical thinking."
    http://en.wiktionary.org/wiki/critical_thinking.
    Wiktionary
    CC BY-SA 3.0.
  - "Statistical literacy."
    http://en.wikipedia.org/wiki/Statistical_literacy.
    Wikipedia
    CC BY-SA 3.0.
  - "statistical literacy."
    http://en.wikipedia.org/wiki/statistical%20literacy.
    Wikipedia
    CC BY-SA 3.0.
  - "Critical thinking."
    http://en.wikipedia.org/wiki/Critical_thinking.

Wikipedia
CC BY-SA 3.0.
    -
    http://thecollaboratory.wdfiles.com/local--files/philosophy-of-thought-and-logic-2011-2012/critical_thinking_skills.jpg.

    http://thecollaboratory.wdfiles.com/local--files/philosophy-of-thought-and-logic-2011-2012/critical_thinking_skills.jpg.
    CC BY-SA.
- Experimental Design
    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "Design of experiments."
      https://en.wikipedia.org/wiki/Design_of_experiments.
      Wikipedia
      CC BY-SA 3.0.
    - "Controlled experiment."
      http://en.wikipedia.org/wiki/Controlled_experiment%23Controlled_experiments.
      Wikipedia
      CC BY-SA 3.0.
    - "Observational study."
      http://en.wikipedia.org/wiki/Observational_study.
      Wikipedia
      CC BY-SA 3.0.
    - "dependent variable."
      http://en.wiktionary.org/wiki/dependent_variable.
      Wiktionary
      CC BY-SA 3.0.
    - "independent variable."
      http://en.wiktionary.org/wiki/independent_variable.
      Wiktionary
      CC BY-SA 3.0.
    - "experiment."
      http://en.wiktionary.org/wiki/experiment.
      Wiktionary

CC BY-SA 3.0.
    - "Balance à tabac 1850."
      https://en.wikipedia.org/wiki/File:Balance_%C3%A0_tabac_1850.JPG.
      Wikipedia
      CC BY-SA.
- Random Samples
    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "Simple random sampling."
      http://en.wikipedia.org/wiki/Simple_random_sampling.
      Wikipedia
      CC BY-SA 3.0.
    - "population."
      http://en.wiktionary.org/wiki/population.
      Wiktionary
      CC BY-SA 3.0.
    - "random sample."
      http://en.wiktionary.org/wiki/random_sample.
      Wiktionary
      CC BY-SA 3.0.
    - "Random sampling."
      http://en.wikipedia.org/wiki/Random_sampling%23Random_sampling.
      Wikipedia
      CC BY-SA 3.0.
    - "MIME types of a random sample of supplementary materials from the Open Access subset in PubMed Central as of October 23, 2012."
      http://commons.wikimedia.org/wiki/File:MIME_types_of_a_random_sample_of_supplementary_materials_from_the_Open_Access_subset_in_PubMed_Central_as_of_October_23,_2012.png.
      Wikimedia
      Public domain.

SECTION II

2. STATISTICS IN PRACTICE

2.1 OBSERVATIONAL STUDIES

2.1: Observational Studies

2.1.1: What are Observational Studies?

An observational study is one in which no variables can be manipulated or controlled by the investigator.

Learning Objectives

Identify situations in which observational studies are necessary and the challenges that arise in their interpretation.

Key Takeaways

Key Points

- An observational study is in contrast with experiments, such as randomized controlled trials, where each subject is randomly assigned to a treated group or a control group.
- Variables may be uncontrollable because 1) a randomized experiment would violate ethical standards, 2) the investigator may simply lack the requisite influence, or 3) a randomized experiment may be impractical.
- Observational studies can never identify causal relationships because even though two

variables are related both might be caused by a third, unseen, variable.

- A major challenge in conducting observational studies is to draw inferences that are acceptably free from influences by overt biases, as well as to assess the influence of potential hidden biases.
- A major challenge in conducting observational studies is to draw inferences that are acceptably free from influences by overt biases, as well as to assess the influence of potential hidden biases.

Key Terms

causality
: the relationship between an event (the cause) and a second event (the effect), where the second event is understood as a consequence of the first

observational study
: a study drawing inferences about the possible effect of a treatment on subjects, where the assignment of subjects into a treated group versus a control group is outside the control of the investigator

A common goal in statistical research is to investigate causality, which is the relationship between an event (the cause) and a second event (the effect), where the second event is understood as a consequence of the first. There are two major types of causal statistical studies: experimental studies and observational studies. An observational study draws inferences about the possible effect of a treatment on subjects, where the assignment of subjects into a treated group versus a control group is outside the control of the investigator. This is in contrast with experiments, such as randomized controlled trials, where each subject is randomly assigned to a treated group or a control group. In other words, observational studies have no independent variables — nothing is manipulated by the experimenter. Rather, observations have the equivalent of two dependent variables.

In an observational study, the assignment of treatments may be beyond the control of the investigator for a variety of reasons:

1. *A randomized experiment would violate ethical standards:* Suppose one wanted to investigate the abortion – breast cancer hypothesis, which postulates a causal link between induced abortion and the incidence of breast cancer. In a hypothetical controlled experiment, one would start with a large subject pool of pregnant women and divide them randomly into a treatment group (receiving induced abortions) and a control group (bearing children), and then conduct regular cancer screenings for

women from both groups. Needless to say, such an experiment would run counter to common ethical principles. The published studies investigating the abortion–breast cancer hypothesis generally start with a group of women who already have received abortions. Membership in this "treated" group is not controlled by the investigator: the group is formed after the "treatment" has been assigned.

2. *The investigator may simply lack the requisite influence:* Suppose a scientist wants to study the public health effects of a community-wide ban on smoking in public indoor areas. In a controlled experiment, the investigator would randomly pick a set of communities to be in the treatment group. However, it is typically up to each community and/or its legislature to enact a smoking ban. The investigator can be expected to lack the political power to cause precisely those communities in the randomly selected treatment group to pass a smoking ban. In an observational study, the investigator would typically start with a treatment group consisting of those communities where a smoking ban is already in effect.
3. *A randomized experiment may be impractical:* Suppose a researcher wants to study the suspected link between a certain medication and a very rare group of symptoms arising as a side effect. Setting aside any ethical considerations, a randomized experiment would be impractical because of the rarity of the effect. There may not be a subject pool large enough for the symptoms to be observed in at least one treated subject. An observational study would typically start with a group of symptomatic subjects and work backwards to find those who were given the medication and later developed the symptoms

Usefulness and Reliability of Observational Studies

Observational studies can never identify causal relationships because even though two variables are related both might be caused by a third, unseen, variable. Since the underlying laws of nature are assumed to be causal laws, observational findings are generally regarded as less compelling than experimental findings.

Observational studies can, however:

1. Provide information on "real world" use and practice
2. Detect signals about the benefits and risks of the use of practices in the general population
3. Help formulate hypotheses to be tested in subsequent experiments
4. Provide part of the community-level data needed to design more informative pragmatic clinical trials
5. Inform clinical practice

A major challenge in conducting observational studies is to draw inferences that are acceptably free from influences by overt biases, as well as to assess the influence of potential hidden biases.

Observational Studies

Nature Observation and Study Hall in The Natural and Cultural Gardens, The Expo Memorial Park, Suita City, Osaka, Japan. Observational studies are a type of experiments in which the variables are outside the control of the investigator.

2.1.2: The Clofibrate Trial

The Clofibrate Trial was a placebo-controlled study to determine the safety and effectiveness of drugs treating coronary heart disease in men.

Learning Objective

Outline how the use of placebos in controlled experiments leads to more reliable results.

Key Takeaways

Key Points

- Clofibrate was one of four lipid-modifying drugs tested in an observational study known as the Coronary Drug Project.
- Placebo-controlled studies are a way of testing a medical therapy in which, in addition to a group of subjects that receives the treatment to be evaluated, a separate control group receives a sham "placebo" treatment which is specifically designed to have no real effect.
- The purpose of the placebo group is to account for the placebo effect — that is, effects from treatment that do not depend on the treatment itself.
- Appropriate use of a placebo in a clinical trial often requires, or at least benefits from, a double-blind study design, which means that neither the experimenters nor the subjects know which subjects are in the "test group" and which are in the "control group. ".
- The use of placebos is a standard control component of most clinical trials which attempt to make some sort of quantitative assessment of the efficacy of medicinal drugs or treatments.

Key Terms

regression to the mean
: the phenomenon by which extreme examples from any set of data are likely to be followed by examples which are less extreme; a tendency towards the average of any sample

placebo
: an inactive substance or preparation used as a control in an experiment or test to determine the effectiveness of a medicinal drug

placebo effect
: the tendency of any medication or treatment, even an inert or ineffective one, to exhibit results simply because the recipient believes that it will work

Clofibrate (tradename Atromid-S) is an organic compound that is marketed as a fibrate. It is a lipid-lowering agent used for controlling the high cholesterol and triacylglyceride level in the blood. Clofibrate was one of four lipid-modifying drugs tested in an observational study known as the Coronary Drug Project. Also known as the World Health Organization Cooperative Trial on Primary Prevention of Ischaemic Heart Disease, the study was a randomized, multi-center, double-blind, placebo-controlled trial that was intended to study the safety and effectiveness of drugs for long-term treatment of coronary heart disease in men.

Placebo-Controlled Observational Studies

Placebo-controlled studies are a way of testing a medical therapy in which, in addition to a group of subjects that receives the treatment to be evaluated, a separate control group receives a sham "placebo" treatment which is specifically designed to have no real effect. Placebos are most commonly used in blinded trials, where subjects do not know whether they are receiving real or placebo treatment.

The purpose of the placebo group is to account for the placebo effect — that is, effects from treatment that do not depend on the treatment itself. Such factors include knowing one is receiving a treatment, attention from health care professionals, and the expectations of a treatment's effectiveness by those running the research study. Without a placebo group to compare against, it is not possible to know whether the treatment itself had any effect.

Appropriate use of a placebo in a clinical trial often requires, or at least benefits from, a double-blind study design, which means that neither the experimenters nor the subjects know which subjects are in the "test group" and which are in the "control group. " This creates a problem in creating placebos that can be mistaken for active treatments. Therefore, it can be necessary to use a psychoactive placebo, a drug that produces physiological effects that encourage the belief in the control groups that they have received an active drug.

Patients frequently show improvement even when given a sham or "fake" treatment. Such intentionally inert placebo treatments can take many forms, such as a pill containing only sugar, a surgery where nothing is actually done, or a medical device (such as ultrasound) that is not actually turned on. Also, due to the body's natural healing ability and statistical effects such as regression to the mean, many patients will get better even when given no treatment at all. Thus, the relevant question when assessing a treatment is not "does the treatment work? " but "does the treatment work better than a placebo treatment, or no treatment at all? "

Therefore, the use of placebos is a standard control component of most clinical trials which attempt to make some sort of quantitative assessment of the efficacy of medicinal drugs or treatments.

Results of The Coronary Drug Project

Those in the placebo group who adhered to the placebo treatment (took the placebo regularly as instructed) showed nearly half the mortality rate as those who were not adherent. A similar study of women found survival was nearly 2.5 times greater for those who adhered to their placebo. This apparent placebo effect may have occurred because:

1. Adhering to the protocol had a psychological effect, i.e. genuine placebo effect.
2. People who were already healthier were more able or more inclined to follow the protocol.
3. Compliant people were more diligent and health-conscious in all aspects of their lives.

The Coronary Drug Project found that subjects using clofibrate to lower serum cholesterol observed excess mortality in the clofibrate-treated group despite successful cholesterol lowering (47% more deaths during treatment with clofibrate and 5% after treatment with clofibrate) than the non-treated high cholesterol group. These deaths were due to a wide variety of causes other than heart disease, and remain "unexplained".

Clofibrate was discontinued in 2002 due to adverse affects.

Placebo-Controlled Observational Studies

Prescription placebos used in research and practice.

2.1.3: Confounding

A confounding variable is an extraneous variable in a statistical model that correlates with both the dependent variable and the independent variable.

Learning Objective

Break down why confounding variables may lead to bias and spurious relationships and what can be done to avoid these phenomenons.

Key Takeaways

Key Points

- A perceived relationship between an independent variable and a dependent variable that has been misestimated due to the failure to account for a confounding factor is termed a spurious relationship.
- Confounding by indication – the most important limitation of observational studies – occurs when prognostic factors cause bias, such as biased estimates of treatment effects in medical trials.
- Confounding variables may also be categorised according to their source: such as operational confounds, procedural confounds or person confounds.
- A reduction in the potential for the occurrence and effect of confounding factors can be obtained by increasing the types and numbers of comparisons performed in an analysis.
- Moreover, depending on the type of study design in place, there are various ways to modify that design to actively exclude or control confounding variables.

Key Terms

peer review
: the scholarly process whereby manuscripts intended to be published in an academic journal are reviewed by independent researchers (referees) to evaluate the contribution, i.e. the

importance, novelty and accuracy of the manuscript's contents

placebo effect
: the tendency of any medication or treatment, even an inert or ineffective one, to exhibit results simply because the recipient believes that it will work

prognostic
: a sign by which a future event may be known or foretold

confounding variable
: an extraneous variable in a statistical model that correlates (positively or negatively) with both the dependent variable and the independent variable

Example

In risk assessments, factors such as age, gender, and educational levels often have impact on health status and so should be controlled. Beyond these factors, researchers may not consider or have access to data on other causal factors. An example is on the study of smoking tobacco on human health. Smoking, drinking alcohol, and diet are lifestyle activities that are related. A risk assessment that looks at the effects of smoking but does not control for alcohol consumption or diet may overestimate the risk of smoking. Smoking and confounding are reviewed in occupational risk assessments such as the safety of coal mining. When there is not a large sample population of non-smokers or non-drinkers in a particular occupation, the risk assessment may be biased towards finding a negative effect on health.

Confounding Variables

A confounding variable is an extraneous variable in a statistical model that correlates (positively or negatively) with both the dependent variable and the independent variable. A perceived relationship between an independent variable and a dependent variable that has been misestimated due to the failure to account for a confounding factor is termed a *spurious relationship*, and the presence of misestimation for this reason is termed *omitted-variable bias*.

As an example, suppose that there is a statistical relationship between ice cream consumption and number of drowning deaths for a given period. These two variables have a positive correlation with each other. An individual might attempt to explain this correlation by inferring a causal relationship between the two variables (either that ice cream causes drowning, or that drowning causes ice cream consumption). However, a more likely explanation is that the relationship between ice cream consumption and drowning is spurious and that a third, confounding, variable (the season) influences both variables: during the summer, warmer temperatures lead to increased ice cream consumption as well as more people swimming and, thus, more drowning deaths.

Types of Confounding

Confounding by indication has been described as the most important limitation of observational studies. Confounding by indication occurs when prognostic factors cause bias, such as biased estimates of treatment effects in medical trials. Controlling for known prognostic factors may reduce this problem, but it is always possible that a forgotten or unknown factor was not included or that factors interact complexly. Randomized trials tend to reduce the effects of confounding by indication due to random assignment.

Confounding variables may also be categorised according to their source:

- The choice of measurement instrument (operational confound) – This type of confound occurs when a measure designed to assess a particular construct inadvertently measures something else as well.
- Situational characteristics (procedural confound) – This type of confound occurs when the researcher mistakenly allows another variable to change along with the manipulated independent variable.
- Inter-individual differences (person confound) – This type of confound occurs when two or more groups of units are analyzed together (e.g., workers from different occupations) despite varying according to one or more other (observed or unobserved) characteristics (e.g., gender).

Decreasing the Potential for Confounding

A reduction in the potential for the occurrence and effect of confounding factors can be obtained by increasing the types and numbers of comparisons performed in an analysis. If a relationship holds among different subgroups of analyzed units, confounding may be less likely. That said, if measures or manipulations of core constructs are confounded (i.e., operational or procedural confounds exist), subgroup analysis may not reveal problems in the analysis.

Peer review is a process that can assist in reducing instances of confounding, either before study implementation or after analysis has occurred. Similarly, study replication can test for the robustness of findings from one study under alternative testing conditions or alternative analyses (e.g., controlling for potential confounds not identified in the initial study). Also, confounding effects may be less likely to occur and act similarly at multiple times and locations.

Moreover, depending on the type of study design in place, there are various ways to modify that design to actively exclude or control confounding variables:

1. Case-control studies assign confounders to both groups, cases and controls, equally. In case-control studies, matched variables most often are age and sex.
2. In cohort studies, a degree of matching is also possible, and it is often done by only admitting certain age groups or a certain sex into the study population. this creates a cohort of people who share similar characteristics; thus, all cohorts are comparable in regard to the possible confounding variable.
3. Double blinding conceals the experiment group membership of the participants from the trial population and the observers. By preventing the participants from knowing if they are receiving treatment or not, the placebo effect should be the same for the control and treatment groups. By preventing the observers from knowing of their membership, there should be no bias from researchers treating the groups differently or from interpreting the outcomes differently.
4. A randomized controlled trial is a method where the study population is divided randomly in order to mitigate the chances of self-selection by participants or bias by the study designers. Before the experiment begins, the testers will assign the members of the participant pool to their groups (control, intervention, parallel) using a randomization process such as the use of a random number generator.

2.1.4: Sex Bias in Graduate Admissions

The Berkeley study is one of the best known real life examples of an experiment suffering from a confounding variable.

Learning Objective

Illustrate how the phenomenon of confounding can be seen in practice via Simpson's Paradox.

Key Takeaways

Key Points

- A study conducted in the aftermath of a law suit filed against the University of California, Berkeley showed that men applying were more likely than women to be admitted.
- Examination of the aggregate data on admissions showed a blatant, if easily misunderstood, pattern of gender discrimination against applicants.
- When examining the individual departments, it appeared that no department was significantly biased against women.
- The study concluded that women tended to apply to competitive departments with low rates of admission even among qualified applicants, whereas men tended to apply to less-competitive departments with high rates of admission among the qualified applicants.
- Simpson's Paradox is a paradox in which a trend that appears in different groups of data disappears when these groups are combined, and the reverse trend appears for the aggregate data.

Key Terms

partition
: a part of something that had been divided, each of its results

Simpson's paradox
: a paradox in which a trend that appears in different groups of data disappears when these groups are combined, and the reverse trend appears for the aggregate data

aggregate
: a mass, assemblage, or sum of particulars; something consisting of elements but considered as a whole

Women have traditionally had limited access to higher education. Moreover, when women began to be admitted to higher education, they were encouraged to major in less-intellectual subjects. For example, the study of English literature in American and British colleges and universities was instituted as a field considered suitable to women's "lesser intellects".

However, since 1991 the proportion of women enrolled in college in the U.S. has exceeded the enrollment rate for men, and that gap has widened over time. As of 2007, women made up the majority — 54 percent — of the 10.8 million college students enrolled in the U.S.

This has not negated the fact that gender bias exists in higher education. Women tend to score lower on graduate admissions exams, such as the Graduate Record Exam (GRE) and the Graduate Management Admissions Test (GMAT). Representatives of the companies that publish these tests have hypothesized that greater number of female applicants taking these tests pull down women's average scores. However, statistical research proves this theory wrong. Controlling for the number of people taking the test does not account for the scoring gap.

Sex Bias at the University of California, Berkeley

On February 7, 1975, a study was published in the journal *Science* by P.J. Bickel, E.A. Hammel, and J.W. O'Connell entitled "Sex Bias in Graduate Admissions: Data from Berkeley. " This study was conducted in the aftermath of a law suit filed against the University, citing admission figures for the fall of 1973, which showed that men applying were more likely than women to be admitted, and the difference was so large that it was unlikely to be due to chance.

Examination of the aggregate data on admissions showed a blatant, if easily misunderstood, pattern of gender discrimination against applicants.

Aggregate Data

	All		**Men**		**Women**	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
Total	12,763	41%	8,442	44%	4,321	35%

When examining the individual departments, it appeared that no department was significantly biased against women. In fact, most departments had a small but statistically significant bias in favor of women. The data from the six largest departments are listed below.

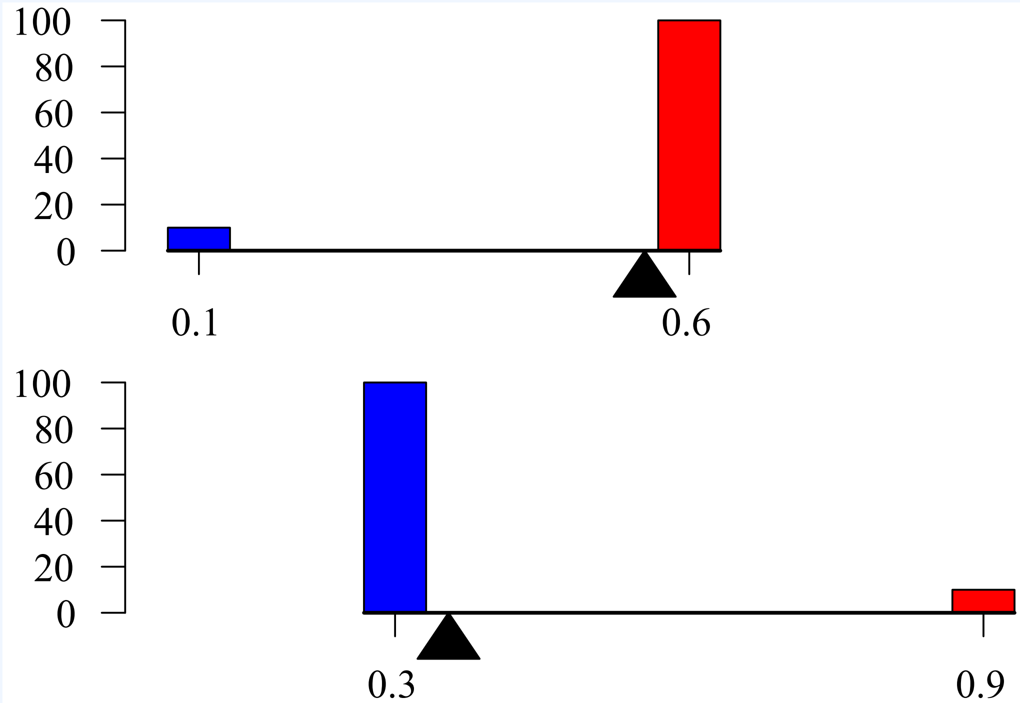
Sex Bias at UC Berkeley by Department

Department	Men (# Applicants)	Men (% Admitted)	Women (# Applicants)	Women (% Admitted)
A	825	62	108	82
B	560	63	25	68
C	325	37	593	34
D	417	33	375	35
E	191	28	393	24
F	272	6	341	7

The research paper by Bickel et al. concluded that women tended to apply to competitive departments with low rates of admission even among qualified applicants (such as in the English Department), whereas men tended to apply to less-competitive departments with high rates of admission among the qualified applicants (such as in engineering and chemistry). The study also concluded that the graduate departments that were easier to enter at the University, at the time, tended to be those that required more undergraduate preparation in mathematics. Therefore, the admission bias seemed to stem from courses previously taken.

Confounding Variables and Simpson's Paradox

The above study is one of the best known real life examples of an experiment suffering from a confounding variable. In this particular case, we can see an occurrence of Simpson's Paradox . Simpson's Paradox is a paradox in which a trend that appears in different groups of data disappears when these groups are combined, and the reverse trend appears for the aggregate data. This result is often encountered in social-science and medical-science statistics, and is particularly confounding when frequency data are unduly given causal interpretations.



Simpson's Paradox: For a full explanation of the figure, visit: [Simpson's Paradox on Wikipedia]

The practical significance of Simpson's paradox surfaces in decision making situations where it poses the following dilemma: *Which data should we consult in choosing an action, the aggregated or the partitioned?* The answer seems to be that one should sometimes follow the partitioned and sometimes the aggregated data, depending on the story behind the data; with each story dictating its own choice.

As to why and how a story, not data, should dictate choices, the answer is that it is the story which encodes the causal relationships among the variables. Once we extract these relationships we can test algorithmically whether a given partition, representing confounding variables, gives the correct answer.

Confounding Variables in Practice

One of the best real life examples of the presence of confounding variables occurred in a study regarding sex bias in graduate admissions here, at the University of California, Berkeley.

Attributions

- What are Observational Studies?
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "causality."

http://en.wikipedia.org/wiki/causality.
Wikipedia
CC BY-SA 3.0.
- “Statistics.”
http://en.wikipedia.org/wiki/Statistics%23Experimental_and_observational_studies.
Wikipedia
CC BY-SA 3.0.
- “Causality.”
http://en.wikipedia.org/wiki/Causality.
Wikipedia
CC BY-SA 3.0.
- “observational study.”
http://en.wikipedia.org/wiki/observational%20study.
Wikipedia
CC BY-SA 3.0.
- “Observational study.”
http://en.wikipedia.org/wiki/Observational_study.
Wikipedia
CC BY-SA 3.0.
- “Statistics/Methods of Data Collection/Observational Studies.”
http://en.wikibooks.org/wiki/Statistics/Methods_of_Data_Collection/Observational_Studies.
Wikibooks
CC BY-SA 3.0.
- “Nature Observation and Study Hall.”
http://commons.wikimedia.org/wiki/File:Nature_Observation_and_Study_Hall.JPG.
Wikimedia
CC BY-SA.

- The Clofibrate Trial
  - “Boundless.”
  http://www.boundless.com/.
  Boundless Learning
  CC BY-SA 3.0.
  - “Coronary Drug Project – Full Text View – ClinicalTrials.gov.”
  http://clinicaltrials.gov/ct2/show/NCT00000482.
  Clinical Trials
  Public domain.
  - “Placebo-controlled study.”

http://en.wikipedia.org/wiki/Placebo-controlled_study.
Wikipedia
CC BY-SA 3.0.

    ◦ "Clofibrate."
      http://en.wikipedia.org/wiki/Clofibrate.
      Wikipedia
      CC BY-SA 3.0.
    ◦ "Placebo-controlled study."
      http://en.wikipedia.org/wiki/Placebo-controlled_study.
      Wikipedia
      CC BY-SA 3.0.
    ◦ "placebo."
      http://en.wikipedia.org/wiki/placebo.
      Wikipedia
      CC BY-SA 3.0.
    ◦ "placebo effect."
      http://en.wiktionary.org/wiki/placebo_effect.
      Wiktionary
      CC BY-SA 3.0.
    ◦ "regression to the mean."
      http://en.wiktionary.org/wiki/regression_to_the_mean.
      Wiktionary
      CC BY-SA 3.0.
    ◦ "Cebocap."
      http://en.wikipedia.org/wiki/File:Cebocap.jpg.
      Wikipedia
      Public domain.
- Confounding
    ◦ "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    ◦ "Confounding."
      http://en.wikipedia.org/wiki/Confounding.
      Wikipedia
      CC BY-SA 3.0.
    ◦ "prognostic."

http://en.wiktionary.org/wiki/prognostic.
Wiktionary
CC BY-SA 3.0.

- "placebo effect."
  http://en.wiktionary.org/wiki/placebo_effect.
  Wiktionary
  CC BY-SA 3.0.
- "confounding variable."
  http://en.wiktionary.org/wiki/confounding_variable.
  Wiktionary
  CC BY-SA 3.0.
- "peer review."
  http://en.wiktionary.org/wiki/peer_review.
  Wiktionary
  CC BY-SA 3.0.

- Sex Bias in Graduate Admissions
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Sexism."
    http://en.wikipedia.org/wiki/Sexism%23Education.
    Wikipedia
    CC BY-SA 3.0.
  - "Simpson's paradox."
    http://en.wikipedia.org/wiki/Simpson's_paradox.
    Wikipedia
    CC BY-SA 3.0.
  - "aggregate."
    http://en.wiktionary.org/wiki/aggregate.
    Wiktionary
    CC BY-SA 3.0.
  - "Simpson's paradox."
    http://en.wiktionary.org/wiki/Simpson's_paradox.
    Wiktionary
    CC BY-SA 3.0.
  - "partition."

http://en.wiktionary.org/wiki/partition.
Wiktionary
CC BY-SA 3.0.

- “Simpson’s paradox.”
  http://commons.wikimedia.org/wiki/File:Simpson’s_paradox.svg.
  Wikimedia
  CC BY-SA.
- “All sizes | Berkeley – Sather Gate and Campanile | Flickr – Photo Sharing!.”
  http://www.flickr.com/photos/24736216@N07/4072064529/sizes/m/in/photostream/.
  Flickr
  CC BY-SA.

2.2 CONTROLLED EXPERIMENTS

2.2: Controlled Experiments

2.2.1: The Salk Vaccine Field Trial

The Salk polio vaccine field trial incorporated a double blind placebo control methodology to determine the effectiveness of the vaccine.

Learning Objective

Demonstrate how controls and treatment groups are used in drug testing.

Key Takeaways

Key Points

- The first effective polio vaccine was developed in 1952 by Jonas Salk at the University of Pittsburgh.
- Roughly 440,000 people received one or more injections of the vaccine, about 210,000 children received a placebo, consisting of harmless culture media, and 1.2 million children received no vaccination and served as a control group, who would then be observed to see if

any contracted polio.
- Two serious issues arose in the original experimental design: selection bias and diagnostic bias.
- The combination of randomized control and double-blind experimental factors, which were implemented in the second version of the experimental design, has become the gold standard for a clinical trial.

Key Terms

control group
: the group of test subjects left untreated or unexposed to some procedure and then compared with treated subjects in order to validate the results of the test

placebo
: an inactive substance or preparation used as a control in an experiment or test to determine the effectiveness of a medicinal drug

The Salk polio vaccine field trials constitute one of the most famous and one of the largest statistical studies ever conducted. The field trials are of particular value to students of statistics because two different experimental designs were used.

Background

The Salk vaccine, or inactivated poliovirus vaccine (IPV), is based on three wild, virulent reference strains:

1. Mahoney (type 1 poliovirus),
2. MEF-1 (type 2 poliovirus), and
3. Saukett (type 3 poliovirus),

grown in a type of monkey kidney tissue culture (Vero cell line), which are then inactivated with formalin. The injected Salk vaccine confers IgG-mediated immunity in the bloodstream, which prevents polio infection from progressing to viremia and protects the motor neurons, thus eliminating the risk of bulbar polio and post-polio syndrome.

The 1954 Field Trial

Statistical tests of new medical treatments almost always have the same basic format. The responses of a treatment group of subjects who are given the treatment are compared to the responses of a control group of subjects who are not given the treatment. The treatment groups and control groups should be as similar as possible.

Beginning February 23, 1954, the vaccine was tested at Arsenal Elementary School and the Watson Home for Children in Pittsburgh, Pennsylvania. Salk's vaccine was then used in a test called the Francis Field Trial, led by Thomas Francis; the largest medical experiment in history. The test began with some 4,000 children at Franklin Sherman Elementary School in McLean, Virginia, and would eventually involve 1.8 million children, in 44 states from Maine to California. By the conclusion of the study, roughly 440,000 received one or more injections of the vaccine, about 210,000 children received a placebo, consisting of harmless culture media, and 1.2 million children received no vaccination and served as a control group, who would then be observed to see if any contracted polio.

The results of the field trial were announced April 12, 1955 (the 10^{th} anniversary of the death of President Franklin D. Roosevelt, whose paralysis was generally believed to have been caused by polio). The Salk vaccine had been 60–70% effective against PV1 (poliovirus type 1), over 90% effective against PV2 and PV3, and 94% effective against the development of bulbar polio. Soon after Salk's vaccine was licensed in 1955, children's vaccination campaigns were launched. In the U.S, following a mass immunization campaign promoted by the March of Dimes, the annual number of polio cases fell from 35,000 in 1953 to 5,600 by 1957. By 1961 only 161 cases were recorded in the United States.

Experimental Design Issues

The original design of the experiment called for second graders (with parental consent) to form the treatment group and first and third graders to form the control group. This design was known as the observed control experiment.

Two serious issues arose in this design: selection bias and diagnostic bias. Because only second graders with permission from their parents were administered the treatment, this treatment group became self-selecting.

Thus, a randomized control design was implemented to overcome these apparent deficiencies. The key distinguishing feature of the randomized control design is that study subjects, after assessment of eligibility and recruitment, but before the intervention to be studied begins, are randomly allocated to receive one or the other of the alternative treatments under study. Therefore, randomized control

The Salk Polio Vaccine Field Trial

Jonas Salk administers his polio vaccine on February 26, 1957 in the Commons Room of the Cathedral of Learning at the University of Pittsburgh where the vaccine was created by Salk and his team.

tends to negate all effects (such as confounding variables) except for the treatment effect.

This design also had the characteristic of being double-blind. Double-blind describes an especially stringent way of conducting an experiment on human test subjects which attempts to eliminate subjective, unrecognized biases carried by an experiment's subjects and conductors. In a double-blind experiment, neither the participants nor the researchers know which participants belong to the control group, as opposed to the test group. Only after all data have been recorded (and in some cases, analyzed) do the researchers learn which participants were which.

This combination of randomized control and double-blind experimental factors has become the gold standard for a clinical trial.

2.2.2: The Portacaval Shunt

Numerous studies have been conducted to examine the value of the portacaval shunt procedure, many using randomized controls.

Learning Objective

Assess the value that the practice of random assignment adds to experimental design.

Key Takeaways

Key Points

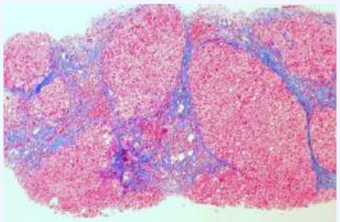
- A portacaval shunt is a treatment for the liver in which a connection is made between the portal vein, which supplies 75% of the liver's blood, and the inferior vena cava, the vein that drains blood from the lower two-thirds of the body.
- Of the studies on portacaval shunts, 63% were conducted without controls, 29% were conducted with non-randomized controls, and 8% were conducted with randomized controls.
- The thinking behind random assignment is that any effect observed between treatment groups can be linked to the treatment effect and cannot be considered a characteristic of the individuals in the group.
- Because most basic statistical tests require the hypothesis of an independent randomly sampled population, random assignment is the desired assignment method.

Key Terms

shunt
: a passage between body channels constructed surgically as a bypass

random assignment
: an experimental technique for assigning subjects to different treatments (or no treatment)

A portacaval shunt is a treatment for high blood pressure in the liver. A connection is made between the portal vein, which supplies 75% of the liver's blood, and the inferior vena cava, the vein that drains blood from the lower two-thirds of the body. The most common causes of liver disease resulting in portal hypertension are cirrhosis , caused by alcohol abuse, and viral hepatitis (hepatitis B and C). Less common causes include diseases such as hemochromatosis, primary biliary cirrhosis (PBC), and portal vein thrombosis. The procedure is long and hazardous .



The Portacaval Shunt

This image is a trichrome stain showing cirrhosis of the liver. Cirrhosis can be combatted by the portacaval shunt procedure, for which there have been numerous experimental trials using randomized assignment.

Numerous studies have been conducted to examine the value of and potential concerns with the surgery. Of these studies, 63% were conducted without controls, 29% were conducted with non-randomized controls, and 8% were conducted with randomized controls.

Randomized Controlled Experiments

Random assignment, or random placement, is an experimental technique for assigning subjects to different treatments (or no treatment). The thinking behind random assignment is that by randomizing treatment assignments, the group attributes for the different treatments will be roughly equivalent; therefore, any effect observed between treatment groups can be linked to the treatment effect and cannot be considered a characteristic of the individuals in the group.

In experimental design, random assignment of participants in experiments or treatment and control groups help to ensure that any differences between and within the groups are not systematic at the outset of the experiment. Random assignment does not guarantee that the groups are "matched" or equivalent, only that any differences are due to chance.

The steps to random assignment include:

1. Begin with a collection of subjects – for example, 20 people.
2. Devise a method of randomization that is purely mechanical (e.g. flip a coin).
3. Assign subjects with "heads" to one group, the control group; assign subjects with "tails" to the other group, the experimental group.

Because most basic statistical tests require the hypothesis of an independent randomly sampled population, random assignment is the desired assignment method. It provides control for all attributes of the members of the samples—in contrast to matching on only one or more variables—and provides the mathematical basis for estimating the likelihood of group equivalence for characteristics one is interested in. This applies both for pre-treatment checks on equivalence and the evaluation of post treatment results using inferential statistics. More advanced statistical modeling can be used to adapt the inference to the sampling method.

2.2.3: Statistical Controls

A scientific control is an observation designed to minimize the effects of variables other than the single independent variable.

Learning Objective

Classify scientific controls and identify how they are used in experiments.

Key Takeaways

Key Points

- Scientific controls increase the reliability of test results, often through a comparison between control measurements and the other measurements.
- Positive and negative controls, when both are successful, are usually sufficient to eliminate most potential confounding variables.
- Negative controls are groups where no phenomenon is expected. They ensure that there is no effect when there should be no effect.
- Positive controls are groups where a phenomenon is expected. That is, they ensure that there is an effect when there should be an effect.

Key Terms

confounding variable
: an extraneous variable in a statistical model that correlates (positively or negatively) with both the dependent variable and the independent variable

scientific control
: an experiment or observation designed to minimize the effects of variables other than the single independent variable

What Is a Control?

A scientific control is an observation designed to minimize the effects of variables other than the single independent variable. This increases the reliability of the results, often through a comparison between control measurements and the other measurements.

For example, during drug testing, scientists will try to control two groups to keep them as identical as possible, then allow one group to try the drug. Another example might be testing plant fertilizer by giving it to only half the plants in a garden: the plants that receive no fertilizer are the control group, because they establish the baseline level of growth that the fertilizer-treated plants will be compared against. Without a control group, the experiment cannot determine whether the fertilizer-treated plants grow more than they would have if untreated.

Ideally, all variables in an experiment will be controlled (accounted for by the control measurements) and none will be uncontrolled. In such an experiment, if all the controls work as expected, it is possible to conclude that the experiment is working as intended and that the results of the experiment are due to the effect of the variable being tested. That is, scientific controls allow an investigator to make a claim like "Two situations were identical until factor X occurred. Since factor X is the only difference between the two situations, the new outcome was caused by factor X. "

Controlled Experiments

Controlled experiments can be performed when it is difficult to exactly control all the conditions in an experiment. In this case, the experiment begins by creating two or more sample groups that are probabilistically equivalent, which means that measurements of traits should be similar among the groups and that the groups should respond in the same manner if given the same treatment. This equivalency is determined by statistical methods that take into account the amount of variation between individuals and the number of individuals in each group. In fields such as microbiology and chemistry, where there is very little variation between individuals and the group size is easily in the millions, these statistical methods are often bypassed and simply splitting a solution into equal parts is assumed to produce identical sample groups.

Types of Controls

The simplest types of control are negative and positive controls. These two controls, when both are successful, are usually sufficient to eliminate most potential confounding variables. This means that the experiment produces a negative result when a negative result is expected and a positive result when a positive result is expected.

Negative Controls

Negative controls are groups where no phenomenon is expected. They ensure that there is no effect when there should be no effect. To continue with the example of drug testing, a negative control is a group that has not been administered the drug. We would say that the control group should show a negative or null effect.

If the treatment group and the negative control both produce a negative result, it can be inferred that the treatment had no effect. If the treatment group and the negative control both produce a positive result, it can be inferred that a confounding variable acted on the experiment, and the positive results are likely not due to the treatment.

Positive Controls

Positive controls are groups where a phenomenon is expected. That is, they ensure that there is an effect when there should be an effect. This is accomplished by using an experimental treatment that is already known to produce that effect and then comparing this to the treatment that is being investigated in the experiment.

Positive controls are often used to assess test validity. For example, to assess a new test's ability to detect a disease, then we can compare it against a different test that is already known to work. The well-established test is the positive control, since we already know that the answer to the question (whether the test works) is yes.

For difficult or complicated experiments, the result from the positive control can also help in comparison to previous experimental results. For example, if the well-established disease test was determined to have the same effectiveness as found by previous experimenters, this indicates that the experiment is being performed in the same way that the previous experimenters did.

When possible, multiple positive controls may be used. For example, if there is more than one disease test that is known to be effective, more than one might be tested. Multiple positive controls also allow finer comparisons of the results (calibration or standardization) if the expected results from the positive controls have different sizes.

Controlled Experiments

An all-female crew of scientific experimenters began a five-day exercise on December 16, 1974. They conducted 11 selected experiments in materials science to determine their practical application for Spacelab missions and to identify integration and operational problems that might occur on actual missions. Air circulation, temperature, humidity and other factors were carefully controlled.

Attributions

- The Salk Vaccine Field Trial
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Jonas Salk."
    http://en.wikipedia.org/wiki/Jonas_Salk.
    Wikipedia
    CC BY-SA 3.0.
  - "Polio vaccine."
    http://en.wikipedia.org/wiki/Polio_vaccine.
    Wikipedia
    CC BY-SA 3.0.
  - "placebo."
    http://en.wikipedia.org/wiki/placebo.
    Wikipedia
    CC BY-SA 3.0.
  - "control group."
    http://en.wiktionary.org/wiki/control_group.
    Wiktionary

CC BY-SA 3.0.

- "PittPolioVaccineCoL." http://commons.wikimedia.org/wiki/File:PittPolioVaccineCoL.jpg. Wikimedia Public domain.

- The Portacaval Shunt
    - "Boundless." http://www.boundless.com/. Boundless Learning CC BY-SA 3.0.
    - "Portacaval shunt." http://en.wikipedia.org/wiki/Portacaval_shunt. Wikipedia CC BY-SA 3.0.
    - "Random assignment." http://en.wikipedia.org/wiki/Random_assignment. Wikipedia CC BY-SA 3.0.
    - "shunt." http://en.wiktionary.org/wiki/shunt. Wiktionary CC BY-SA 3.0.
    - "Cirrhosis of the liver (trichrome stain) | Flickr – Photo Sharing!." http://www.flickr.com/photos/euthman/5690946257/. Flickr CC BY.
- Statistical Controls
    - "Boundless." http://www.boundless.com/. Boundless Learning CC BY-SA 3.0.
    - "Experiment." http://en.wikipedia.org/wiki/Experiment. Wikipedia CC BY-SA 3.0.
    - "scientific control." http://en.wikipedia.org/wiki/scientific%20control.

Wikipedia
CC BY-SA 3.0.

- "Scientific control."
  http://en.wikipedia.org/wiki/Scientific_control.
  Wikipedia
  CC BY-SA 3.0.
- "confounding variable."
  http://en.wiktionary.org/wiki/confounding_variable.
  Wiktionary
  CC BY-SA 3.0.
- "Materials Science Experiments Conducted at MSFC – GPN-2002-000198."
  http://commons.wikimedia.org/wiki/
  File:Materials_Science_Experiments_Conducted_at_MSFC_-_GPN-2002-000198.jpg.
  Wikimedia
  Public domain.

SECTION III

2.XLSX – EXCEL CHALLENGE - FUNDAMENTAL SKILLS

Microsoft® Excel® is a tool that can be used in virtually all careers and is valuable in both professional and personal settings. Whether you need to keep track of medications in inventory for a hospital or create a financial plan for your retirement, Excel enables you to do these activities efficiently and accurately. The following trainings and Excel Challenge assignment introduce the fundamental skills necessary to get you started in using Excel. You will find that just a few skills can make you very productive in a short period of time.

Attribution

Adapted from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

2.XLSX.1 OVERVIEW OF MICROSOFT EXCEL

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Learning Objectives

1. Examine the value of using Excel to make decisions.
2. Learn how to start Excel.
3. Become familiar with the Excel workbook.
4. Understand how to navigate worksheets.
5. Examine the Excel Ribbon.
6. Examine the right-click menu options.
7. Learn how to save workbooks.
8. Examine the Status Bar.
9. Become familiar with the features in the Excel Help window.

Microsoft® Office contains a variety of tools that help people accomplish many personal and professional objectives. Microsoft Excel is perhaps the most versatile and widely used of all the Office applications. No matter which career path you choose, you will likely need to use Excel to accomplish your professional objectives, some of which may occur daily. This chapter provides an overview of the Excel application along with an orientation for accessing the commands and features of an Excel workbook.

Making Decisions with Excel

Taking a very simple view, Excel is a tool that allows you to enter quantitative data into an electronic spreadsheet to apply one or many mathematical computations. These computations ultimately convert that quantitative data into information. The information produced in Excel can be used to make decisions in both professional and personal contexts. For example, employees can use Excel to determine how much inventory to buy for a clothing retailer, how much medication to administer to a patient, or how much money to spend to stay within a budget. With respect to personal decisions, you can use Excel to determine how much money you can spend on a house, how much you can spend on car lease payments, or how much you need to save to reach your retirement goals. We will demonstrate how you can use Excel to make these decisions and many more throughout this text.

Figure 1.1 shows a completed Excel worksheet that will be constructed in this chapter. The information shown in this worksheet contains sales data for a hypothetical merchandise retail company. The worksheet data can help a retailer analyze the business and determine the number of salespeople needed for each month for example.

	A	B	C	D
1	***Merchandise City, USA Retail Sales***			
2	**Month**	**Unit Sales**	**Average Price**	**Sales Dollars**
3	January	2,670	$ 9.99	$ 26,685
4	February	2,160	$ 12.49	$ 26,937
5	March	515	$ 14.99	$ 7,701
6	April	590	$ 17.49	$ 10,269
7	May	1,030	$ 14.99	$ 15,405
8	June	2,875	$ 12.49	$ 35,916
9	July	2,700	$ 9.99	$ 26,937
10	August	900	$ 19.99	$ 17,958
11	September	775	$ 19.99	$ 15,708
12	October	1,180	$ 19.99	$ 23,562
13	November	1,800	$ 17.49	$ 31,416
14	December	4,560	$ 14.99	$ 75,125
15	**Total Sales**	21,755		$ 313,619
16				
17				

Figure 1.1 Example of an Excel Worksheet

Starting Excel

1. Locate Excel on your computer.
2. Click Microsoft Excel to launch the Excel application where you are presented with workbook options to help get you started.
3. Click the first option; "Blank Workbook".

Excel for Windows vs Excel for Mac

The Excel for Windows and Excel for Mac software versions are very similar. Most of the features, tools and commands are available in both versions. There are, however, some differences with the Excel interface. There are also a few features that are not available in the Excel for Mac version. The screenshots and step-by-step instructions in this textbook are specific to Excel for Windows. We have attempted to provide alternate screenshots and instructions for the Mac version when the differences are significant. When you see this icon , it means we are providing information specific to Mac users.

The Excel Workbook

A workbook is an Excel file that contains one or more worksheets (referred to as spreadsheets). Excel will assign a file name to the workbook, such as **Book1**, **Book2**, **Book3**, and so on, depending on how many new workbooks are opened. **Figure 1.2** shows a blank workbook after starting Excel. Take some time to familiarize yourself with this screen. Your screen may be slightly different based on the version you're using.

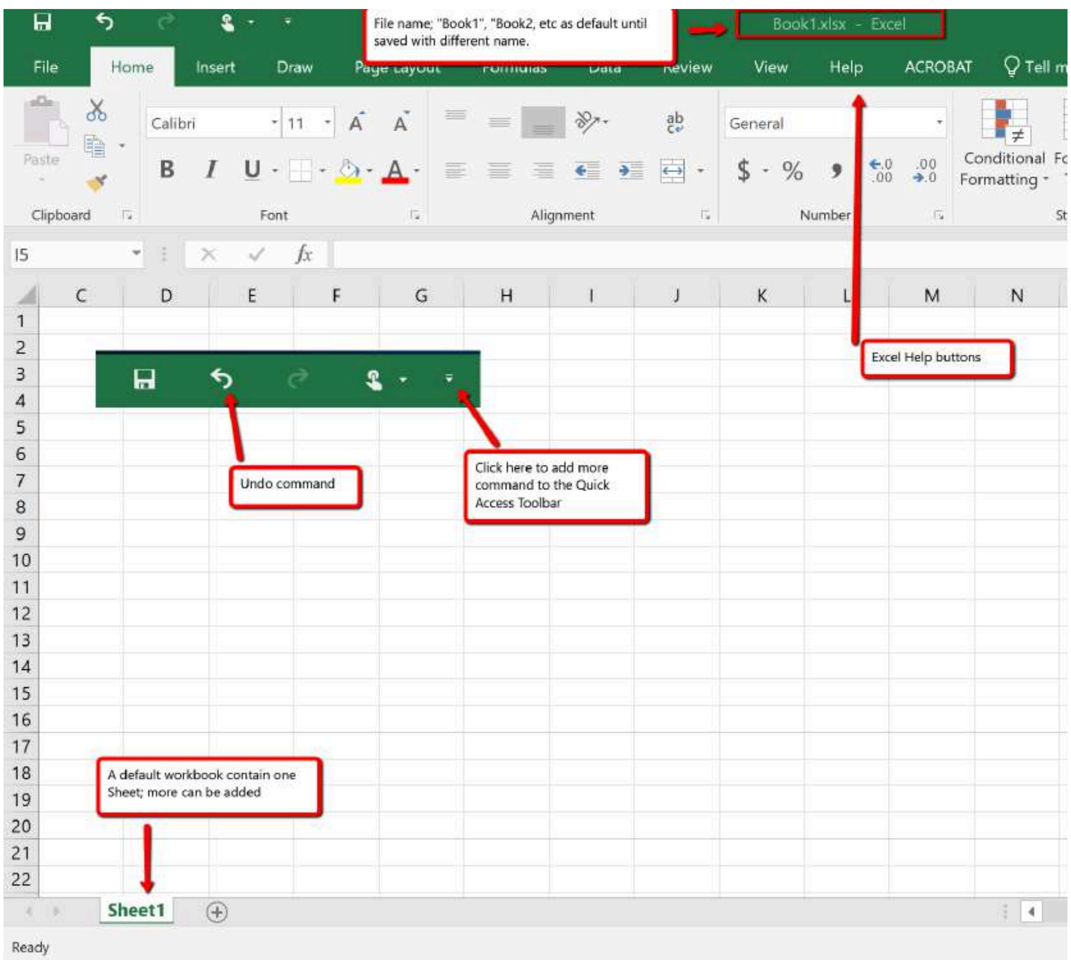


Figure 1.2 Blank Workbook

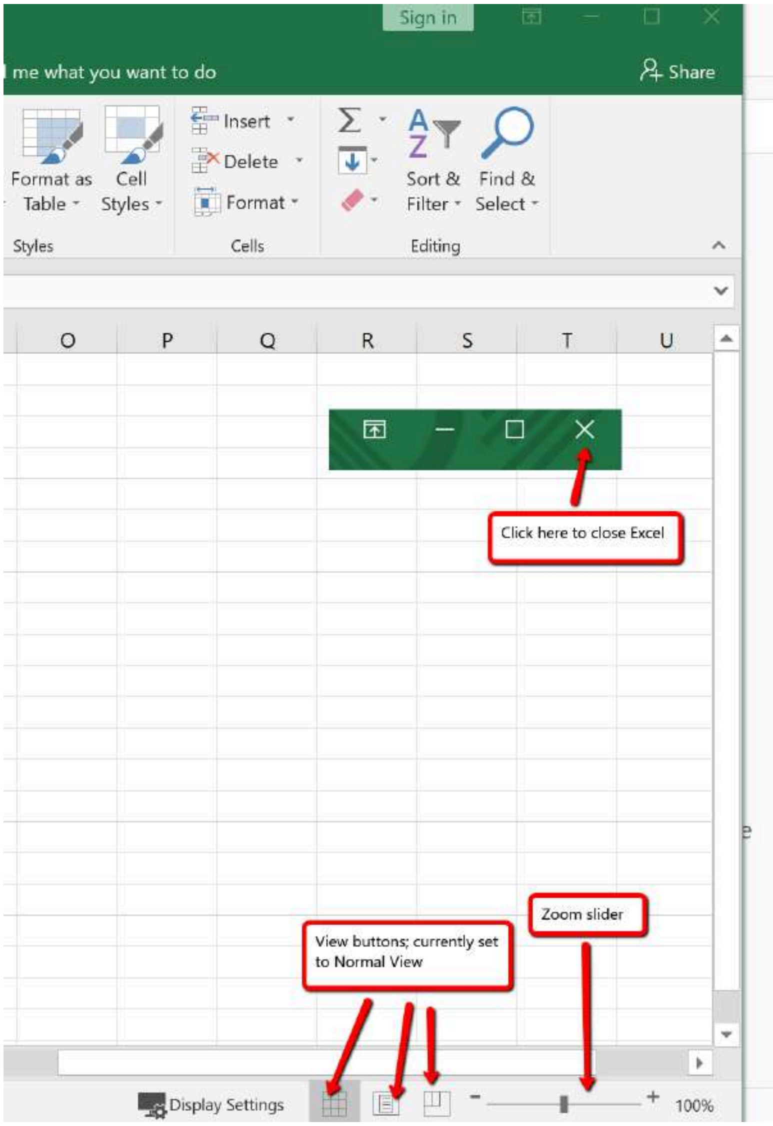


Figure 1.2a Blank Workbook (right-side)

Your workbook should already be maximized (or shown at full size) once Excel is started, as shown in **Figure 1.2**. However, if your screen looks like **Figure 1.3** after starting Excel, you should click the Maximize button, as shown in the figure.

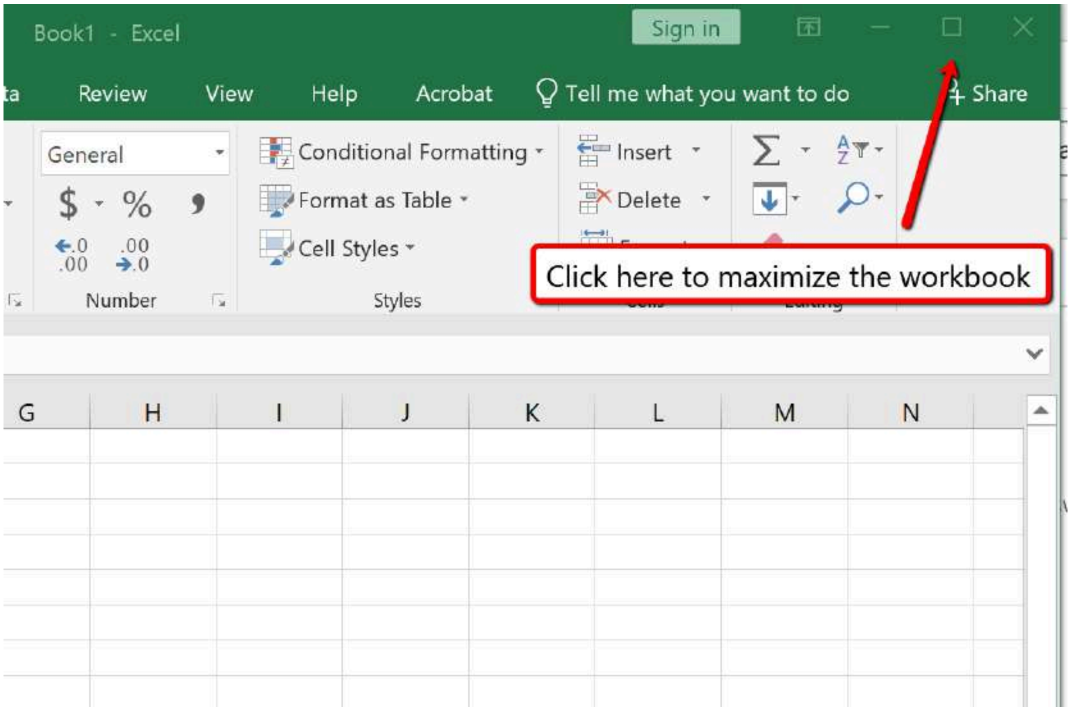


Figure 1.3 Restored Worksheet

Navigating Worksheets

Data are entered and managed in an Excel worksheet. The worksheet contains several rectangles called cells for entering numeric and non-numeric data. Each cell in an Excel worksheet contains an address, which is defined by a column letter followed by a row number. For example, the cell that is currently activated in **Figure 1.3** is **A1**. This would be referred to as cell location A1 or cell reference A1. The following steps explain how you can navigate in an Excel worksheet:

1. Place your mouse pointer over cell D5 and click.
2. Check to make sure column letter D and row number 5 are highlighted, as shown in **Figure 1.4**.

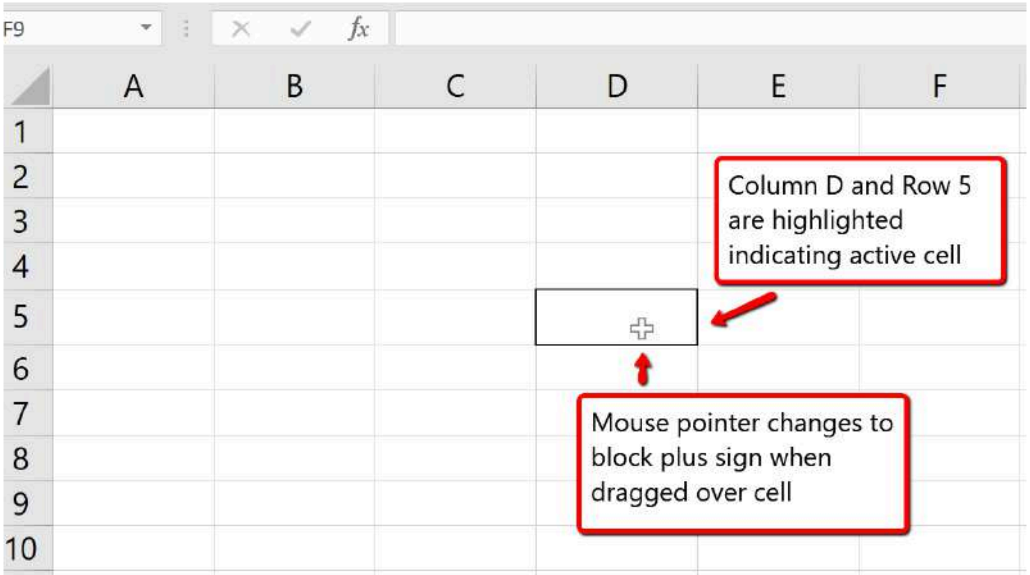


Figure 1.4 Activating a Cell Location

1. Move the mouse pointer to cell A1.
2. Click and hold the left mouse button and drag the mouse pointer back to cell D5.
3. Release the left mouse button. You should see several cells highlighted, as shown in **Figure 1.5**.

This is referred to as a ***cell range*** and is documented as follows: **A1:D5**. Any two cell locations separated by a colon are known as a cell range. The first cell is the top left corner of the range, and the second cell is the lower right corner of the range.

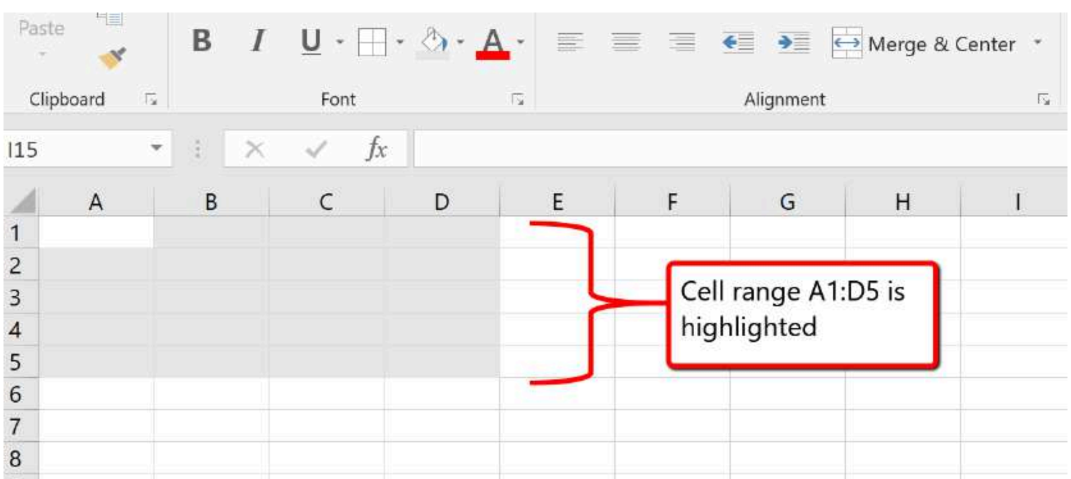


Figure 1.5 Highlighting a Range of Cells

1. At the bottom of the screen, you'll see a sheet tab indicated by "**Sheet1**". Clicking on the + adds additional worksheets. This is how you open or add a worksheets within a workbook. To see how this works, click on the + to add another worksheet so that you now have two sheets
2. Click the Sheet1 worksheet tab at the bottom of the worksheet to return to the worksheet shown in **Figure 1.5**.

Keyboard Shortcuts

Basic Worksheet Navigation

- Use the arrow keys on your keyboard to activate cells on the worksheet.
- Hold the SHIFT key and press the arrow keys on your keyboard to highlight a range of cells in a worksheet.
- Hold the CTRL key while pressing the PAGE DOWN or PAGE UP keys to open other worksheets in a workbook.
- Mac Users: Hold down the Fn and Command keys and press the left or right arrow keys

The Excel Ribbon

Excel's features and commands are found in the Ribbon, which is the upper area of the Excel screen that contains several tabs running across the top. Each tab provides access to a different set of Excel commands. **Figure 1.6** shows the commands available in the Home tab of the Ribbon. **Table 1.1 "Command Overview for Each Tab of the Ribbon"** provides an overview of the commands that are found in each tab of the Ribbon.

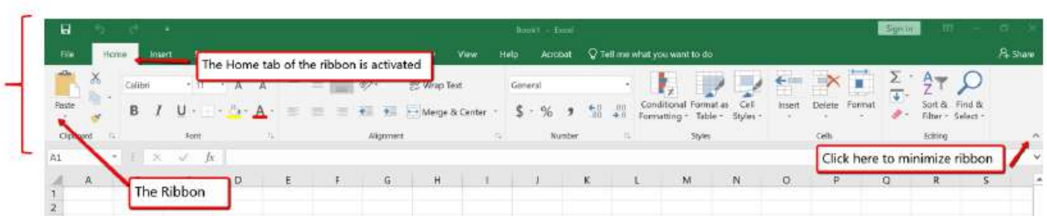


Figure 1.6 Home Tab of Ribbon

The Excel for Mac ribbon, as shown in **Figure 1.6a** below, has two primary differences:

- The older dropdown menu structure is still available with Excel for Mac.
- The specific commands and tools within each tab are slightly different between the two Excel Ribbons. Some of the commands found within the Excel for Windows Ribbon tabs are located within the dropdown menu structure in the Excel for Mac version. So, if you can't find the tool on the Excel for Mac Ribbon, then try to find the tool by looking through the dropdown menu instead.

Figure 1.6a Home tab of Excel for Mac Ribbon with dropdown menu structure

Group Title Names on the Ribbon

If you look closely at the Excel Ribbon (See Figure 1.6 above), you will see that the Ribbon is separated in groups of tool buttons, and each group has a title name. On Home tab, the group title names are "Clipboard", "Font", "Alignment", "Number", "Styles". "Cells", "Editing", etc. The tool buttons within each group are all related to the group title.

Mac Users Only: The default "View" for the Excel for Mac ribbon **does not display** these "group title names". Notice in Figure 1.6a above, there are no group title names. It is a good idea to change this "view" so you can see the group title names. Here are the steps:

1. Click the **"Excel" menu option at top left above the Ribbon**
2. Choose **"Preferences"**
3. Click the **"View" button**
4. Scroll down and check the box for **"Group Titles"**
5. Close the "View" dialog box. The group title names should now display as shown in Figure 6.1 (not Figure 6.1a) above

Table 1.1 Command Overview for Each Tab of the Ribbon

Tab Name	Description of Commands
File	Also known as the Backstage view of the Excel workbook. Contains all commands for opening, closing, saving, and creating new Excel workbooks. Includes print commands, document properties, e-mailing options, and help features. The default settings and options are also found in this tab.
Home	Contains the most frequently used Excel commands. Formatting commands are found in this tab along with commands for cutting, copying, pasting, and for inserting and deleting rows and columns.
Insert	Used to insert objects such as charts, pictures, shapes, PivotTables, Internet links, symbols, or text boxes.
Page Layout	Contains commands used to prepare a worksheet for printing. Also includes commands used to show and print the gridlines on a worksheet.
Formulas	Includes commands for adding mathematical functions to a worksheet. Also contains tools for auditing mathematical formulas.
Data	Used when working with external data sources such as Microsoft® Access®, text files, or the Internet. Also contains sorting commands and access to scenario tools.
Review	Includes Spelling and Track Changes features. Also contains protection features to password protect worksheets or workbooks.
View	Used to adjust the visual appearance of a workbook. Common commands include the Zoom and Page Layout view.
Help	This tab provides access to help and support features such as contacting Microsoft support, sending feedback, suggesting a new feature, and community discussion groups. **This tab is not available with Excel for Mac.**
Draw	Provides drawing options for using a digital pen, mouse or finger depending on the type of device (laptop with touch screen, tablet, computer, etc). This tab is not visible by default. See below on how to customize the Ribbon to add or remove tabs.
Developer	Provides access to some advanced features such as macros, form controls, and XML commands. This tab is not visible by default. See below on how to customize the Ribbon to add or remove tabs.

The Ribbon shown in **Figure 1.6 and Figure 1.6a** (above) is full, or maximized. The benefit of having a full Ribbon is that the commands are always visible while you are developing a worksheet. However, depending on the screen dimensions of your computer, you may find that the Ribbon takes up too much vertical space on your worksheet. If this is the case, you can minimize the Ribbon by clicking the button shown in **Figure 1.6**. When minimized, the Ribbon will show only the tabs and not the command buttons. When you click on a tab, the command buttons will appear until you select a command or click anywhere on your worksheet.

To hide the Ribbon with Excel for Mac you can use the keyboard shortcut:

Hold down the “Command and Option” keys and tap the “R” key

The same keyboard shortcut will unhide the Ribbon as well.

How to Customize the Excel Ribbon

Here are the steps to add additional tabs to the Excel Ribbon

1. Click the File tab and choose Options
2. Click on "Customize Ribbon" at the left side of the Options screen
3. Click the checkbox next to the Tab name that you want to add (See Figure 1.7 below)

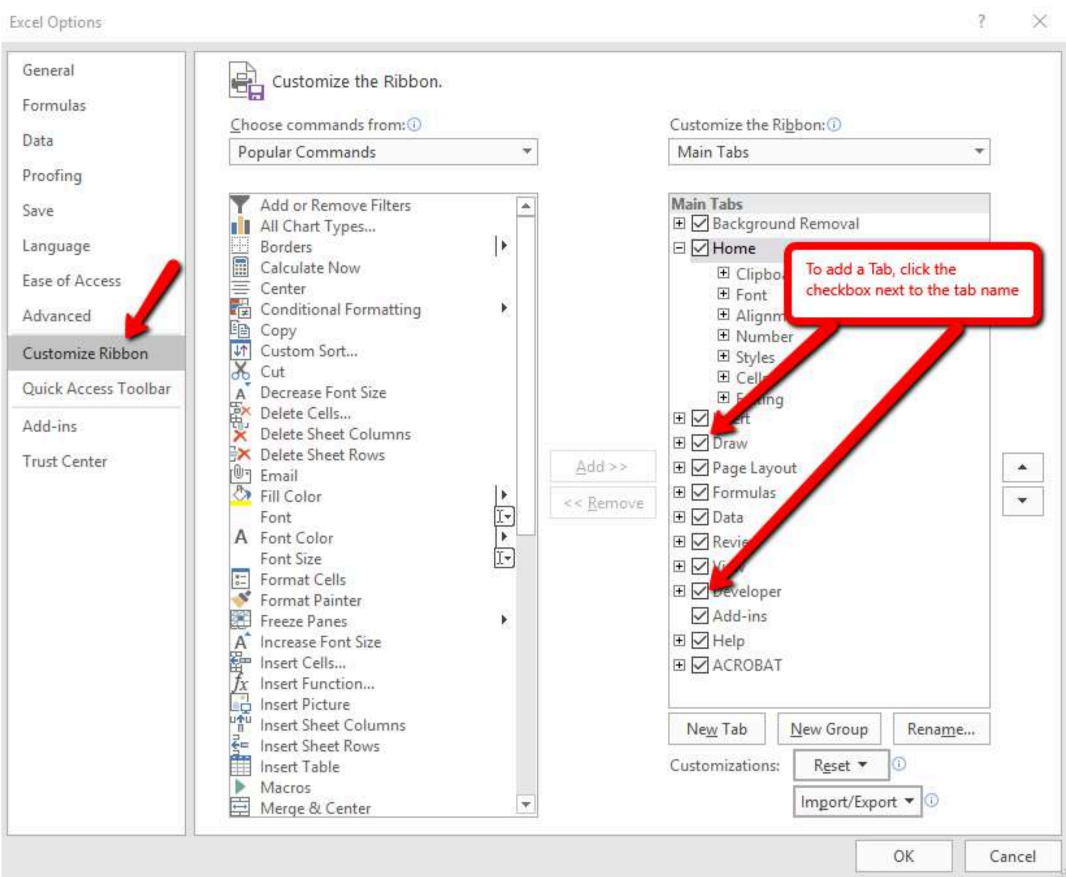


Figure 1.7 Customize the Ribbon Dialog Box

Keyboard Shortcuts

Minimizing or Maximizing the Ribbon

- Hold down the CTRL key and press the F1 key.
- Hold down the CTRL key and press the F1 key again to maximize the Ribbon.

- Mac Users: Hold down the Command and Option keys and press R

Quick Access Toolbar and Right-Click Menu

The Quick Access Toolbar is found at the upper left side of the Excel screen above the Ribbon, as shown in **Figure 1.7**. This area provides access to the most frequently used commands, such as Save and Undo. You also can customize the Quick Access Toolbar by adding commands that you use on a regular basis. By placing these commands in the Quick Access Toolbar, you do not have to navigate through the Ribbon to find them. To customize the Quick Access Toolbar, click the down arrow as shown in **Figure 1.8**. This will open a menu of commands that you can add to the Quick Access Toolbar. If you do not see the command you are looking for on the list, select the More Commands option.

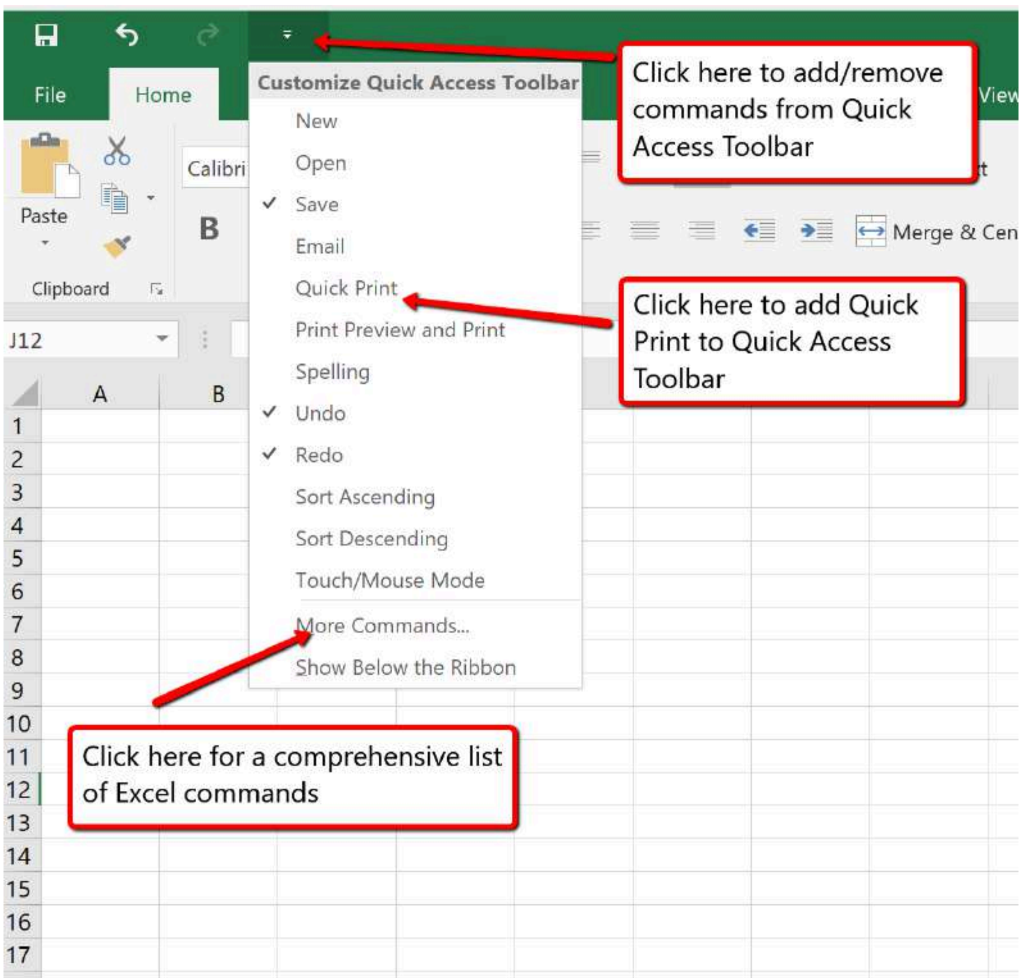


Figure 1.8 Customizing the Quick Access Toolbar

In addition to the Ribbon and Quick Access Toolbar, you can also access many commands by right clicking anywhere on the worksheet. **Figure 1.9** shows an example of the commands available in the right-click menu.

There is no "Right-click" option for Excel for Mac. To access the same commands with Excel for Mac, hold down the Control key and click the mouse button.

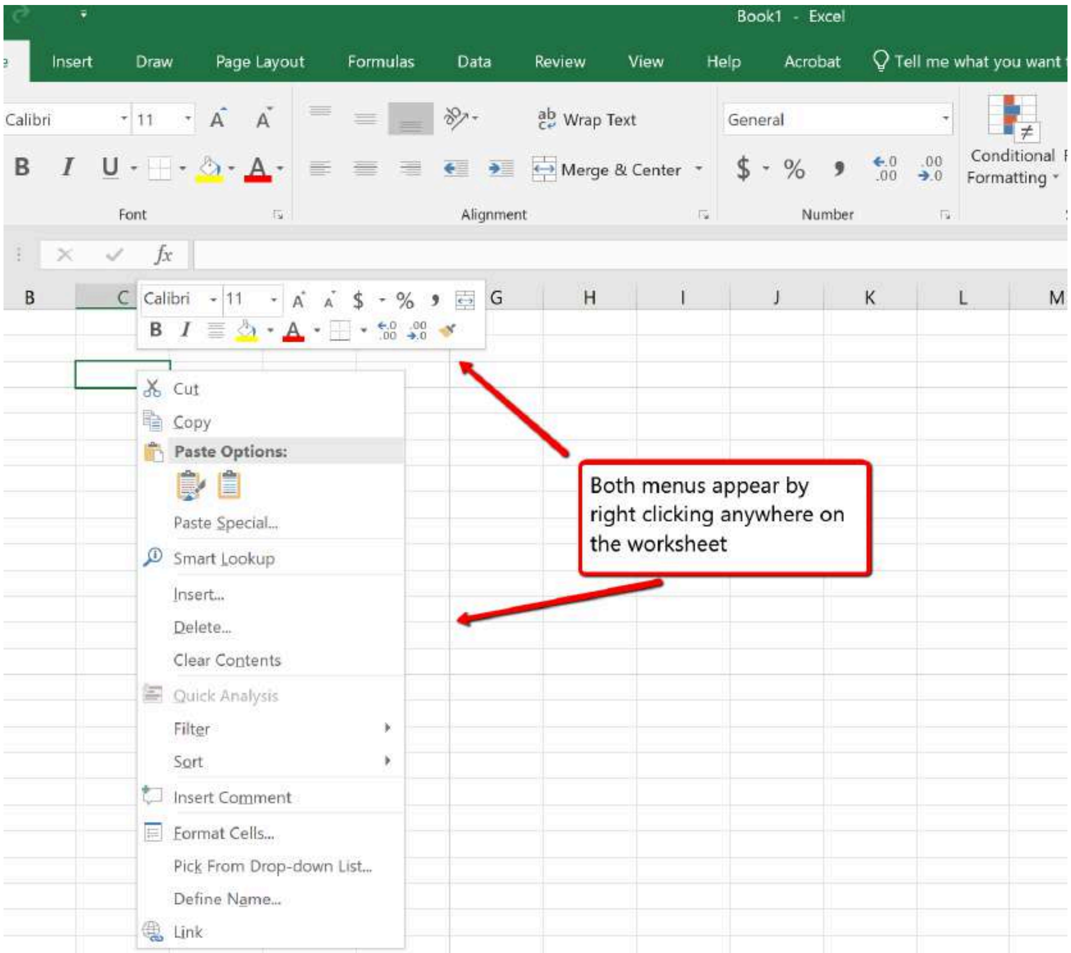


Figure 1.9
Right-Click Menu

The File Tab

The File tab is also known as the **Backstage** view of the workbook. It contains a variety of features and commands related to the workbook that is currently open, new workbooks, or workbooks stored in other locations on your computer or network. **Figure 1.10** shows the options available in the File tab or Backstage view. To leave the Backstage view and return to the worksheet, click the arrow in the upper left-hand corner as shown below.

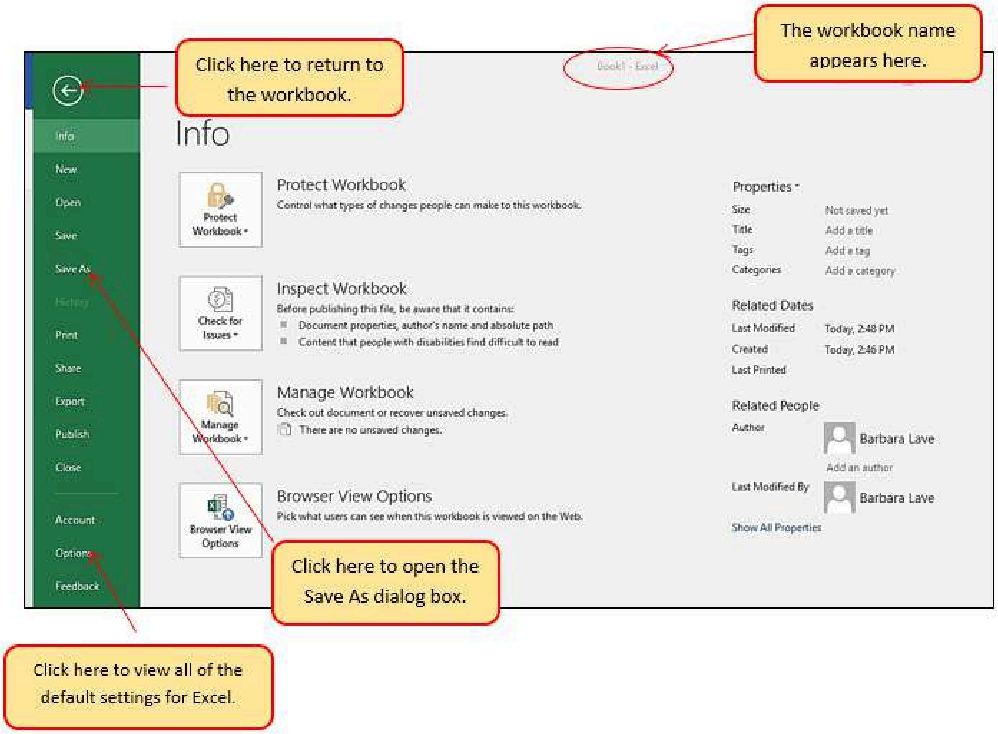


Figure 1.10 File Tab or Backstage View of a Workbook

Included in the File tab are the default settings for the Excel application that can be accessed and modified by clicking the Options button. **Figure 1.11** shows the Excel Options window, which gives you access to settings such as the default font style, font size, and the number of worksheets that appear in new workbooks.

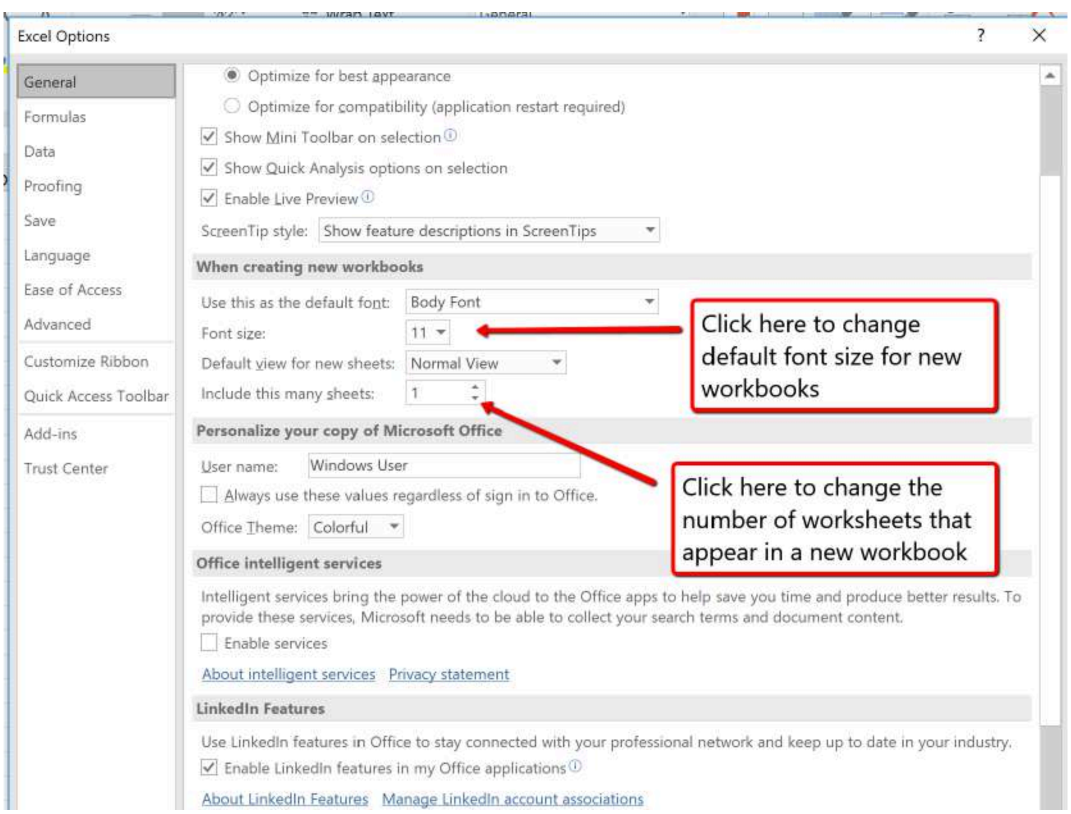


Figure 1.11 Excel Options Window

To access these same options in Excel for Mac, you must click the “Excel” menu option and choose “Preferences” (see **Figure 1.12** below)

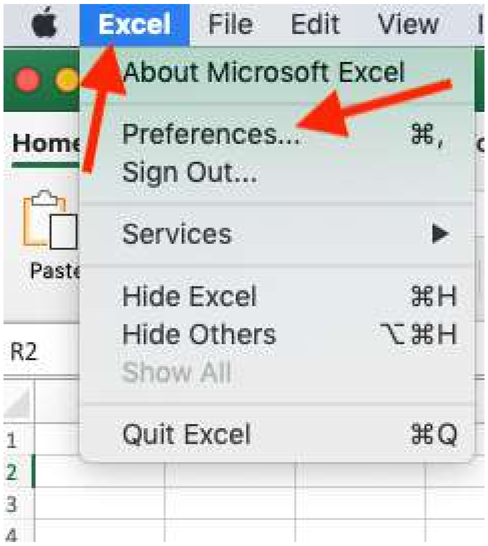


Figure 1.12 The Excel for Mac “Excel” menu option

Saving Workbooks (Save As)

Once you create a new workbook, you will need to change the file name and choose a location on your computer or network to save that file. It is important to remember where you save this workbook on your computer or network as you will be using this file in the **Section 1.2 “Entering, Editing, and Managing Data”** to construct the workbook shown in **Figure 1.1**. The process of saving can be different with different versions of Excel. Please be sure you follow the steps for the version of Excel you are using. The following steps explain how to save a new workbook and assign it a file name.

Saving Workbooks in Excel 365

1. If you have not done so already, open a blank workbook in Excel.
2. Click the File tab and then the **Save As** button in the left side of the Backstage view window. This will open the **Save As** dialog box.
3. Determine a location for saving on your computer by clicking **Browse** on the left side to open the **Save As** dialog box.
4. Click in the File Name box near the bottom of the **Save As** dialog box. Type the new file name: **CH1 Merchandise City Sales Data**
5. Review the settings in the screen for correctness and click the Save button.

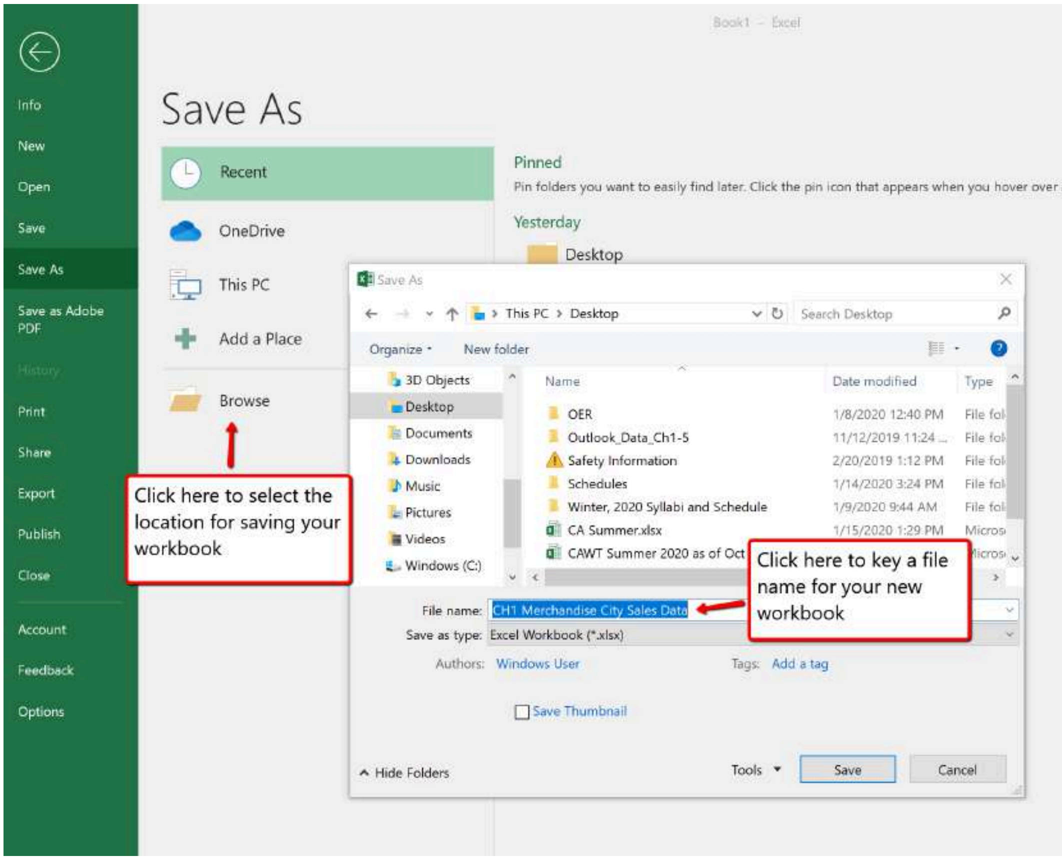


Figure 1.13 Save As Dialog entries for Excel 365

Keyboard Shortcuts

Save As

- Press the F12 key and use the tab and arrow keys to navigate around the Save As dialog box. Use the ENTER key to make a selection.
- Or press the ALT key on your keyboard. You will see letters and numbers, called Key Tips, appear on the Ribbon. Press the F key on your keyboard for the File tab and then the A key. This will open the Save As dialog box.
- The Mac shortcut is: Hold down the Command and Shift keys and press S

Skill Refresher

Saving Workbooks (Save As)

1. Click the File tab on the Ribbon.
2. Click the Save As option.
3. Click on Browse to select a location on your PC to save.
4. Click in the File name box and type a new file name if needed.
5. Click the down arrow next to the "Save as type" box and select the appropriate file type if needed. Excel will default to the file type of .xlsx
6. Click the Save button.

The Status Bar

The Status Bar is located below the worksheet tabs on the Excel screen (**see Figure 1.13**). It displays a variety of information, such as the status of certain keys on your keyboard (e.g., CAPS LOCK), the available views for a workbook, the magnification of the screen, and mathematical functions that can be performed when data are highlighted on a worksheet. You can customize the Status Bar as follows:

1. Place the mouse pointer over any area of the Status Bar and right click to display the "Customize Status Bar" list of options (**see Figure 1.14**).
   Mac Users: use "Control-click" on the Status Bar to display the "Customize Status Bar" options.
2. Select the Caps Lock option from the menu (see **Figure 1.14**).
3. Press the CAPS LOCK key on your keyboard. You will see the Caps Lock indicator on the lower right side of the Status Bar.
4. Press the CAPS LOCK on your keyboard again. The indicator on the Status Bar goes away.

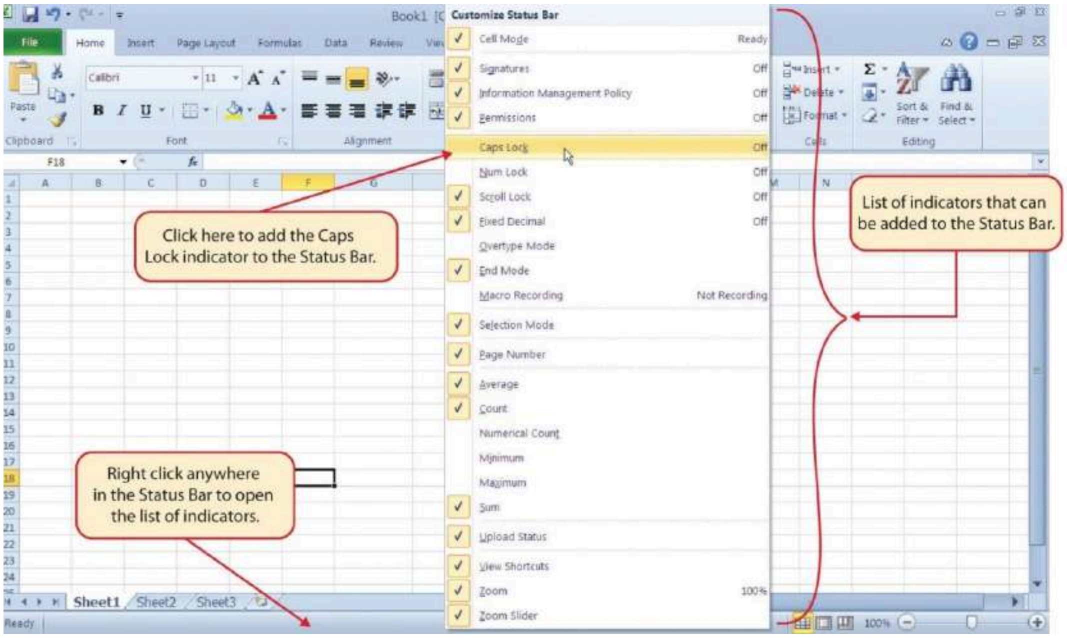


Figure 1.14 Customizing the Status Bar

Excel Help

The Help feature provides extensive information about the Excel application. Although some of this information may be stored on your computer, the Help window will automatically connect to the Internet, if you have a live connection, to provide you with resources that can answer most of your questions. You can open the Excel Help window by clicking the question mark in the upper right area of the screen or ribbon. With newer versions of Excel, use the query box to enter your question and select from helpful option links or select the question mark from the dropdown list to launch Excel Help windows.

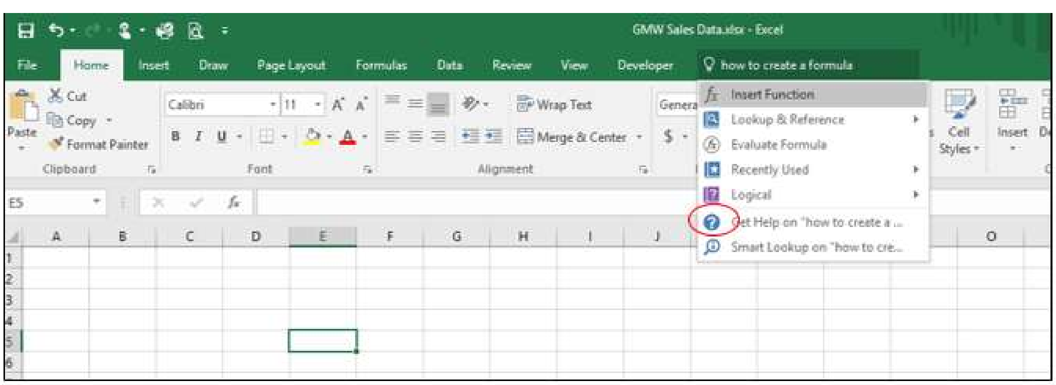


Figure 1.15 Excel Help Window

Keyboard Shortcuts

Excel Help

- Press the F1 key on your keyboard.
- Mac Users: Press F1 or hold down the Command key and press /

Key Takeaways

- Excel is a powerful tool for processing data for the purposes of making decisions.
- You can find Excel commands throughout the tabs in the Ribbon.
- You can customize the Quick Access Toolbar by adding commands you frequently use.
- You can add or remove the information that is displayed on the Status Bar.
- The Help window provides you with extensive information about Excel.

Attribution

Adapted by Barbara Lave from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

2.XLSX.2 ENTERING, EDITING, AND MANAGING DATA

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Learning Objectives

1. Understand how to enter data into a worksheet.
2. Examine how to edit data in a worksheet.
3. Examine how Auto Fill is used when entering data.
4. Understand how to delete data from a worksheet and use the Undo command.
5. Examine how to adjust column widths and row heights in a worksheet.
6. Understand how to hide columns and rows in a worksheet.
7. Examine how to insert columns and rows into a worksheet.
8. Understand how to delete columns and rows from a worksheet.
9. Learn how to move data to different locations in a worksheet.

In this section, we will begin the development of the workbook shown in **Figure 1.1**. The skills covered in this section are typically used in the early stages of developing one or more worksheets in a workbook.

Entering Data

You will begin building the workbook shown in **Figure 1.1** by manually entering data into the worksheet. The following steps explain how the column headings in Row 2 are typed into the worksheet:

1. Click cell location A2 on the worksheet.
2. Type the word **Month**.
3. Press the RIGHT ARROW key. This will enter the word into cell A2 and activate the next cell to the right.
4. Type **Unit Sales** and press the RIGHT ARROW key.

5. Repeat step 4 for the words **Average Price** and then again for **Sales Dollars**.

Figure 1.15 shows how your worksheet should appear after you have typed the column headings into Row 2. Notice that the word **Price** in cell location C2 is not visible. This is because the column is too narrow to fit the entry you typed. We will examine formatting techniques to correct this problem in the next section.

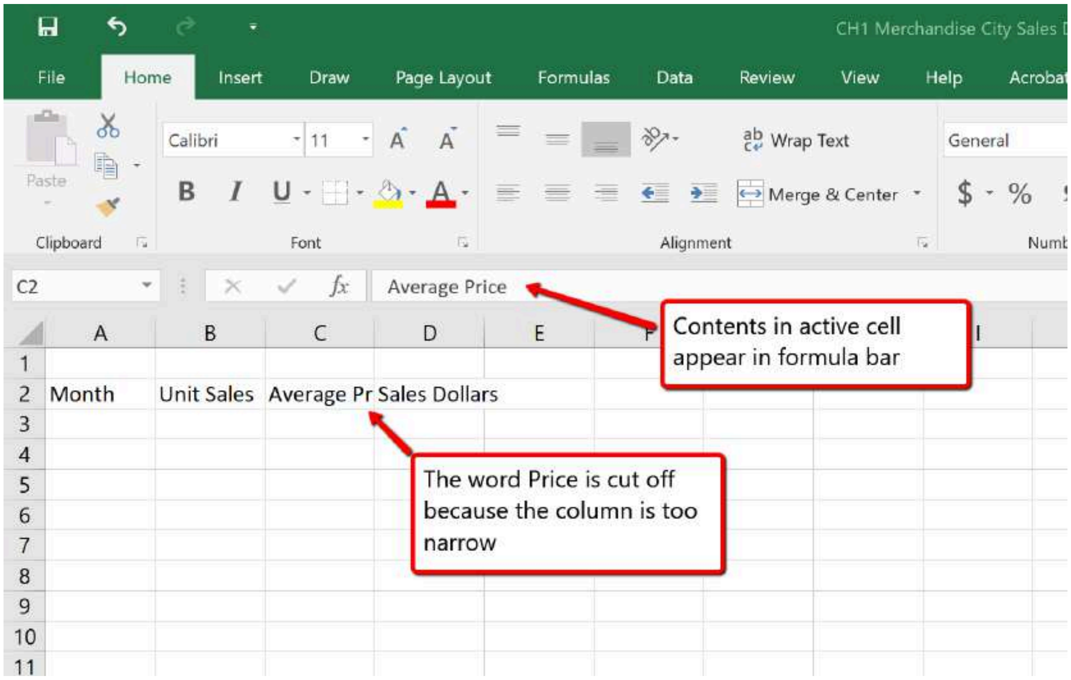


Figure 1.15 Entering Column Headings into a Worksheet

Integrity Check

Column Headings

It is critical to include column headings that accurately describe the data in each column of a worksheet. In professional environments, you will likely be sharing Excel workbooks with coworkers. Good column headings reduce the chance of someone misinterpreting the data contained in a worksheet, which could lead to costly errors depending on your career.

1. Click cell B3.

2. Type the number **2670** and press the ENTER key. After you press the ENTER key, cell B4 will be activated. Using the ENTER key is an efficient way to enter data vertically down a column.
3. Enter the following numbers in cells B4 through B14: **2160**, **515, 590, 1030, 2875, 2700, 900, 775, 1180, 1800**, and **4560**.
4. Click cell C3.
5. Type the number **9.99** and press the ENTER key.
6. Enter the following numbers in cells C4 through C14: **12.49, 14.99, 17.49, 14.99, 12.49, 9.99, 19.99, 19.99, 19.99, 17.49**, and **14.99**.
7. Click cell D3.
8. Type the number 26685 and press the ENTER key.
9. Enter the following numbers in cells D4 through D14: **26937, 7701, 10269, 15405, 35916, 26937, 17958, 15708, 23562, 31416**, and **75125**.
10. When finished, check that the data you entered matches **Figure 1.16**.

Why?

Avoid Formatting Symbols When Entering Numbers

When typing numbers into an Excel worksheet, it is best to avoid adding any formatting symbols such as dollar signs and commas. Although Excel allows you to add these symbols while typing numbers, it slows down the process of entering data. It is more efficient to use Excel's formatting features to add these symbols to numbers after you type them into a worksheet.

Integrity Check

Data Entry

It is very important to proofread your worksheet carefully, especially when you have entered numbers. Transposing numbers when entering data manually into a worksheet is a common error. For example, the number **563** could be transposed to **536.** Such errors can seriously compromise the integrity of your workbook.

Integrity Check

Figure 1.16 shows how your worksheet should appear after entering the data. Check your numbers carefully to make sure they are accurately entered into the worksheet.

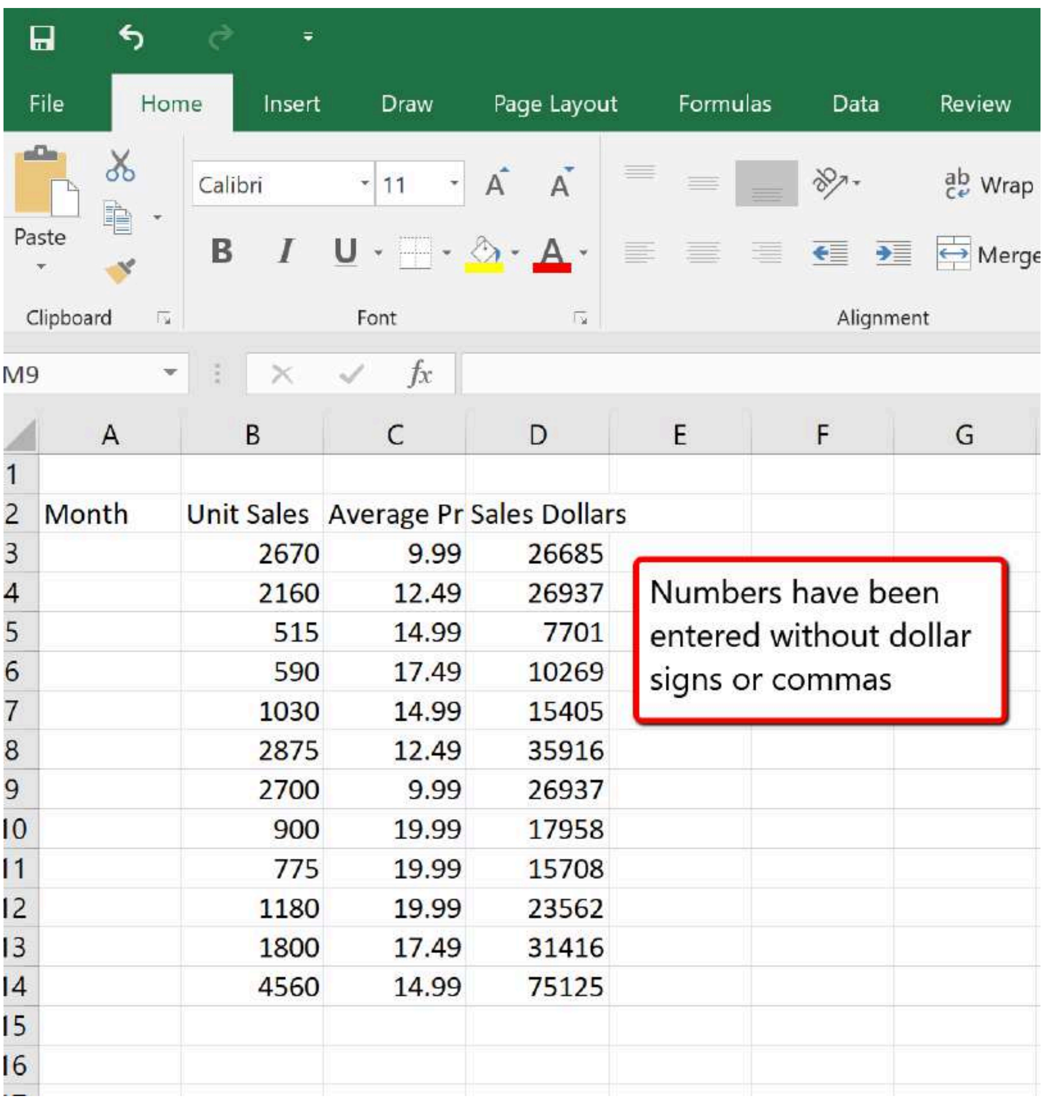


Figure 1.16 Completed Data Entry for Columns B, C, and D

Editing Data

Data that has been entered in a cell can be changed by double clicking the cell location or using the Formula Bar. You may have noticed that as you were typing data into a cell location, the data you typed appeared in the Formula Bar. The Formula Bar can be used for entering data into cells as well as for editing data that already exists in a cell. The following steps provide an example of entering and then editing data that has been entered into a cell location:

1. Click cell A15 in the Sheet1 worksheet.
2. Type the abbreviation **Tot** and press the ENTER key.
3. Click cell A15.
4. Move the mouse pointer up to the Formula Bar. You will see the pointer turn into a cursor. Move the cursor to the end of the abbreviation **Tot** and left click.
5. Type the letters **al** to complete the word Total.
6. Click the check mark to the left of the Formula Bar (see **Figure 1.17**). This will enter the change into the cell.

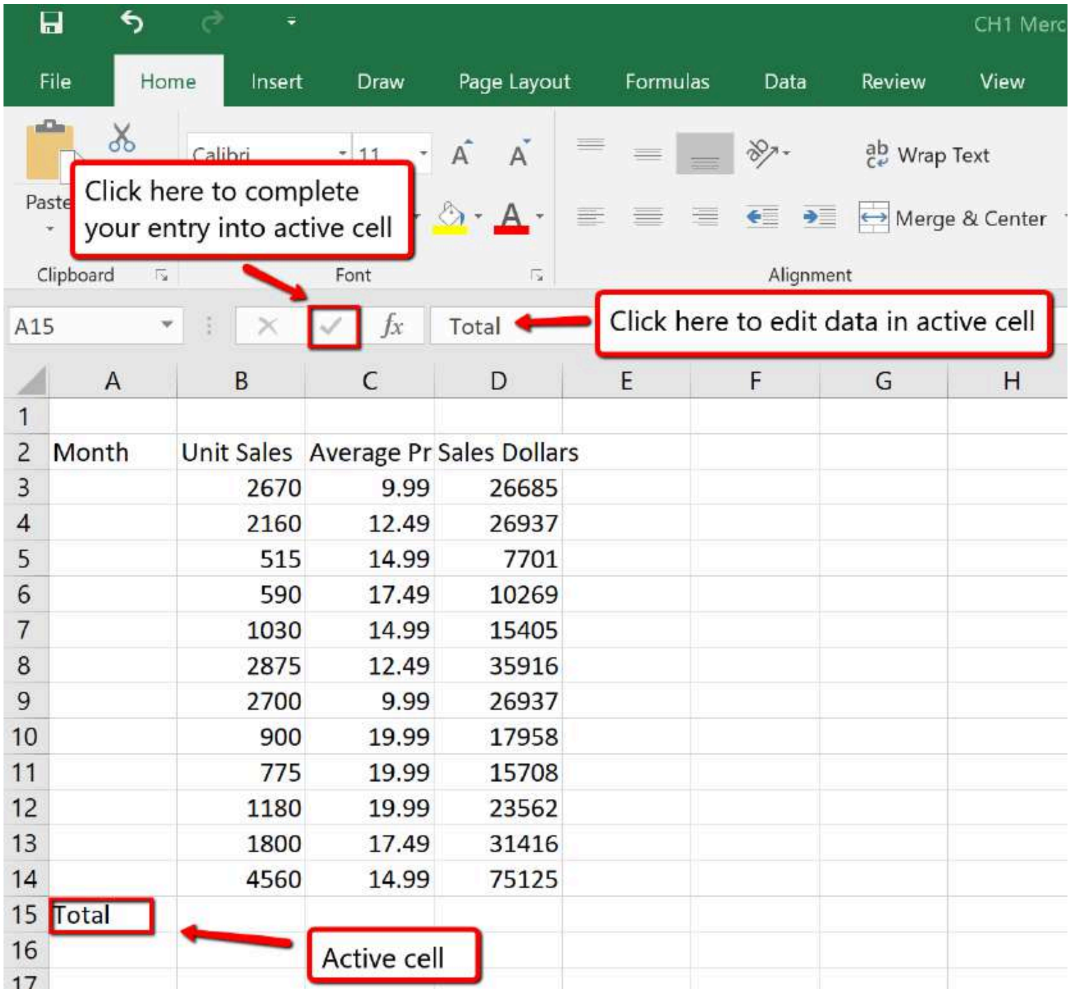


Figure 1.17 Using the Formula Bar to Edit and Enter Data

1. Double click cell A15.
2. Add a space after the word Total and type the word **Sales**.
3. Press the ENTER key.

Keyboard Shortcuts

Editing Data in a Cell

- Activate the cell that is to be edited and press the F2 key on your keyboard.
    - Same for Mac Users

Auto Fill

The Auto Fill feature is a valuable tool when manually entering data into a worksheet. This feature has many uses, but it is most beneficial when you are entering data in a defined sequence, such as the numbers 2, 4, 6, 8, and so on, or nonnumeric data such as the days of the week or months of the year. The following steps demonstrate how Auto Fill can be used to enter the months of the year in Column A:

1. Click cell A3 in the Sheet1 worksheet.
2. Type the word **January** and press the ENTER key.
3. Click cell A3 again.
4. Move the mouse pointer to the lower right corner of cell A3. You will see a small square in this corner of the cell; this is called the Fill Handle (See **Figure 1.18**) When the mouse pointer gets close to the Fill Handle, the white block plus sign will turn into a black plus (+) sign.

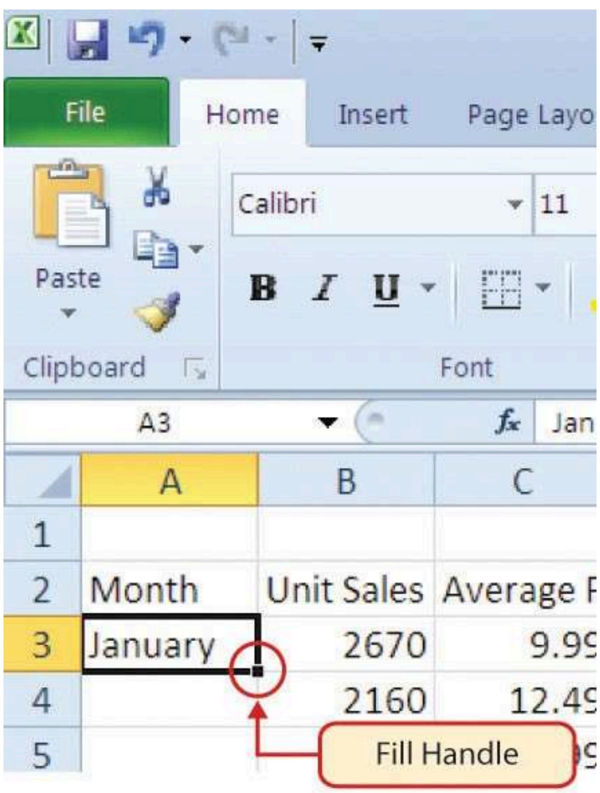


Figure 1.18 Fill Handle

Left click and drag the Fill Handle to cell A14. Notice that the Auto Fill tip box indicates what month will be placed into each cell (see **Figure 1.19**). Release the mouse button when the tip box reads "December."

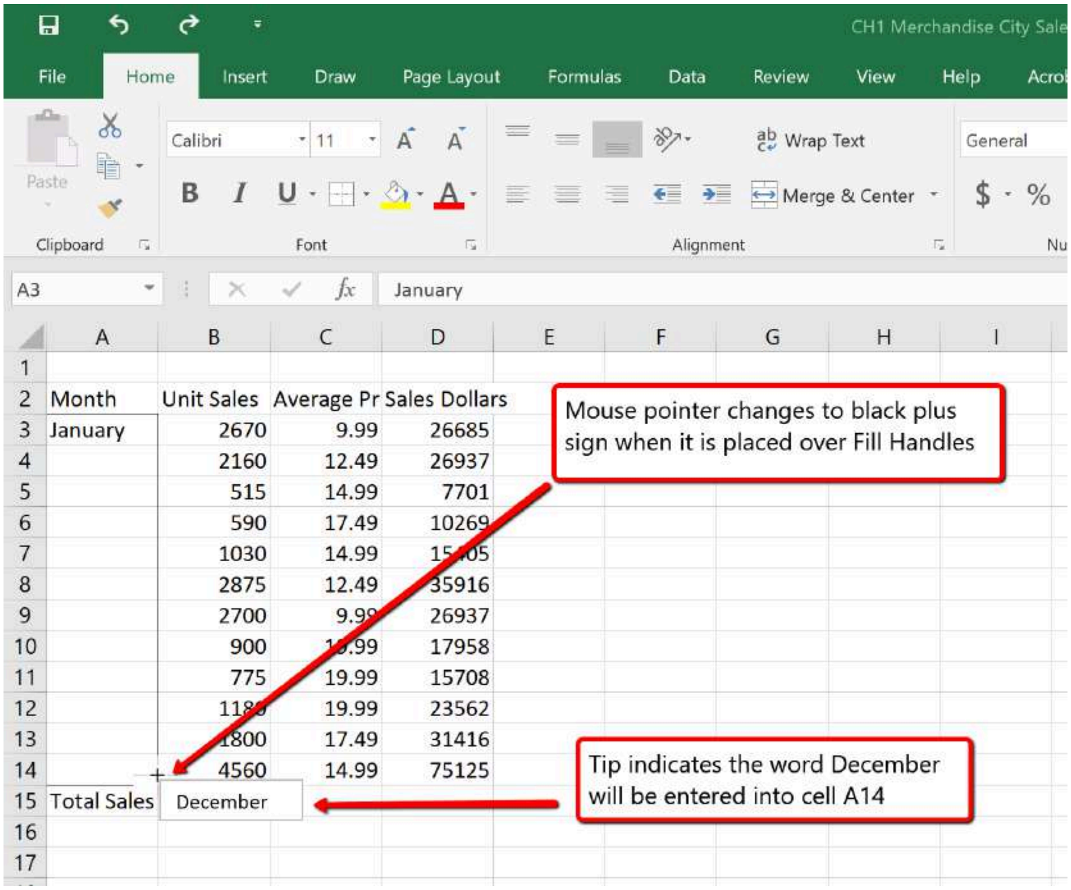


Figure 1.19 Using Auto Fill to Enter the Months of the Year

Once you release the left mouse button, all twelve months of the year should appear in the cell range A3:A14, as shown in **Figure 1.20**. You will also see the Auto Fill Options button. By clicking this button, you have several options for inserting data into a group of cells.

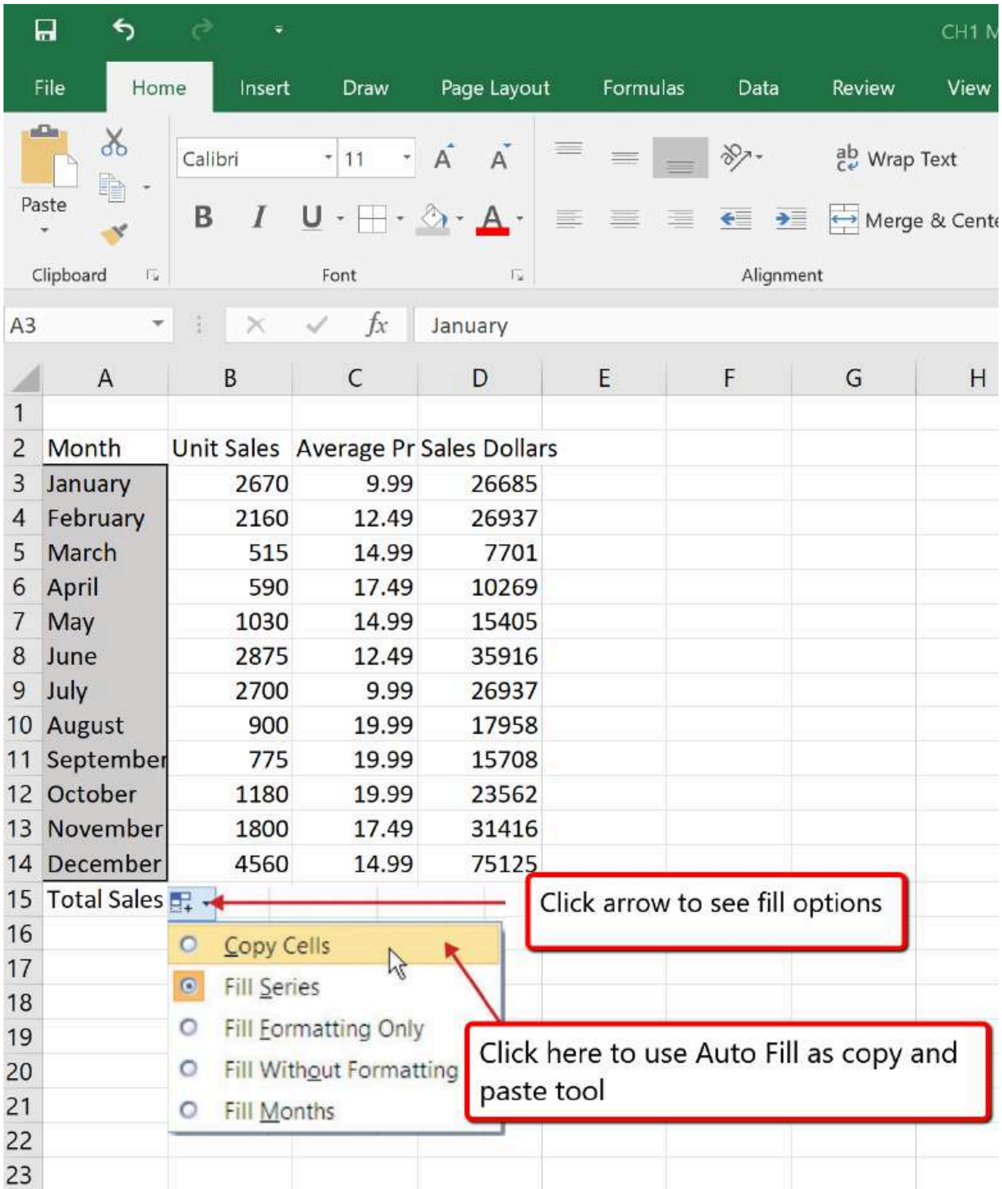


Figure 1.20 Auto Fill Options Button

1. Click the Auto Fill Options button.
2. Click the Copy Cells option. This will change the months in the range A4:A14 to January.
3. Click the Auto Fill Options button again.
4. Click the Fill Months option to return the months of the year to the cell range A4:A14. The Fill Series option will provide the same result.

Deleting Data and the Undo Command

There are several methods for removing data from a worksheet, a few of which are demonstrated here. With each method, you use the Undo command. This is a helpful command in the event you mistakenly remove data from your worksheet. The following steps demonstrate how you can delete data from a cell or range of cells:

1. Click cell C2.
2. Press the **DELETE** key on your keyboard. This removes the contents of the cell.
   Mac Users: Hold down the Fn key and press the Delete key
3. Highlight the range C3:C14. Then left click and drag the mouse pointer down to cell C14.
4. Place the mouse pointer over the Fill Handle. You will see the white block plus sign change to a black plus sign (+).
5. Click and drag the mouse pointer up to cell C3 (see **Figure 1.21**). Release the mouse button. The contents in the range C3:C14 will be removed.

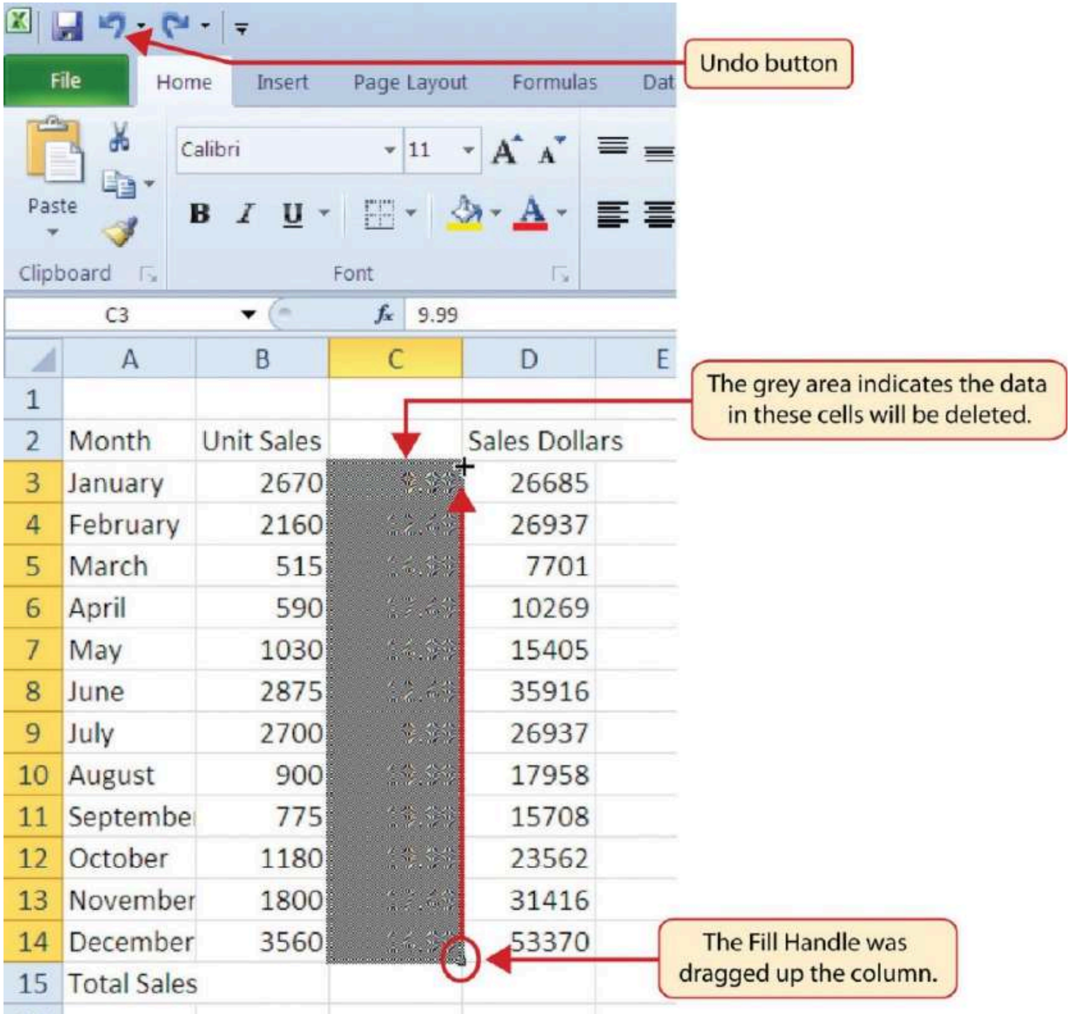


Figure 1.21 Using Auto Fill to Delete Contents of Cell

1. Click the Undo button in the Quick Access Toolbar (see **Figure 1.2**). This should replace the data in the range C3:C14.
2. Click the Undo button again. This should replace the data in cell C2.

Keyboard Shortcuts

Undo Command

- Hold down the CTRL key while pressing the letter Z on your keyboard.
  - Same for Mac Users.
- Highlight the range C2:C14 by placing the mouse pointer over cell C2. Then left click and drag

the mouse pointer down to cell C14.

- Click the Clear button in the Home tab of the Ribbon, which is next to the Cells group of commands (see **Figure 1.22**). This opens a drop-down menu that contains several options for removing or clearing data from a cell. Notice that you also have options for clearing just the formats in a cell or the hyperlinks in a cell.
- Click the Clear All option. This removes the data in the cell range.
- Click the Undo button. This replaces the data in the range C2:C14.

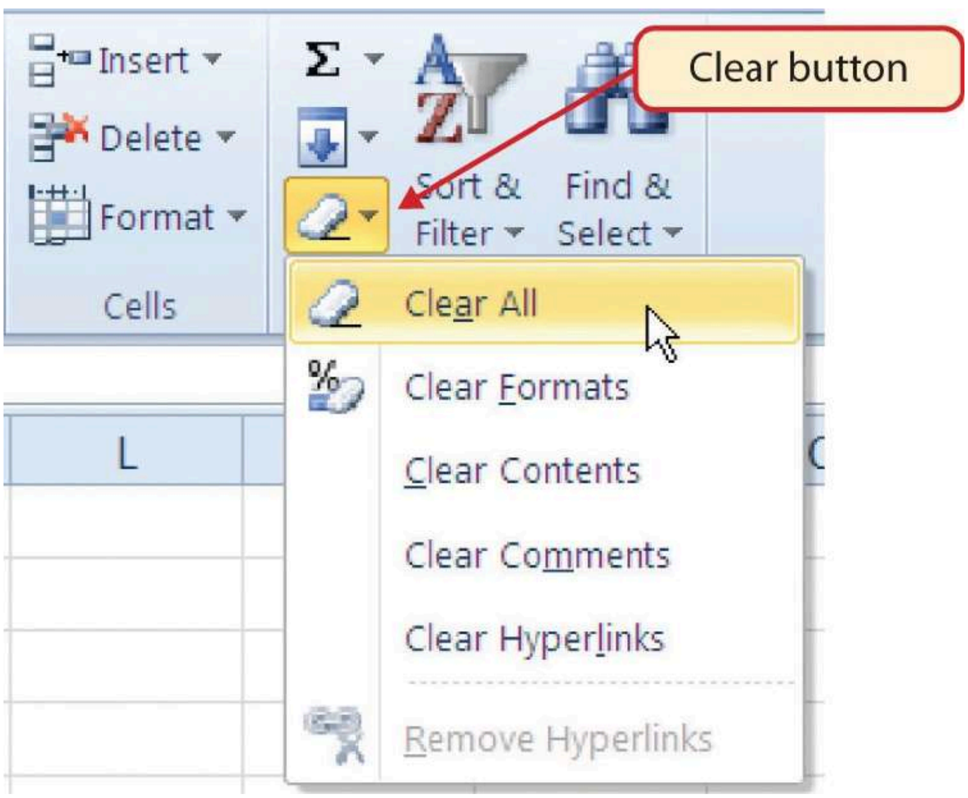


Figure 1.22 Clear Command Drop-Down Menu

Adjusting Columns and Rows

There are a few entries in the worksheet that appear cut off. For example, the last letter of the word September cannot be seen in cell A11. This is because the column is too narrow for this word. The columns and rows on an Excel worksheet can be adjusted to accommodate the data that is being entered into a cell using three different methods. The following steps explain how to adjust the column widths and row heights in a worksheet:

1. Bring the mouse pointer between Column A and Column B in the Sheet1 worksheet, as shown in **Figure 1.23**. You will see the white block plus sign turn into double arrows.
2. Click and drag the column to the right so the entire word September in cell A11 can be seen. As you drag the column, you will see the column width tip box. This box displays the number of characters that will fit into the column using the Calibri 11-point font which is the default setting for font/size.
3. Release the left mouse button.

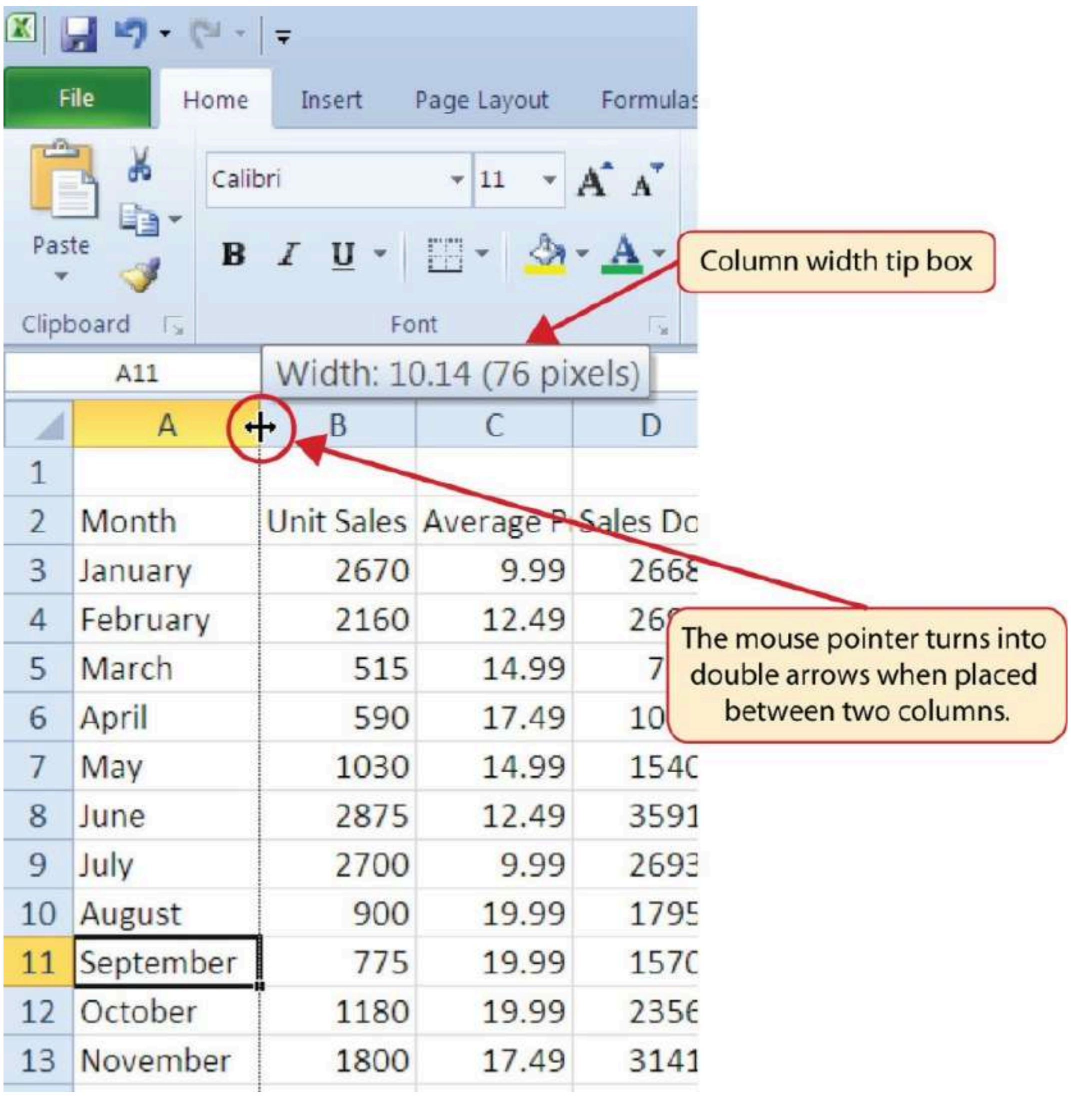


Figure 1.23 Adjusting Column Widths

You may find that using the click-and-drag method is inefficient if you need to set a specific character width for one or more columns. Steps 1 through 6 illustrate a second method for adjusting column widths when using a specific number of characters:

1. Click any cell location in Column A by moving the mouse pointer over a cell location and clicking the

left mouse button. You can highlight cell locations in multiple columns if you are setting the same character width for more than one column.

2. In the Home tab of the Ribbon, left click the Format button in the Cells group.
3. Click the Column Width option from the drop-down menu. This will open the Column Width dialog box.
4. Type the number **13** and click the OK button on the Column Width dialog box. This will set Column A to this character width (see **Figure 1.24**).
5. Once again bring the mouse pointer between Column A and Column B so that the double arrow pointer displays and then double-click to activate AutoFit. This features adjusts the column width based on the longest entry in the column.
6. Use the Column Width dialog box (step 6 above) to reset the width to 13.

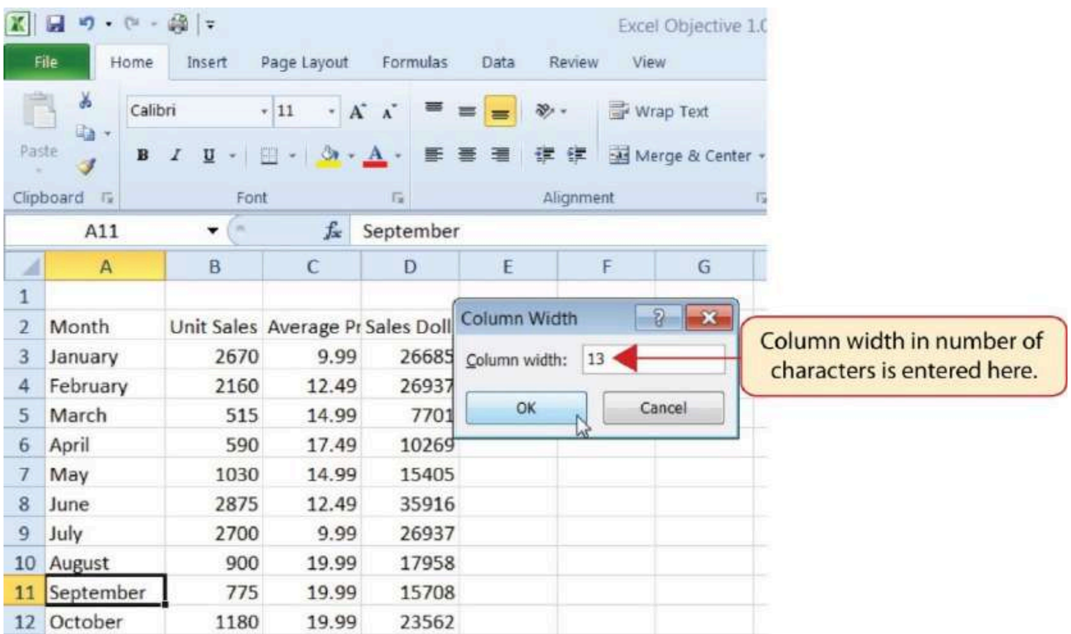


Figure 1.24 Column Width Dialog Box

Keyboard Shortcuts

Column Width

- Press the ALT key on your keyboard, then press the letters H, O, and W one at a time.
- This keyboard shortcut is not available for Excel for Mac

Steps 1 through 4 demonstrate how to adjust row height, which is similar to adjusting column width:

1. Click cell A15.
2. In the Home tab of the Ribbon, left click the Format button in the Cells group.
3. Click the Row Height option from the drop-down menu. This will open the Row Height dialog box.
4. Type the number **24** and click the OK button on the Row Height dialog box. This will set Row 15 to a height of 24 points. A point is equivalent to approximately 1/72 of an inch. This adjustment in row height was made to create space between the totals for this worksheet and the rest of the data.

Keyboard Shortcuts

Row Height

- Press the ALT key on your keyboard, then press the letters H, O, and H one at a time.
- This keyboard shortcut is not available for Excel for Mac

Figure 1.25 shows the appearance of the worksheet after Column A and Row 15 are adjusted.

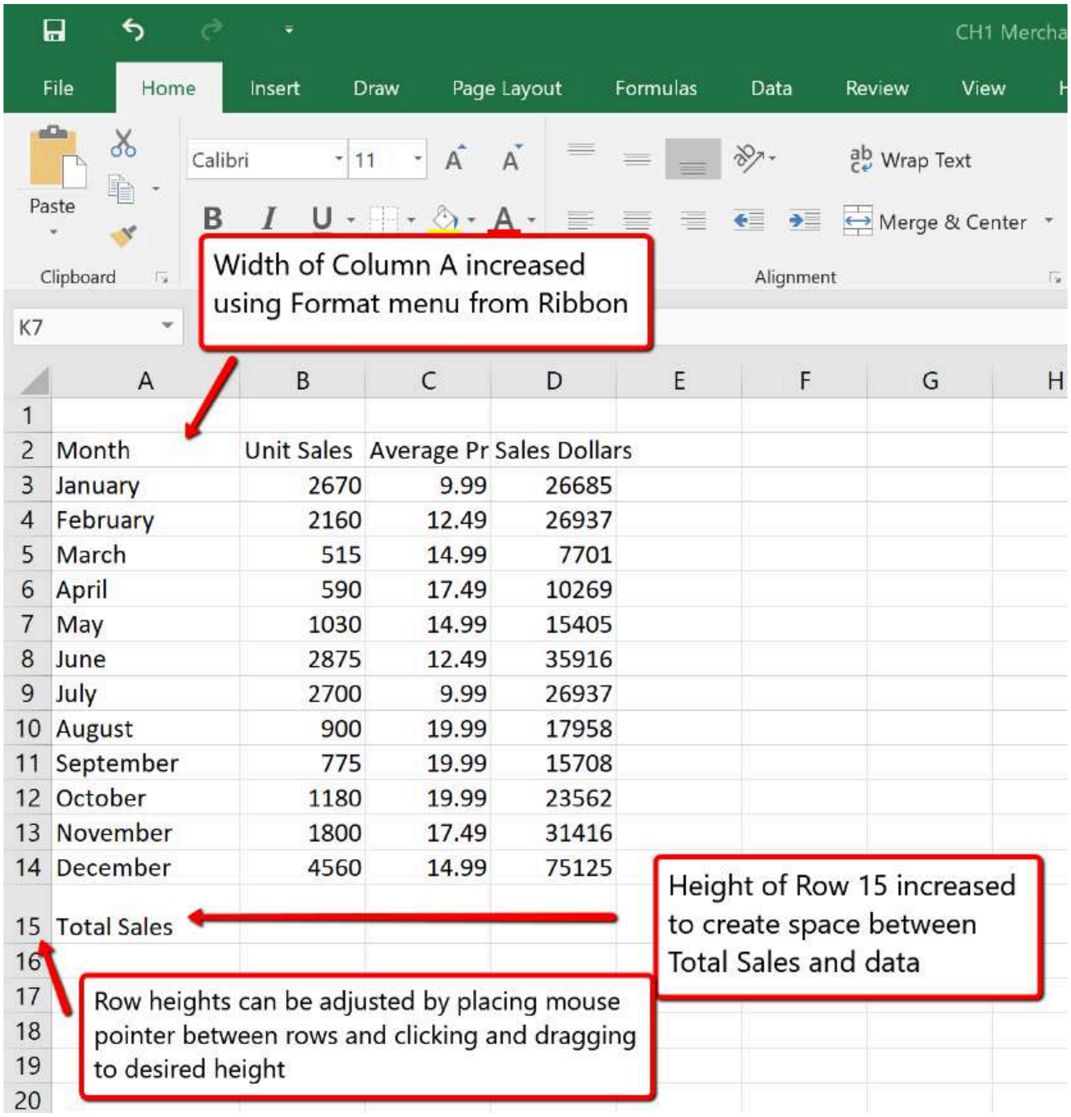


Figure 1.25 Sales Data with Column A and Row 15 Adjusted

Skill Refresher

Adjusting Columns and Rows

1. Activate at least one cell in the row or column you are adjusting.
2. Click the Home tab of the Ribbon.
3. Click the Format button in the Cells group.
4. Click either Row Height or Column Width from the drop-down menu.
5. Enter the Row Height in points or Column Width in characters in the dialog box.
6. Click the OK button.

Hiding Columns and Rows

In addition to adjusting the columns and rows on a worksheet, you can also hide columns and rows. This is a useful technique for enhancing the visual appearance of a worksheet that contains data that is not necessary to display. These features will be demonstrated using the GMW Sales Data workbook. However, there is no need to have hidden columns or rows for this worksheet. The use of these skills here will be for demonstration purposes only.

1. Click cell C1.
2. Click the Format button in the Home tab of the Ribbon.
3. Place the mouse pointer over the Hide & Unhide option in the drop-down menu. This will open a submenu of options.
4. Click the Hide Columns option in the submenu of options (see **Figure 1.26**). This will hide Column C.

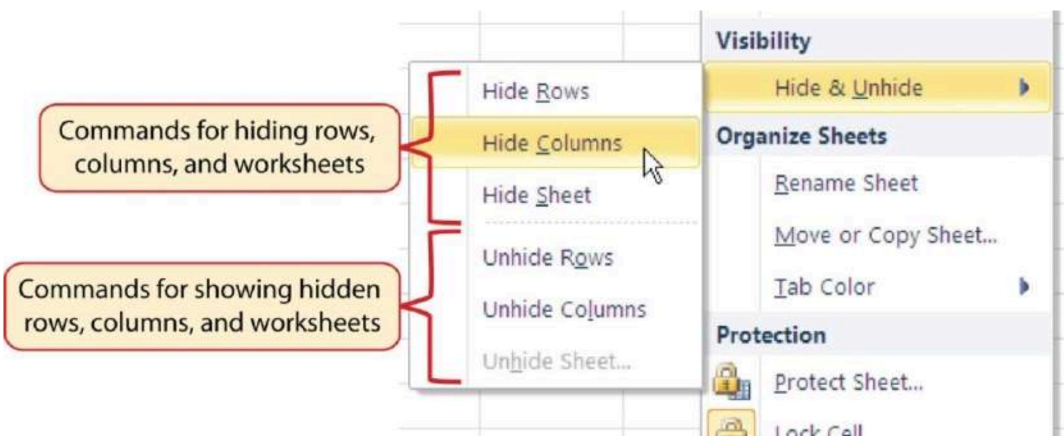


Figure 1.26 Hide & Unhide Submenu

Keyboard Shortcuts

Hiding Columns

- Hold down the CTRL key while pressing the number 0 on your keyboard.
  - Same for Mac Users

Figure 1.27 shows the workbook with Column C hidden in the Sheet1 worksheet. You can tell a column is hidden by the missing letter C.

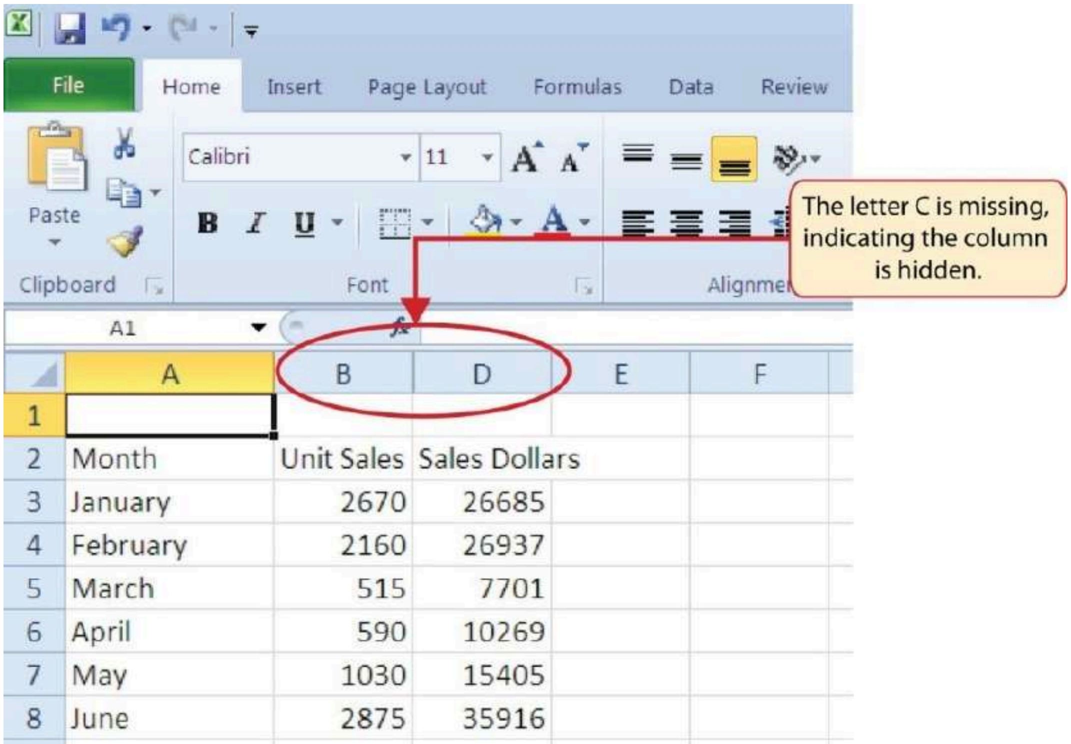


Figure 1.27 Hidden Column

To unhide a column, follow these steps:

1. Select the range B1:D1.
2. Click the Format button in the Home tab of the Ribbon.
3. Place the mouse pointer over the Hide & Unhide option in the drop-down menu.
4. Click the Unhide Columns option in the submenu of options. Column C will now be visible on the worksheet.

Keyboard Shortcuts

Unhiding Columns

- Highlight cells on either side of the hidden column(s), then hold down the CTRL key and the SHIFT key while pressing the close parenthesis key ()) on your keyboard.

- Mac Users: Hold down Control and Shift keys and press the number 0

The following steps demonstrate how to hide rows, which is similar to hiding columns:

1. Click cell A3.
2. Click the Format button in the Home tab of the Ribbon.
3. Place the mouse pointer over the Hide & Unhide option in the drop-down menu. This will open a submenu of options.
4. Click the Hide Rows option in the submenu of options. This will hide Row 3.

Keyboard Shortcuts

Hiding Rows

- Hold down the CTRL key while pressing the number 9 key on your keyboard.
  - Same for Mac Users

To unhide a row, follow these steps:

1. Select the range A2:A4.
2. Click the Format button in the Home tab of the Ribbon.
3. Place the mouse pointer over the Hide & Unhide option in the drop-down menu.
4. Click the Unhide Rows option in the submenu of options. Row 3 will now be visible on the worksheet.

Keyboard Shortcuts

Unhiding Rows

- Highlight cells above and below the hidden row(s), then hold down the CTRL key and the SHIFT key while pressing the open parenthesis key (() on your keyboard.
- Mac Users: Hold down Control and Shift keys and press the number 9

Integrity Check

Hidden Rows and Columns

In most careers, it is common for professionals to use Excel workbooks that have been designed by a coworker. Before you use a workbook developed by someone else, always check for hidden rows and columns. You can quickly see whether a row or column is hidden if a row number or column letter is missing.

Skill Refresher

Hiding Columns and Rows

1. Activate at least one cell in the row(s) or column(s) you are hiding.
2. Click the Home tab of the Ribbon.
3. Click the Format button in the Cells group.
4. Place the mouse pointer over the Hide & Unhide option.
5. Click either the Hide Rows or Hide Columns option.

Skill Refresher

Unhiding Columns and Rows

1. Highlight the cells above and below the hidden row(s) or to the left and right of the hidden column(s).
2. Click the Home tab of the Ribbon.
3. Click the Format button in the Cells group.
4. Place the mouse pointer over the Hide & Unhide option.
5. Click either the Unhide Rows or Unhide Columns option.

Inserting Columns and Rows

Using Excel workbooks that have been created by others is a very efficient way to work because it eliminates the need to create data worksheets from scratch. However, you may find that to accomplish your goals, you need to add additional columns or rows of data. In this case, you can insert blank columns or rows into a worksheet. The following steps demonstrate how to do this:

1. Click cell C1.
2. Click the down arrow on the Insert button in the Home tab of the Ribbon (see **Figure 1.28**).

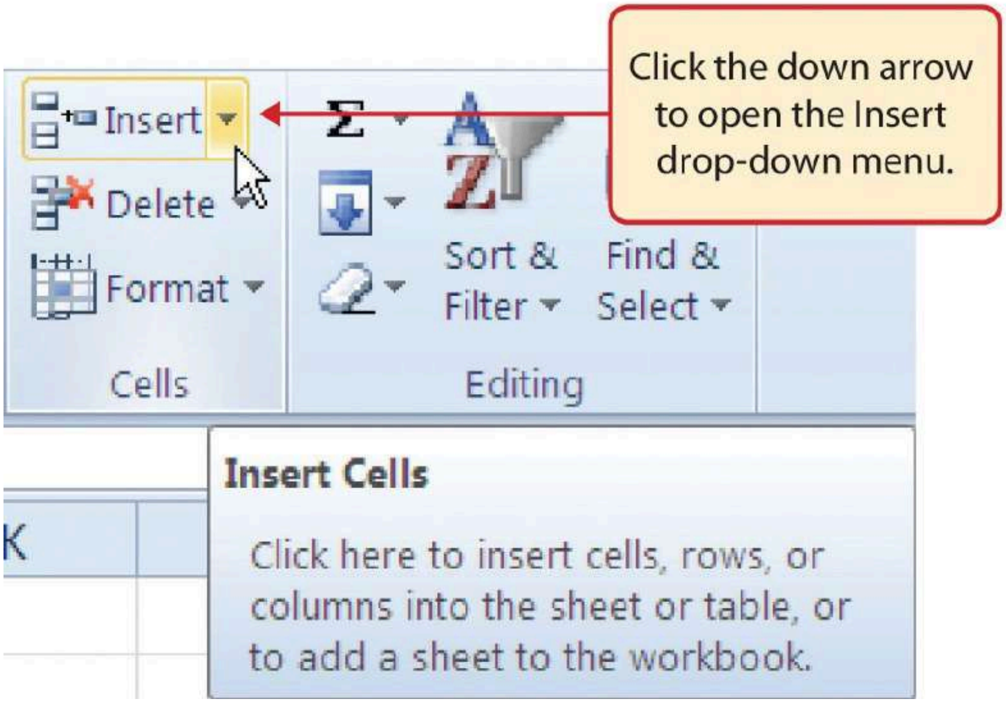


Figure 1.28 Insert Button (Down Arrow)

3. Click the Insert Sheet Columns option from the drop-down menu (see **Figure 1.29**). A blank column will be inserted to the left of Column C. The contents that were previously in Column C now appear in Column D. Note that columns are always inserted to the left of the activated cell.

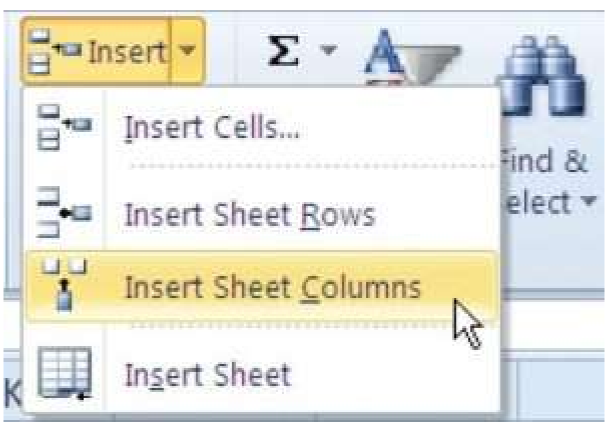


Figure 1.29 Insert Drop-Down Menu

Keyboard Shortcuts

Inserting Columns

- Press the ALT key and then the letters H, I, and C one at a time. A column will be

inserted to the left of the activated cell.
  - Mac Users: First hold down the Control key and press the spacebar to select the column; then hold down the Shift and Controls keys and press the + symbol

4. Click cell A3.
5. Click the down arrow on the Insert button in the Home tab of the Ribbon (see **Figure 1.28**).
6. Click the Insert Sheet Rows option from the drop-down menu (see **Figure 1.29**). A blank row will be inserted above Row 3. The contents that were previously in Row 3 now appear in Row 4. Note that rows are always inserted above the activated cell.

Keyboard Shortcuts

Inserting Rows

- Press the ALT key and then the letters H, I, and R one at a time. A row will be inserted above the activated cell.
- Mac Users: First hold down the Shift key and press the spacebar to select the row; then hold down the Shift and Controls keys and press the + symbol

Skill Refresher

Inserting Columns and Rows

1. Activate the cell to the right of the desired blank column or below the desired blank row.
2. Click the Home tab of the Ribbon.
3. Click the down arrow on the Insert button in the Cells group.
4. Click either the Insert Sheet Columns or Insert Sheet Rows option.

Moving Data

Once data are entered into a worksheet, you have the ability to move it to different locations. The following steps demonstrate how to move data to different locations on a worksheet:

1. Select the range D2:D15.
2. Bring the mouse pointer to the left edge of cell D2. You will see the white block plus sign change to cross arrows (see **Figure 1.30**). This indicates that you can left click and drag the data to a new location.

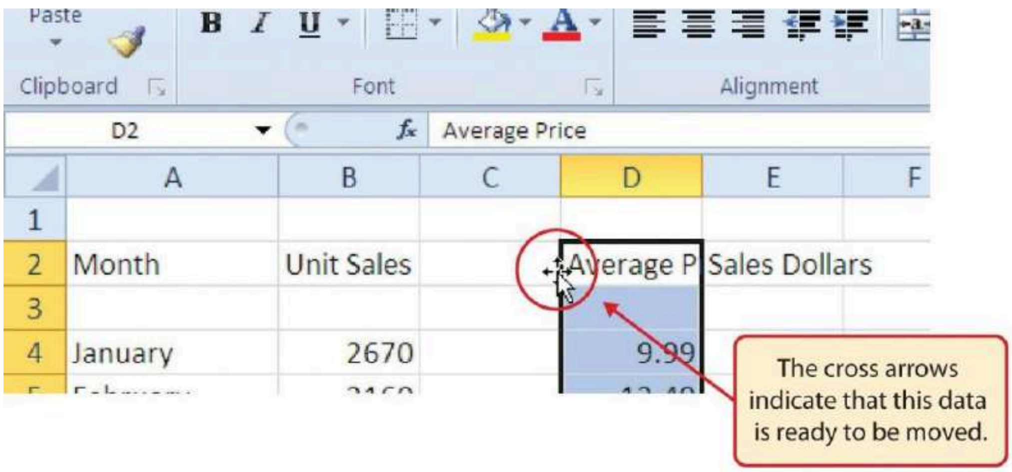


Figure 1.30 Moving Data

Mac Users: when the mouse hovers over the left edge of cell D2, the pointer will turn into a small hand that looks like this: 

3. Left Click and drag the mouse pointer to cell C2.
4. Release the left mouse button. The data now appears in Column C.
5. Click the Undo button in the Quick Access Toolbar. This moves the data back to Column D.

Integrity Check

Moving Data

Before moving data on a worksheet, make sure you identify all the components that belong with the series you are moving. For example, if you are moving a column of data, make sure the column heading is included. Also, make sure all values are highlighted in the column before moving it.

Deleting Columns and Rows

You may need to delete entire columns or rows of data from a worksheet. This need may arise if you need to remove either blank columns or rows from a worksheet or columns and rows that contain data. The methods for removing cell contents were covered earlier and can be used to delete unwanted data. However, if you do not want a blank row or column in your workbook, you can delete it using the following steps:

1. Click cell A3.
2. Click the down arrow on the Delete button in the Cells group in the Home tab of the Ribbon.
3. Click the Delete Sheet Rows option from the drop-down menu (see **Figure 1.31**). This removes Row 3 and shifts all the data (below Row 2) in the worksheet up one row.

Keyboard Shortcuts

Deleting Rows

- Press the ALT key and then the letters H, D, and R one at a time. The row with the activated cell will be deleted.
- Mac Users: First hold down the Shift key and press the spacebar to select the row; then hold down Control key and press the – symbol

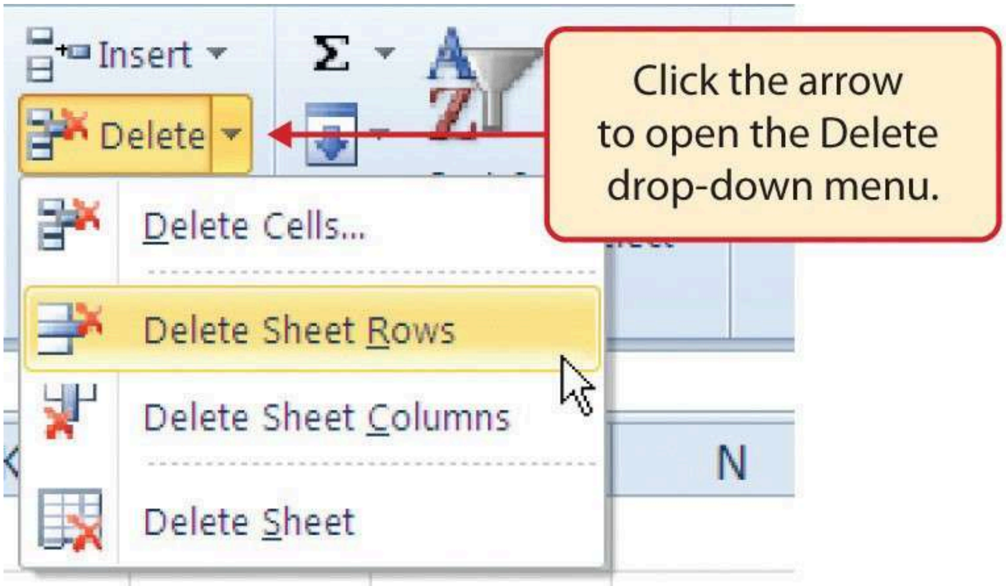


Figure 1.31 Delete Drop-Down Menu

4. Click cell C1.
5. Click the down arrow on the Delete button in the Cells group in the Home tab of the Ribbon.
6. Click the Delete Sheet Columns option from the drop-down menu (see **Figure 1.31**). This removes Column C and shifts all the data in the worksheet (to the right of Column B) over one column to the left.
7. Save the changes to your workbook by clicking either the **Save** button on the Home ribbon; or by selecting the **Save** option from the File menu.

Keyboard Shortcuts

Deleting Columns

- Press the ALT key and then the letters H, D, and C one at a time. The column with the activated cell will be deleted.
- Mac Users: First hold down the Control key and press the spacebar to select the column; then hold down Control key and press the – symbol

Skill Refresher

Deleting Columns and Rows

1. Activate any cell in the row or column that is to be deleted.
2. Click the Home tab of the Ribbon.
3. Click the down arrow on the Delete button in the Cells group.
4. Click either the Delete Sheet Columns or the Delete Sheet Rows option.

Key Takeaways

- Column headings should be used in a worksheet and should accurately describe the data contained in each column.
- Using symbols such as dollar signs when entering numbers into a worksheet can slow down the data entry process.
- Worksheets must be carefully proofread when data has been manually entered.
- The Undo command is a valuable tool for recovering data that was deleted from a worksheet.
- When using a worksheet that was developed by someone else, look carefully for hidden columns or rows.

Attribution

Adapted by Barbara Lave from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

2.XLSX.3 FORMATTING AND DATA ANALYSIS

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

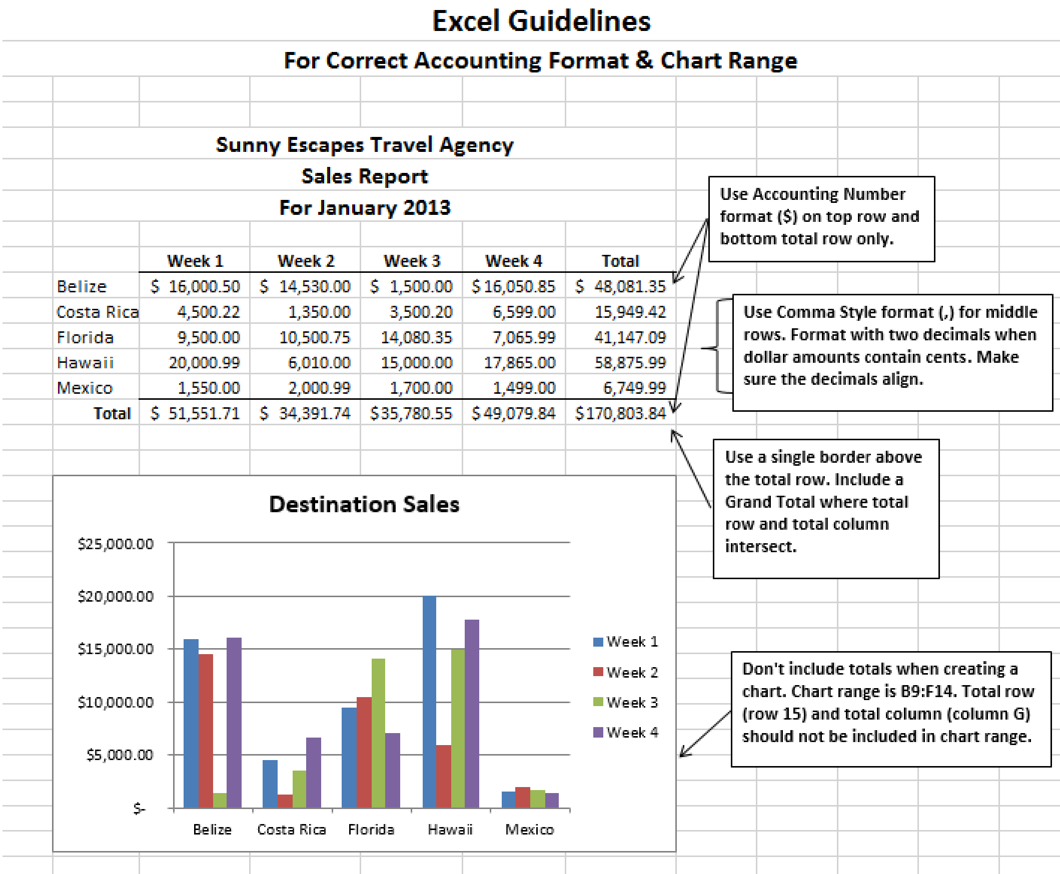
Learning Objectives

1. Use formatting techniques as introduced in the Excel Spreadsheet Guidelines to enhance the appearance of a worksheet.
2. Understand how to align data in cell locations.
3. Examine how to enter multiple lines of text in a cell location.
4. Understand how to add borders to a worksheet.
5. Examine how to use the AutoSum feature to calculate totals.
6. Use the Cut, Copy, and Paste commands to manipulate the data on a worksheet.
7. Understand how to move, rename, insert, and delete worksheet tabs.

This section addresses formatting commands that can be used to enhance the visual appearance of a worksheet. It also provides an introduction to mathematical calculations. The skills introduced in this section will give you powerful tools for analyzing the data that we have been working with in this workbook and will highlight how Excel is used to make key decisions in virtually any career. Additionally, Excel Spreadsheet Guidelines for format and appearance will be introduced as a format for the course and spreadsheets submitted.

Formatting Data and Cells

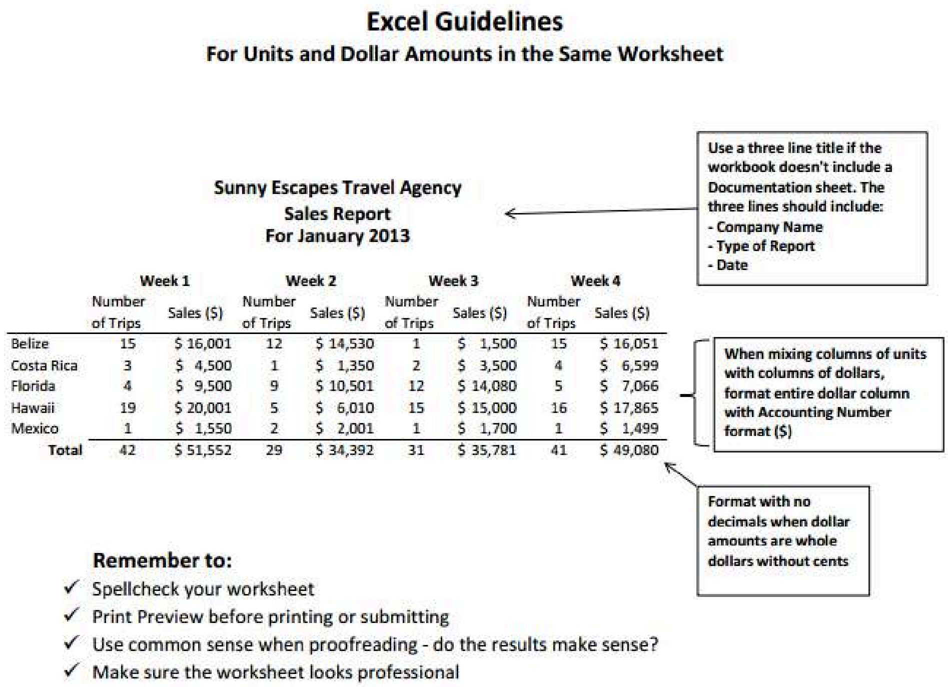
Enhancing the visual appearance of a worksheet is a critical step in creating a valuable tool for you or your coworkers when making key decisions. There are accepted professional formatting standards when spreadsheets contain only currency data. For this course, we will use the following Excel Guidelines for Formatting. The first figure displays how to use Accounting number format when ALL figures are currency. Only the first row of data and the totals should be formatted with the Accounting format. The other data should be formatted with Comma style. There also needs to be a Top Border above the numbers in the total row. If any of the numbers have cents, you need to format all of the data with two decimal places.



	Week 1	Week 2	Week 3	Week 4	Total
Belize	$ 16,000.50	$ 14,530.00	$ 1,500.00	$ 16,050.85	$ 48,081.35
Costa Rica	4,500.22	1,350.00	3,500.20	6,599.00	15,949.42
Florida	9,500.00	10,500.75	14,080.35	7,065.99	41,147.09
Hawaii	20,000.99	6,010.00	15,000.00	17,865.00	58,875.99
Mexico	1,550.00	2,000.99	1,700.00	1,499.00	6,749.99
Total	$ 51,551.71	$ 34,391.74	$35,780.55	$ 49,079.84	$170,803.84

Figure 1.31a

Often, your Excel spreadsheet will contain values that are both currency and non-currency in nature. When that is the case, you'll want to use the guidelines in the following figure:



Excel Guidelines
For Units and Dollar Amounts in the Same Worksheet

Sunny Escapes Travel Agency
Sales Report
For January 2013

	Week 1		Week 2		Week 3		Week 4	
	Number of Trips	Sales ($)	Number of Trips	Sales ($)	Number of Trips	Sales ($)	Number of Trips	Sales ($)
Belize	15	$ 16,001	12	$ 14,530	1	$ 1,500	15	$ 16,051
Costa Rica	3	$ 4,500	1	$ 1,350	2	$ 3,500	4	$ 6,599
Florida	4	$ 9,500	9	$ 10,501	12	$ 14,080	5	$ 7,066
Hawaii	19	$ 20,001	5	$ 6,010	15	$ 15,000	16	$ 17,865
Mexico	1	$ 1,550	2	$ 2,001	1	$ 1,700	1	$ 1,499
Total	42	$ 51,552	29	$ 34,392	31	$ 35,781	41	$ 49,080

Use a three line title if the workbook doesn't include a Documentation sheet. The three lines should include:
- Company Name
- Type of Report
- Date

When mixing columns of units with columns of dollars, format entire dollar column with Accounting Number format ($)

Format with no decimals when dollar amounts are whole dollars without cents

Remember to:

✓ Spellcheck your worksheet
✓ Print Preview before printing or submitting
✓ Use common sense when proofreading - do the results make sense?
✓ Make sure the worksheet looks professional

Figure 1.31b

The following steps demonstrate several fundamental formatting skills that will be applied to the workbook that we are developing for this chapter. Several of these formatting skills are identical to ones that you may have already used in other Microsoft applications such as Microsoft® Word® or Microsoft® PowerPoint®.

1. Select the range A2:D2. Click the Bold button in the Font group of commands in the Home tab of the ribbon.
2. Click the Border button in the Font group of commands in the Home tab of the Ribbon (see **Figure 1.32**). Select the Bottom Border option from the list to achieve the goal of a border on the bottom of row 2 below the column headings.

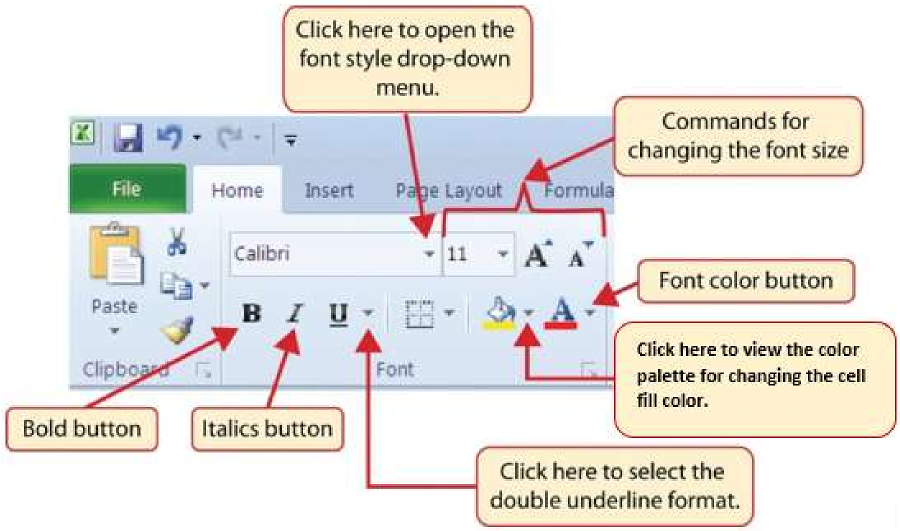


Figure 1.32 Font Group of Commands

Keyboard Shortcuts

Bold Format

- Hold down the CTRL key while pressing the letter B on your keyboard.
- Mac Users: Hold the Control key and press the letter B **or** hold down the Command key and press the letter B

3. Select the range A15:D15.
4. Click the Bold button in the Font group of commands in the Home tab of the Ribbon.
5. Click the Border button in the Font group of commands in the Home tab of the Ribbon (see **Figure 1.32**). Select the Top Border option from the list to achieve the goal of a border on the top of row 15 where totals will eventually display.

Keyboard Shortcuts

Italics Format

- Hold the CTRL key while pressing the letter I on your keyboard.
- Mac Users: Hold the Control key and press the letter I **or** hold down the Command key and press the letter I

Keyboard Shortcuts

Underline Format

- Hold the CTRL key while pressing the letter U on your keyboard.
- Mac Users: Hold the Control key and press the letter U **or** hold down the Command key and press the letter U

Why?

Format Column Headings and Totals

Applying formatting enhancements to the column headings and column totals in a worksheet is a very important technique, especially if you are sharing a workbook with other people. These formatting techniques allow users of the worksheet to clearly see the column headings that define the data. In addition, the column totals usually contain the most important data on a worksheet with respect to making decisions, and formatting techniques allow users to quickly see this information.

1. Select the range B3:B14.
2. Click the Comma Style button in the Number group of commands in the Home tab of the Ribbon. This feature adds a comma as well as two decimal places. (see **Figure 1.33**).

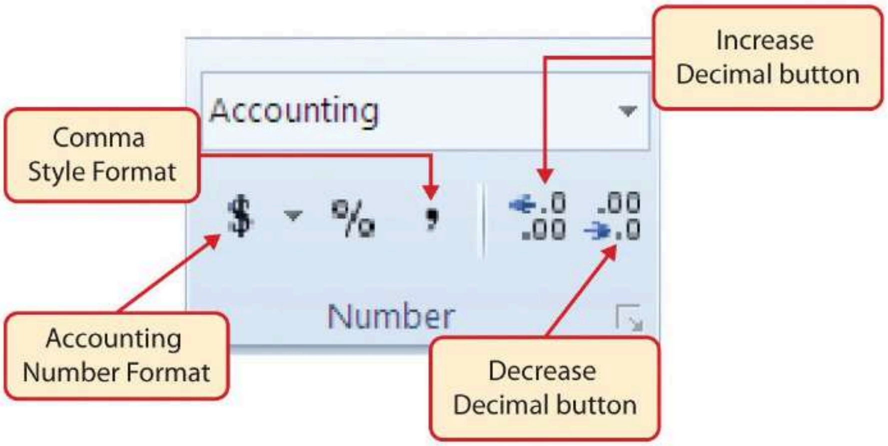


Figure 1.33 Number Group of Commands

3. Since the figures in this range do not include cents, click the Decrease Decimal button in the Number group of commands in the Home tab of the Ribbon two times (see **Figure 1.33**).
4. The numbers will also be reduced to zero decimal places.
5. Select the range C3:C14.
6. Click the Accounting Number Format button in the Number group of commands in the Home tab of the Ribbon (see **Figure 1.33**). This will add the US currency symbol and two decimal places to the values. This format is common when working with pricing data. As discussed above in the Formatting Data and Cells section, you will want to use Accounting format on all values in this range since the worksheet contains non-currency as well as currency data.
7. Select the range D3:D14.
8. Again, select the Accounting Number Format; this will add the US currency symbol to the values as well as two decimal places.
9. Click the Decrease Decimal button in the Number group of commands in the Home tab of the Ribbon.
10. This will add the US currency symbol to the values and reduce the decimal places to zero since there are no cents in these figures.
11. Select the range A1:D1.
12. Click the down arrow next to the Fill Color button in the Font group of commands in the Home tab of the Ribbon (see **Figure 1.34**). This will add background fill color the range for a worksheet title when entered.

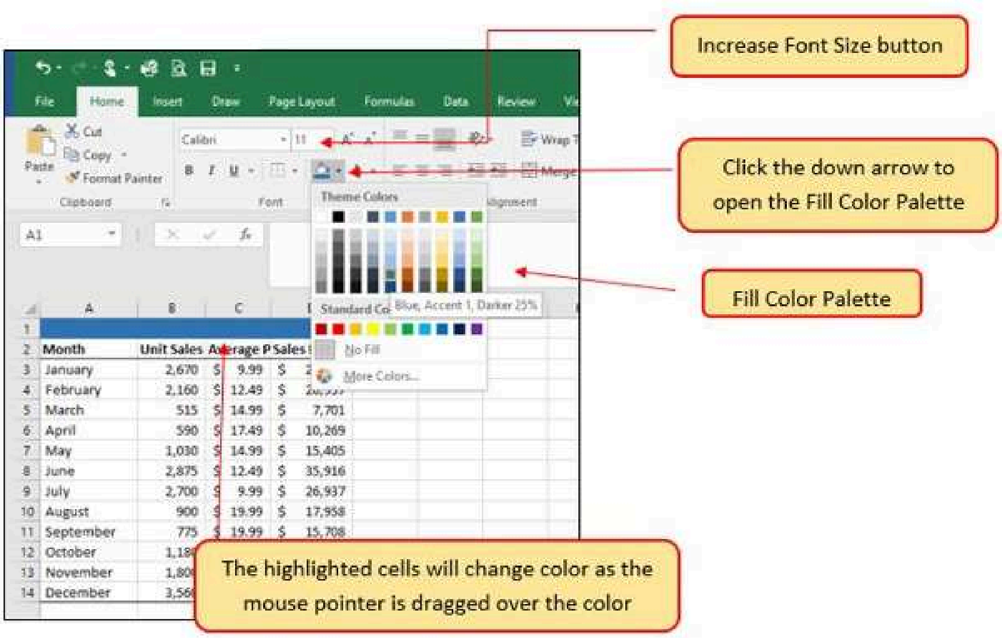


Figure 1.34 Fill Color Palette

13. Click the Blue, Accent 1, Darker 25% color from the palette (see **Figure 1.34**). Notice that as you move the mouse pointer over the color palette, you will see a preview of how the color will appear in the highlighted cells. Experiment with this feature.
14. Click on A1 and enter the worksheet title: **Merchandise City, USA** and click on the check mark in the formula bar to enter this information.
15. Since the black font is difficult to read on the blue background, you'll change the font color to be more visible. Click the down arrow next to the Font Color button in the Font group of commands in the Home tab of the Ribbon; select **White** as the font color for this range (see **Figure 1.32**).
16. Select the range A1:D1 and format for **Italics** by clicking on "I" in the Font group.
17. Click the drop-down arrow on the right side of the Font button in the Home tab of the Ribbon; select Arial as the font for this range and format for **Bold** click on "B**"** in the Font group. (see **Figure 1.32**).
18. Notice that as you move the mouse pointer over the font style options, you can see the font change in the highlighted cells.
19. Expand the column width of Column D to 14 characters.

Why?

Pound Signs (####) Appear in Columns

When a column is too narrow for a long number, Excel will automatically convert the number to a series of pound signs (####). In the case of words or text data, Excel will only show the characters that fit in the column. However, this is not the case with numeric data because it can give the appearance of a number that is much smaller than what is actually in the cell. To remove the pound signs, increase the width of the column.

Figure 1.35 shows how the Sheet1 worksheet should appear after the formatting techniques are applied.

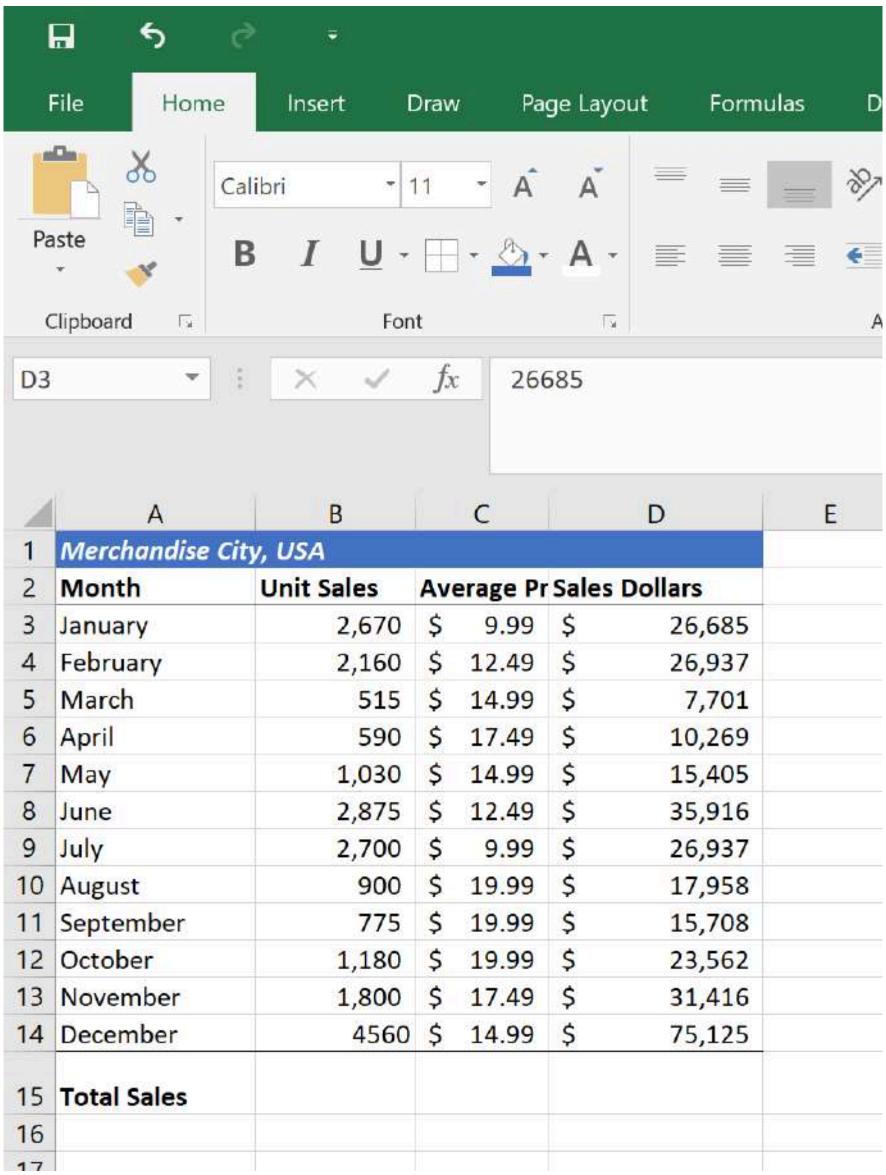


Figure 1.35 Formatting Techniques Applied

Data Alignment (Wrap Text, Merge Cells, and Center)

The skills presented in this segment show how data are aligned within cell locations. For example, text and numbers can be centered in a cell location, left justified, right justified, and so on. In some cases you may want to stack multiword text entries vertically in a cell instead of expanding the width of a column. This is referred to as wrapping text. These skills are demonstrated in the following steps:

1. Select the range A2:D2.
2. Click the Center button in the Alignment group of commands in the Home tab of the Ribbon (see **Figure 1.36**). This will center the column headings in each cell location.

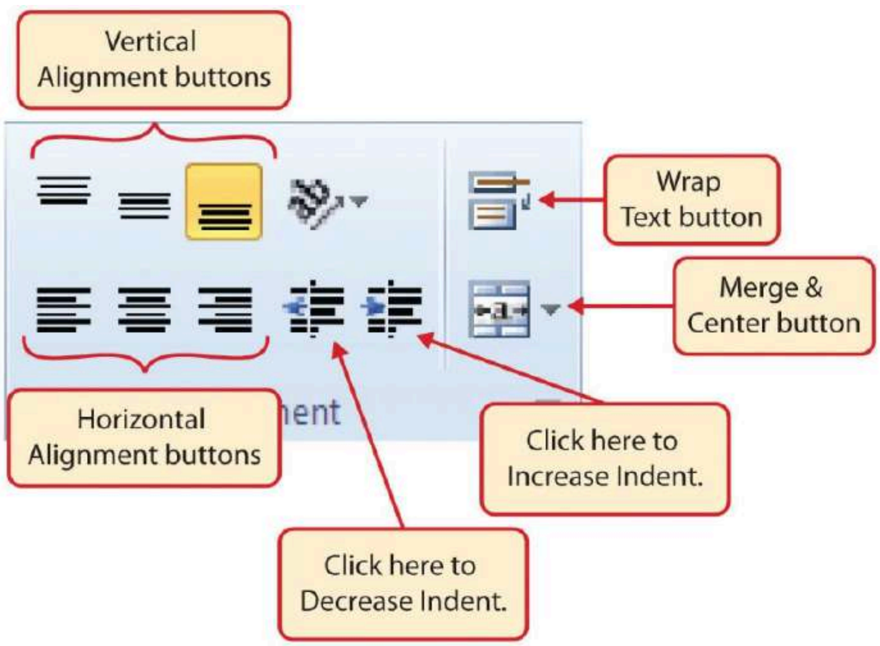


Figure 1.36 Alignment Group in Home Tab

3. Click the Wrap Text button in the Alignment group (see **Figure 1.36**). The height of Row 2 automatically expands, and the words that were cut off because the columns were too narrow are now stacked vertically.

Keyboard Shortcuts

Wrap Text

- Press the ALT key and then the letters H and W one at a time.

- There is no equivalent shortcut for Excel for Mac

Why?

Wrap Text

The benefit of using the Wrap Text command is that it significantly reduces the need to expand the column width to accommodate multiword column headings. The problem with increasing the column width is that you may reduce the amount of data that can fit on a piece of paper or one screen. This makes it cumbersome to analyze the data in the worksheet and could increase the time it takes to make a decision.

4. Select the range A1:D1.
5. Click the down arrow on the right side of the Merge & Center button in the Alignment group of commands in the Home tab of the Ribbon.
6. Click the Merge & Center option (see **Figure 1.37**). This will create one large cell location running across the top of the data set and center the text in that cell.

Keyboard Shortcuts

Merge Commands

- Merge & Center: Press the ALT key and then the letters H, M, and C one at a time.
- Merge Cells: Press the ALT key and then the letters H, M, and M one at a time.
- Unmerge Cells: Press the ALT key and then the letters H, M, and U one at a time.
    - There are no equivalent shortcuts for Excel for Mac

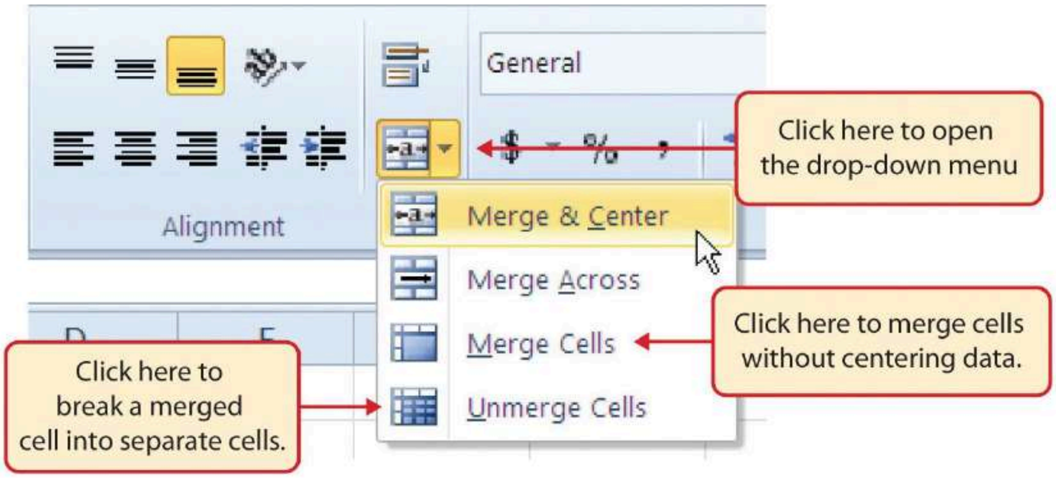


Figure 1.37 Merge Cell Drop-Down Menu

Why?

Merge & Center

One of the most common reasons the Merge & Center command is used is to center the title of a worksheet directly above the columns of data. Once the cells above the column headings are merged, a title can be centered above the columns of data. It is very difficult to center the title over the columns of data if the cells are not merged.

Figure 1.38 shows the Sheet1 worksheet with the data alignment commands applied. The reason for merging the cells in the range A1:D1 will become apparent in the next segment.

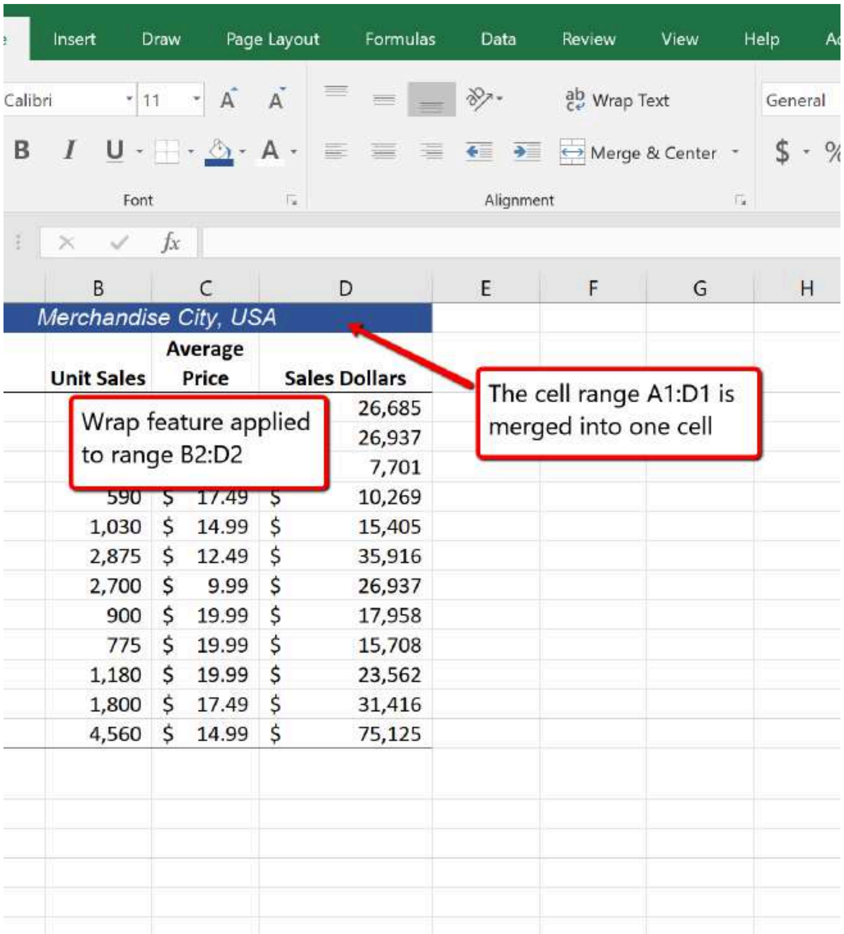


Figure 1.38 Data Alignment Features Added

Skill Refresher

Wrap Text

1. Activate the cell or range of cells that contain text data.
2. Click the Home tab of the Ribbon.
3. Click the Wrap Text button.

Skill Refresher

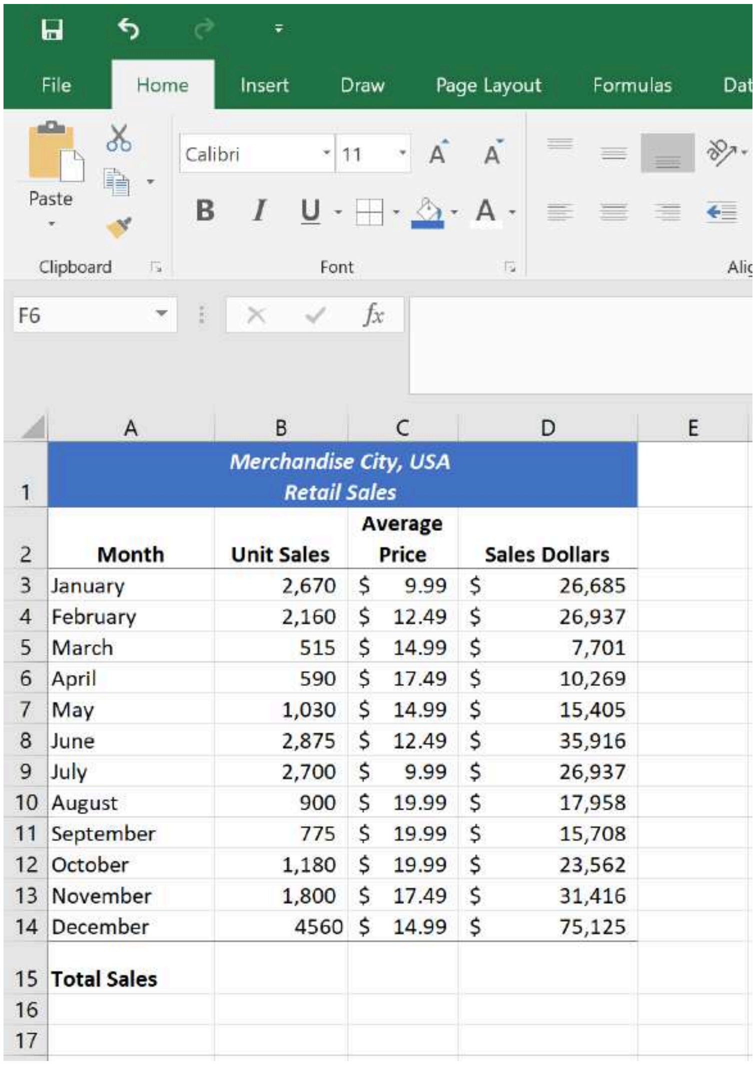
Merge Cells

1. Highlight a range of cells that will be merged.
2. Click the Home tab of the Ribbon.
3. Click the down arrow next to the Merge & Center button.
4. Select an option from the Merge & Center list.

Entering Multiple Lines of Text

In the Sheet1 worksheet, the cells in the range A1:D1 were merged for the purposes of adding a title to the worksheet. This worksheet will contain both a title and a subtitle. The following steps explain how you can enter text into a cell and determine where you want the second line of text to begin:

1. Click cell A1. Since the cells were merged, clicking cell A1 will automatically activate the range A1:D1. Position your mouse to the end of the title, directly after the “A” in the word “USA” and double-click to get a cursor (flashing I-beam).
2. Hold down the ALT key and press the ENTER key. This will start a new line of text in this cell location.
3. Type the text **Retail Sales** and press the ENTER key.
4. Select cell A1. Then click the Bold buttons in the Font group of commands in the Home tab of the Ribbon so that the titles are now in Bold and Italics.
5. Increase the height of Row 1 to 30 points. Once the row height is increased, all the text typed into the cell will be visible (see **Figure 1.39**).



	A	B	C	D	E
1	*Merchandise City, USA Retail Sales*				
2	**Month**	**Unit Sales**	**Average Price**	**Sales Dollars**	
3	January	2,670	$ 9.99	$ 26,685	
4	February	2,160	$ 12.49	$ 26,937	
5	March	515	$ 14.99	$ 7,701	
6	April	590	$ 17.49	$ 10,269	
7	May	1,030	$ 14.99	$ 15,405	
8	June	2,875	$ 12.49	$ 35,916	
9	July	2,700	$ 9.99	$ 26,937	
10	August	900	$ 19.99	$ 17,958	
11	September	775	$ 19.99	$ 15,708	
12	October	1,180	$ 19.99	$ 23,562	
13	November	1,800	$ 17.49	$ 31,416	
14	December	4560	$ 14.99	$ 75,125	
15	**Total Sales**				
16					
17					

Figure 1.39 Title & Subtitle Added to the Worksheet

Skill Refresher

Entering Multiple Lines of Text

1. Activate a cell location.
2. Type the first line of text.
3. Hold down the ALT key and press the ENTER key.
4. Type the second line of text and press the ENTER key.

Borders (Adding Lines to a Worksheet)

In Excel, adding custom lines to a worksheet is known as adding borders. Borders are different from the grid lines that appear on a worksheet and that define the perimeter of the cell locations. The Borders command lets you add a variety of line styles to a worksheet that can make reading the worksheet much easier. The following steps illustrate methods for adding preset borders and custom borders to a worksheet:

1. Click the down arrow to the right of the Borders button in the Font group of commands in the Home page of the Ribbon to view border options. (see **Figure 1.40**).

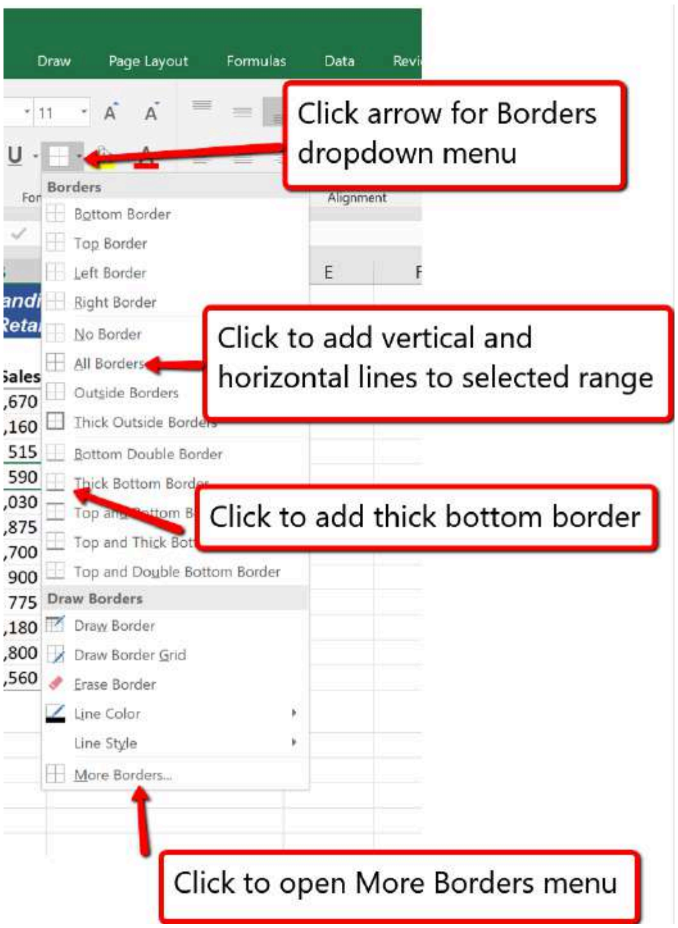


Figure 1.40 Borders Dropdown Menu

2. Select the range A1:D15. Left click the All Borders option from the Borders drop-down menu (see **Figure 1.40**). This will add vertical and horizontal lines to the range A1:D15.
3. Select the range A2:D2 .
4. Click the down arrow to the right of the Borders button.
5. Left click the Thick Bottom Border option from the Borders drop-down menu.
6. Select the range A14:D14 and apply a Thick Bottom Border from the drop-down menu. The thick border will help maintain the Excel Formatting Guidelines.

7. Select the range A1:D15.
8. Click the down arrow to the right of the Borders button.
9. Click More Borders… at the bottom of the List.
10. This will open the Format Cells dialog box (see **Figure 1.41**). You can access all formatting commands in Excel through this dialog box.
11. In the Style section of the Borders tab, click the thickest line style (see **Figure 1.41**).
12. Click the Outline button in the Presets section (see **Figure 1.41**).
13. Click the OK button at the bottom of the dialog box (see **Figure 1.41**).

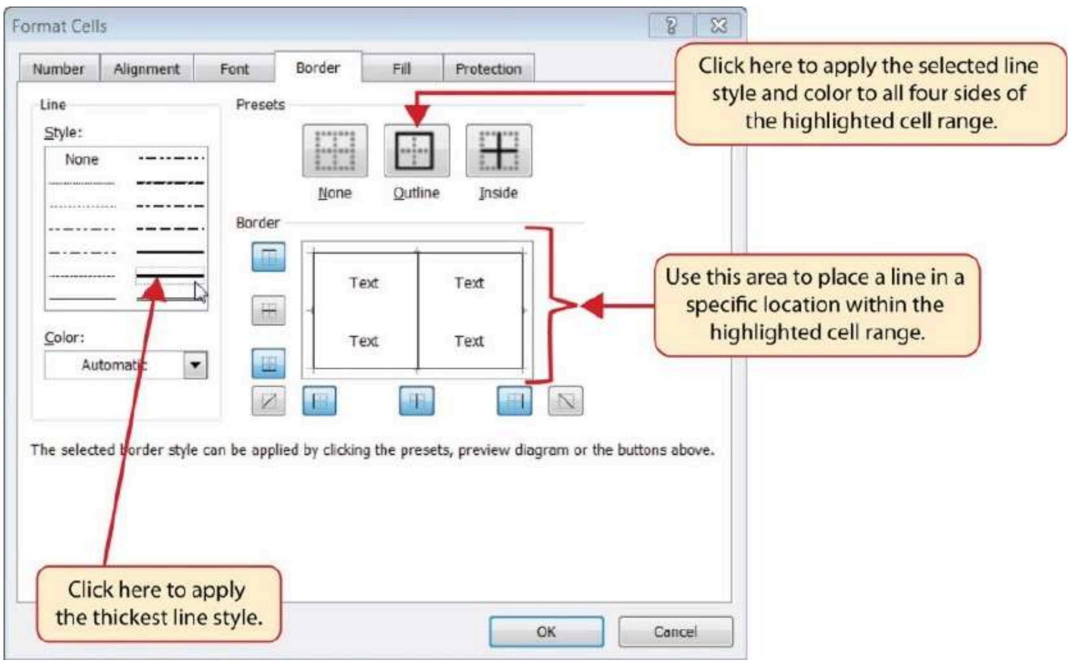


Figure 1.41 Borders Tab of the Format Cells Dialog Box

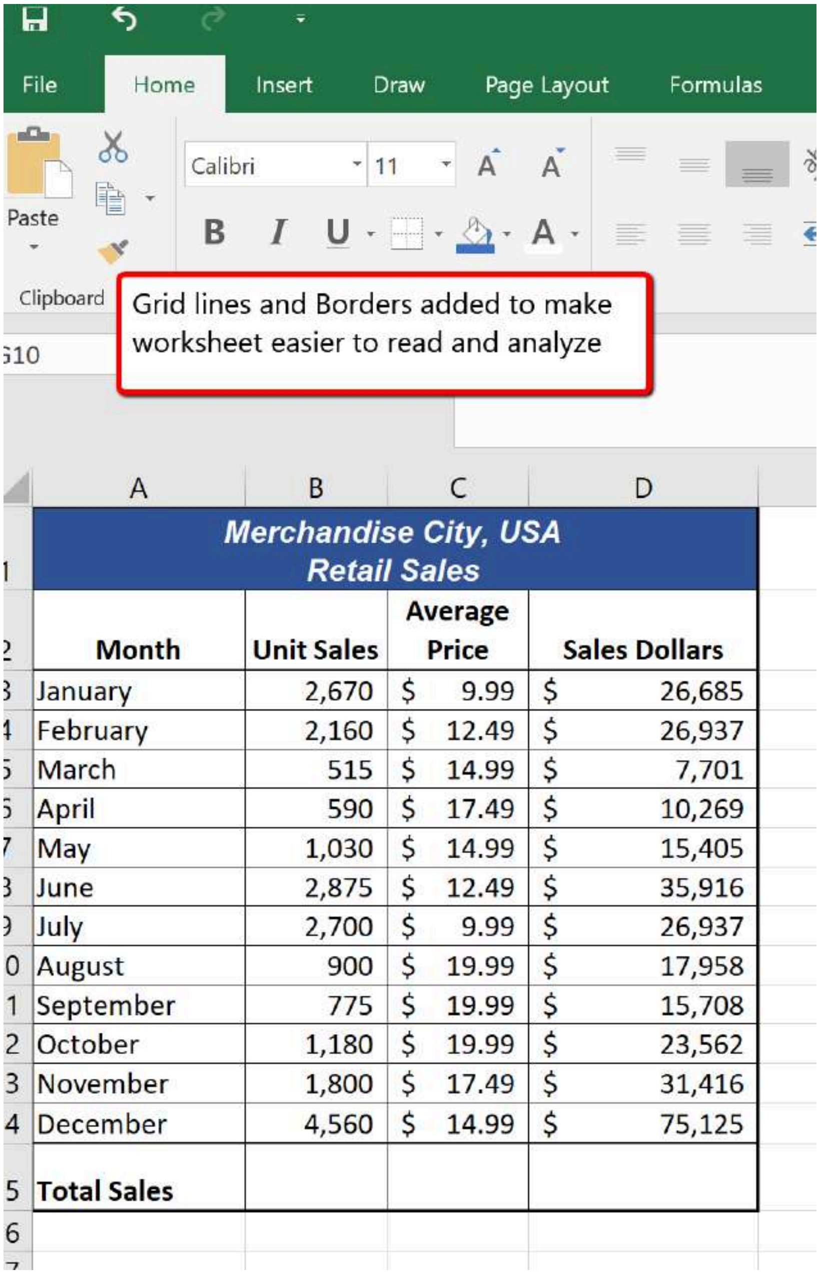


Figure 1.42 Borders Added to the Worksheet

Skill Refresher

Preset Borders

1. Highlight a range of cells that require borders.
2. Click the Home tab of the Ribbon.
3. Click the down arrow next to the Borders button.
4. Select an option from the preset borders list.

Custom Borders

1. Highlight a range of cells that require borders.
2. Click the Home tab of the Ribbon.
3. Click the down arrow next to the Borders button.
4. Select the More Borders option at the bottom of the options list.
5. Select a line style and line color.
6. Select a placement option.
7. Click the OK button on the dialog box.

AutoSum

You will see at the bottom of Figure 1.42 that Row 15 is intended to show the totals for the data in this worksheet. Applying mathematical computations to a range of cells is accomplished through functions in Excel. Chapter 2 will review mathematical formulas and functions in detail. However, the following steps will demonstrate how you can quickly sum the values in a column of data using the AutoSum command:

1. Click cell B15 in the Sheet1 worksheet.
2. Click the Formulas tab of the Ribbon.
3. Click the down arrow below the AutoSum button in the Function Library group of commands (see **Figure 1.43**). ***Note that the AutoSum button can also be found in the Editing group of commands in the Home tab of the Ribbon.***

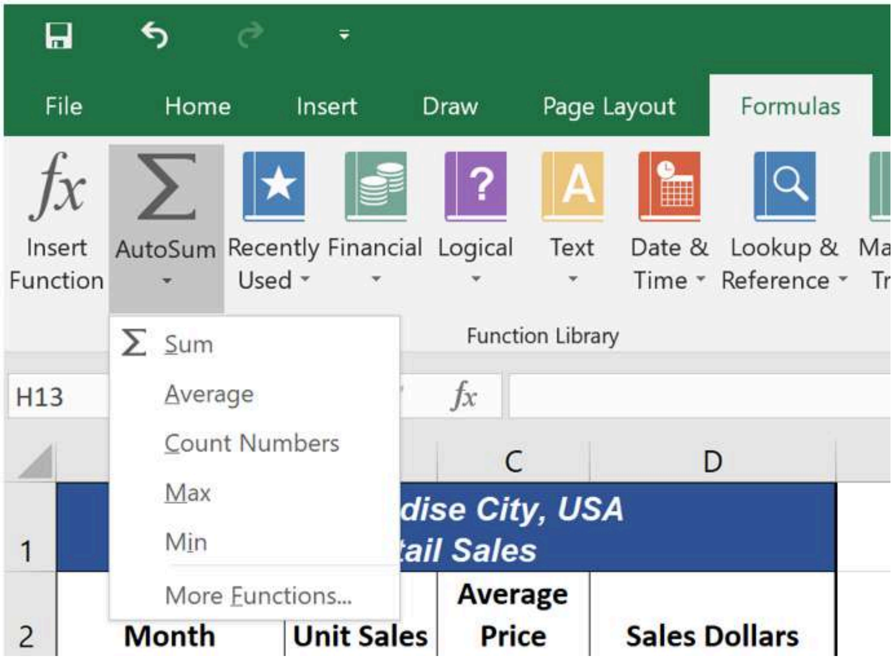


Figure 1.43 AutoSum List

4. Click the Sum option from the AutoSum drop down menu. The first click will display a flashing marquee around the range. Click the check mark next to the Formula bar to complete the function.
5. Excel will provide a total for the values in the Unit Sales column.
6. Click cell D15. It would not make sense to total the averages in column C so C15 will be left blank.
7. Repeat steps 3 through 5 to sum the values in the Sales Dollars column (see **Figure 1.44**).
8. Click cell C15 to explore other AutoSum selections. Select the COUNT function from the list; Excel will return "12" for the number of months (rows). Excel will also display indicators of a green arrow in the corner of C15 and an exclamation point in yellow. These indicate that the function in this cell varies from the other functions in row 15. They can be ignored and do not print.
9. Click cell C15 again; this time selecting the MAX option from the list. Excel will display $19.99. This reflects the Maximum Average Price in column C.
10. Click cell C15 and delete the contents in this cell.

Merchandise City, USA
Retail Sales

Month	Unit Sales	Average Price	Sales Dollars
January	2,670	$ 9.99	$ 26,685
February	2,160	$ 12.49	$ 26,937
March	515	$ 14.99	$ 7,701
April	590	$ 17.49	$ 10,269
May	1,030	$ 14.99	$ 15,405
June	2,875	$ 12.49	$ 35,916
July	2,700	$ 9.99	$ 26,937
August	900	$ 19.99	$ 17,958
September	775	$ 19.99	$ 15,708
October	1,180	$ 19.99	$ 23,562
November	1,800	$ 17.49	$ 31,416
December	4,560	$ 14.99	$ 75,125
Total Sales	21,755		$ 313,619

Figure 1.44 Totals Added to the Sheet1 Worksheet

Skill Refresher

AutoSum

1. Highlight a cell location below or to the right of a range of cells that contain numeric values.
2. Click the Formulas tab of the Ribbon.
3. Click the down arrow below the AutoSum button.
4. Select a mathematical function from the list.

Moving, Renaming, Inserting, and Deleting Worksheets

The default names for the worksheet tabs at the bottom of workbook are Sheet1, Sheet2, and so on. However, you can change the worksheet tab names to identify the data you are using in a workbook. Additionally, you can change the order in which the worksheet tabs appear in the workbook. The following steps explain how to rename and move the worksheets in a workbook:

1. Double click the Sheet1 worksheet tab at the bottom of the workbook (see **Figure 1.45**). Type the name **Sales by Month**.
2. Press the ENTER key on your keyboard.
3. Click the + to the right of the newly named worksheet.
4. Type the name **Unit Sales Rank** to prepare the worksheet for future use.
5. Press the ENTER key on your keyboard.

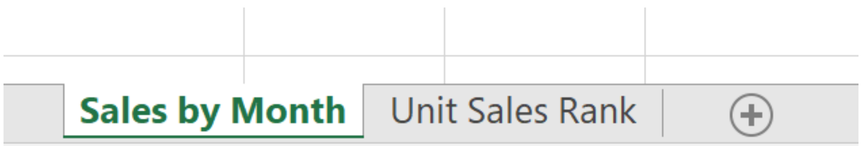


Figure 1.45 Renaming a Worksheet Tab

1. Click the + to add another worksheet tab.
2. Click the Home tab of the Ribbon.
3. Click the down arrow on the Delete button in the Cells group of commands.
4. Click the Delete Sheet option from the drop-down list. This removes the unneeded worksheet.
5. Click the Delete button on the Delete warning box (if a warning box appears).
6. Complete the steps above to delete the newly named **Unit Sales Rank** worksheet since it's decided that worksheet is also unnecessary so that you are left with just one worksheet.
7. Excel incorporates **Spell Check** which is located on the **Review** Ribbon. Clicking on the tool will allow Excel to check Spelling of alphabetic entries and allow for corrections. It's a good habit to always use Spell Check your work before saving/printing.

Figure 1.45a Spell Check Tool

8. Save the changes to your workbook by clicking either the **Save** button on the Home ribbon; or by selecting the **Save** option from the File menu.

Integrity Check

Deleting Worksheets

Be very cautious when deleting worksheets that contain data. Once a worksheet is deleted, you cannot use the Undo command to bring the sheet back. Deleting a worksheet is a permanent command.

Keyboard Shortcuts

Inserting New Worksheets

- Press the SHIFT key and then the F11 key on your keyboard. Same for Excel for Mac.

Figure 1.46 shows the final appearance of the Merchandise City, USA workbook.

Merchandise City, USA Retail Sales			
Month	**Unit Sales**	**Average Price**	**Sales Dollars**
January	2,670	$ 9.99	$ 26,685
February	2,160	$ 12.49	$ 26,937
March	515	$ 14.99	$ 7,701
April	590	$ 17.49	$ 10,269
May	1,030	$ 14.99	$ 15,405
June	2,875	$ 12.49	$ 35,916
July	2,700	$ 9.99	$ 26,937
August	900	$ 19.99	$ 17,958
September	775	$ 19.99	$ 15,708
October	1,180	$ 19.99	$ 23,562
November	1,800	$ 17.49	$ 31,416
December	4,560	$ 14.99	$ 75,125
Total Sales	21,755		$ 313,619

Figure 1.46 Final Appearance of the Merchandise City, USA Workbook

Skill Refresher

Renaming Worksheets

1. Double click the worksheet tab.
2. Type the new name.
3. Press the ENTER key.

Moving Worksheets

1. Left click the worksheet tab.
2. Drag it to the desired position.

Deleting Worksheets

1. Open the worksheet to be deleted.
2. Click the Home tab of the Ribbon.
3. Click the down arrow on the Delete button.
4. Select the Delete Sheet option.
5. Click Delete on the warning box.

Key Takeaways

- Formatting skills are critical for creating worksheets that are easy to read and have a professional appearance.
- A series of pound signs (####) in a cell location indicates that the column is too narrow to display the number entered.
- Using the Wrap Text command allows you to stack multiword column headings vertically in a cell location, reducing the need to expand column widths.
- Use the Merge & Center command to center the title of a worksheet directly over the columns that contain data.
- Adding borders or lines will make your worksheet easier to read and helps to separate the data in each column and row.
- You cannot use the Undo command to bring back a worksheet that has been deleted.

Attribution

Adapted by Barbara Lave from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

2.XLSX.4 PRINTING

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Learning Objectives

- Use the Page Layout tab to prepare a worksheet for printing.
- Add headers and footers to a printed worksheet.
- Examine how to print worksheets and workbooks.

Once you have completed a workbook, it is good practice to select the appropriate settings for printing. These settings are in the Page Layout tab of the Ribbon and discussed in this section of the chapter.

Page Setup

Before you can properly print the worksheets in a workbook, you must establish appropriate settings. The following steps explain several of the commands in the Page Layout tab of the Ribbon used to prepare a worksheet for printing:

1. Open the **CH1 Merchandise City Sales Data** workbook, if it is not already open.
2. Click the Page Layout tab of the Ribbon.
3. Click the Margins button in the Page Setup group of commands. This will open a drop-down list of options for setting the margins of your printed document.
4. Click the Wide option from the Margins drop-down list. (see **Figure 1.47**)
5. Click the Orientation button in the Page Setup and select Landscape.
6. Click on the arrow to the bottom right of the Page Setup category to launch the Page Setup options dialog box. Mac Users: there is no "arrow at the bottom right of the Page Setup category". Simply click the Page Setup button 

on the Page Layout tab.
7. Click the Margins tab and locate "Center on Page". Click the boxes to **Horizontally** and **Vertically**

center the data on the worksheet. Click OK.

Why?

Use Print Settings

Because professionals often share Excel workbooks, it is a good practice to select the appropriate print settings in the Page Layout tab even if you do not intend to print the worksheets in a workbook. It can be extremely frustrating for recipients of a workbook who wish to print your worksheets to find that the necessary print settings have not been selected. This may reflect poorly on your attention to detail, especially if the recipient of the workbook is your boss.

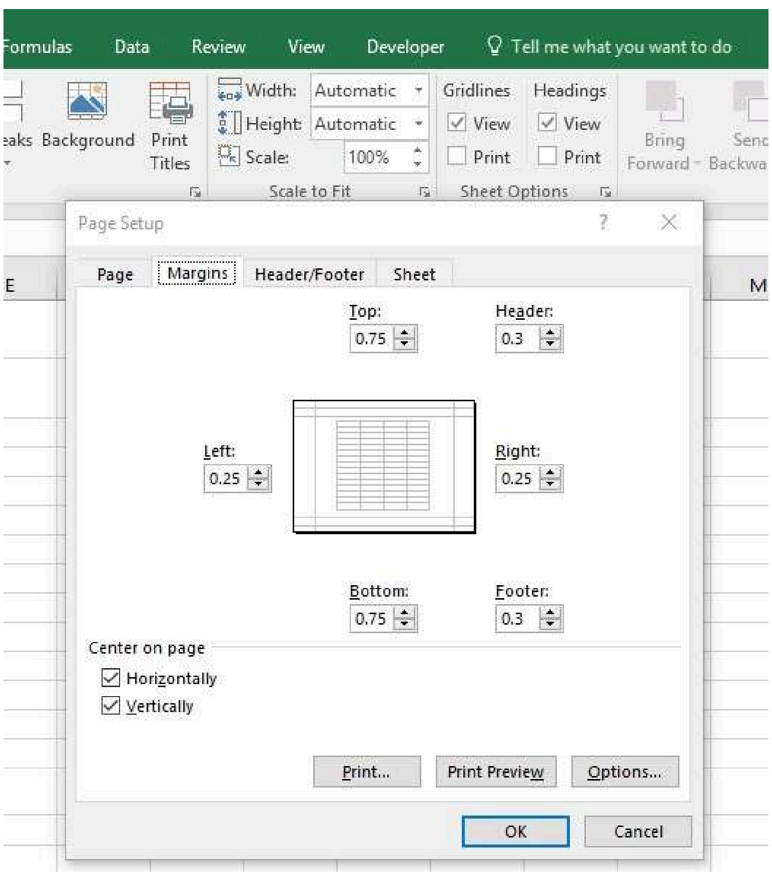


Figure 1.47 Page Layout Commands for Printing

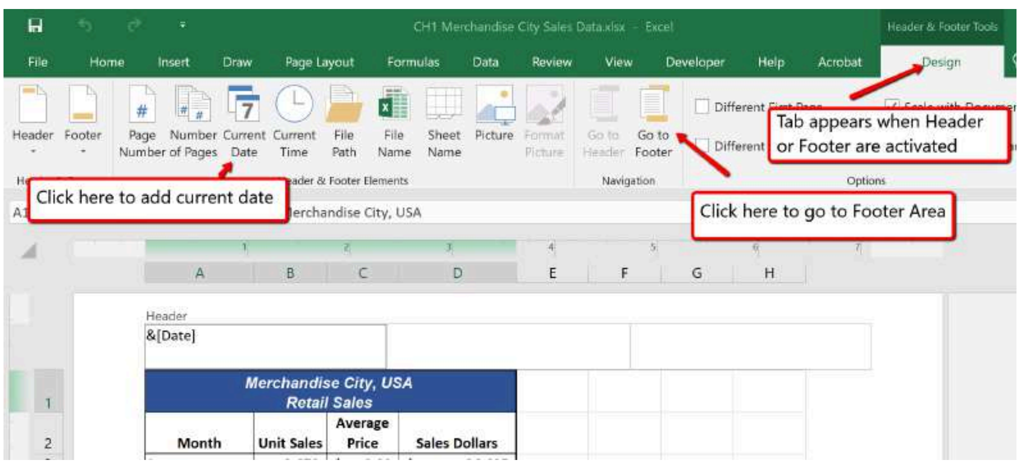
Table 1.2 Printing Resources: Purpose and Use for Page Setup Commands

Command	Purpose	Use
Margins	Sets the top, bottom, right, and left margin space for the printed document	1. Click the Page Layout tab of the Ribbon. 2. Click the Margin button. 3. Click one of the preset margin options or click Custom Margins.
Orientation	Sets the orientation of the printed document to either portrait or landscape	1. Click the Page Layout tab of the Ribbon. 2. Click the Orientation button. 3. Click one of the preset orientation options.
Size	Sets the paper size for the printed document	1. Click the Page Layout tab of the Ribbon. 2. Click the Size button. 3. Click one of the preset paper size options or click More Paper Sizes.
Print Area	Used for printing only a specific area or range of cells on a worksheet	1. Highlight the range of cells on a worksheet that you wish to print. 2. Click the Page Layout tab of the Ribbon. 3. Click the Print Area button. 4. Click the Set Print Area option from the drop-down list.
Breaks	Allows you to manually set the page breaks on a worksheet	1. Activate a cell on the worksheet where the page break should be placed. Breaks are created above and to the left of the activated cell. 2. Click the Page Layout tab of the Ribbon. 3. Click the Breaks button. 4. Click the Insert Page Break option from the drop-down list.
Background	Adds a picture behind the cell locations in a worksheet	1. Click the Page Layout tab of the Ribbon. 2. Click the Background button. 3. Select a picture stored on your computer or network.
Print Titles	Used when printing large data sets that are several pages long. This command will repeat the column headings at the top of each printed page.	1. Click the Page Layout tab of the Ribbon. 2. Click the Print Titles button. 3. Click in the Rows to Repeat at Top input box in the Page Setup dialog box. 4. Click any cell in the row that contains the column headings for your worksheet. 5. Click the OK button at the bottom of the Page Setup dialog box.

Headers and Footers

When printing worksheets from Excel, it is common to add headers and footers to the printed document. Information in the header or footer could include the date, page number, file name, company name, and so on. The following steps explain how to add headers and footers to the Merchandise City, USA Retail Sales worksheet.

1. Click the **Insert** Ribbon and click on **Header & Footer** at the right end of the ribbon (located in the **Text** group). You will see the Design tab added to the Ribbon; this is used for creating the headers and footers for the printed worksheet. Also, this will convert the view of the worksheet from Normal to Page Layout (see **Figure 1.48**). This Page Layout view makes adding Headers & Footers easy and provides key features to incorporate.
2. Click on the Current Date icon to add the date to the left section of the worksheet Header. The &[Date] symbols which will toggle to a Date format when you click outside of this area.



1.48 Design Tab for Creating Headers and Footers

Figure 1.48 Design Tab for Creating Headers and Footers

3. Type your name in the center section of the Header.
4. Place the mouse pointer over the left section of the Header and left click (see **Figure 1.48**).
5. Click the Go to Footer button in the Navigation group of commands in the Design tab of the Ribbon.
6. Place the mouse pointer over the far right section of the footer and left click.
7. Click the Page Number button (you may need to click on the Design tab again) in the Header & Footer Elements group of commands in the Design tab of the Ribbon. This view will display as &[Page] until printed or until you return to normal view.
8. Click any cell location outside the header or footer area. The Design tab for creating headers and footers will disappear.
9. Click the Normal view button in the lower right side of the Status Bar (see **Figure 1.49**).

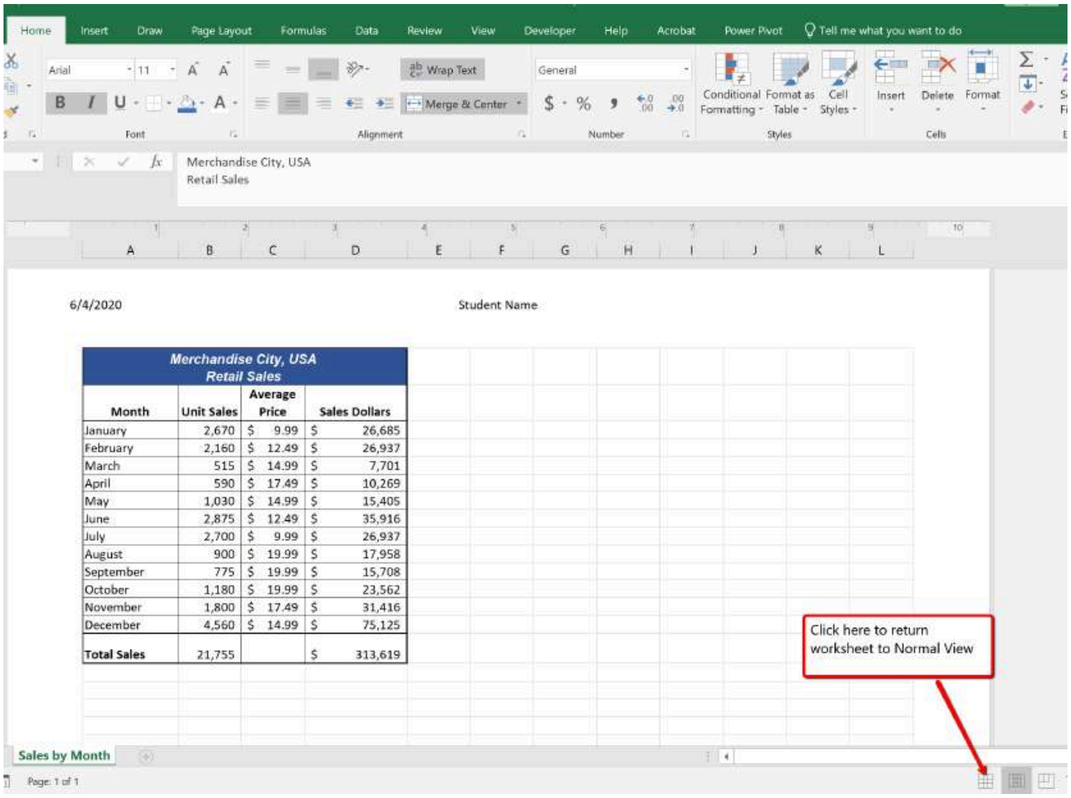


Figure 1.49 Worksheet in Page Layout View

Printing Worksheets and Workbooks

Once you have established the print settings for the worksheets in a workbook and have added headers and footers, you are ready to print your worksheets. The following steps explain how to print the worksheets in the **Merchandise City, USA Sales** workbook:

1. Click the File tab on the Ribbon.
2. Click the Print option on the left side of the Backstage view (see **Figure 1.50**). On the right side of the Backstage view, you will be able to see a preview of your printed worksheet.

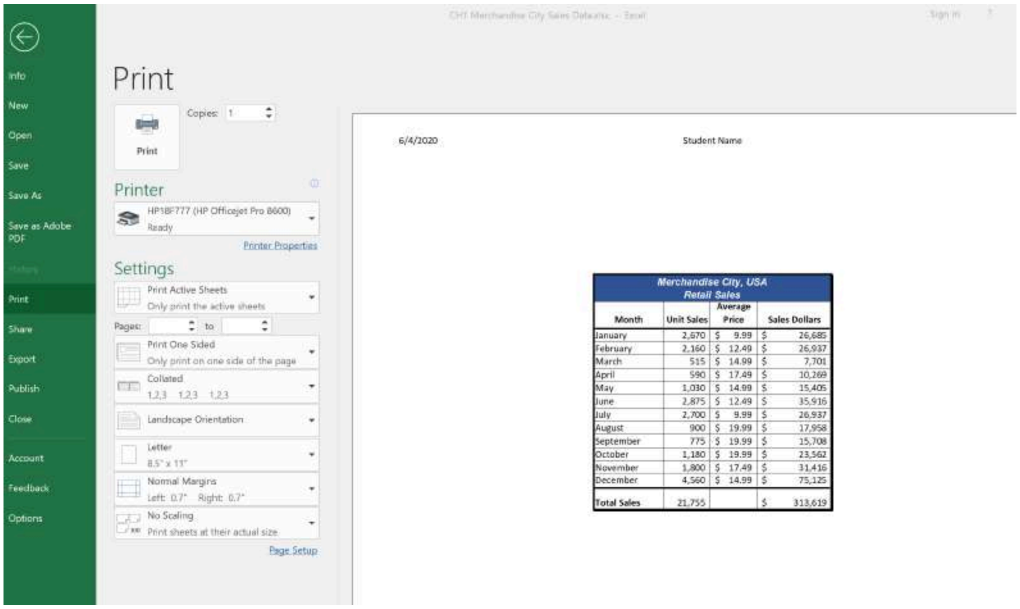


Figure 1.50 Backstage View Print option

3. Click the Print Active Sheets button in the Print section of the Backstage view (see **Figure 1.50**).
4. If your instructor has asked you to print your work, click the Print button.
5. Click the Home tab of the Ribbon.
6. Save and close the workbook.

Key Takeaways

- The commands in the Page Layout tab of the Ribbon are used to prepare a worksheet for printing.
- You can add headers and footers to a worksheet to show key information such as page numbers, the date, the file name, your name, and so on.
- The Print commands are in the File tab of the Ribbon.

Attribution

Adapted by Barbara Lave from [How to Use Microsoft Excel: The Careers in Practice Series](), adapted by [The Saylor Foundation]() without attribution as requested by the work's original creator or licensee, and licensed under [CC BY-NC-SA 3.0]().

2.XLSX.5 CHAPTER PRACTICE

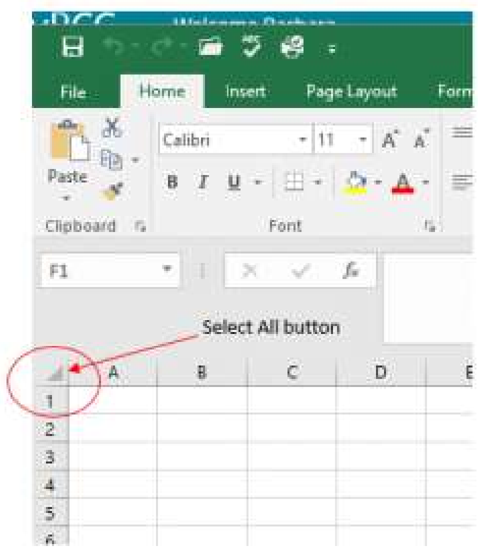
Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

To assess your understanding of the material covered in the chapter, complete the following assignment.

Basic Monthly Budget for Medical Office

Download Data File: PR1 Data

Creating and maintaining budgets are common practices in many careers. Budgets play a critical role in helping a business or household control expenditures. In this exercise you will create a budget for a hypothetical medical office while reviewing the skills covered in this chapter.

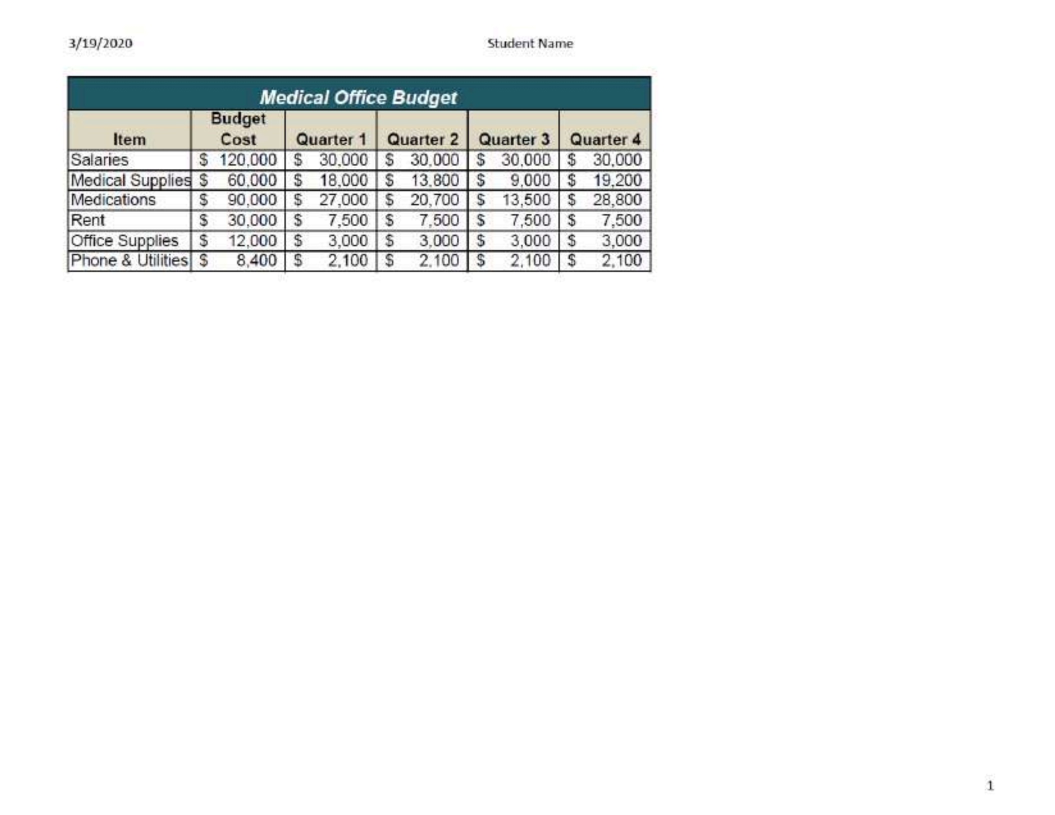
1. Open the file name **PR1 Data**, then Save As **PR1 Medical Office Budget**.
2. Activate all the cell locations in the Sheet1 worksheet by clicking the Select All button in the upper left 

corner of the worksheet.
3. In the Home tab of the Ribbon, set the font style to Arial and the font size to 12 points. Then click any cell to Deselect.
4. Increase the width of Column A so all the entries in the range A3:A8 are visible. Place the mouse pointer between the letter A and letter B of Column A and Column B. When the mouse pointer changes to a double arrow, left click and drag it to the right until the character width is approximately 18.00.
5. Enter **Quarter 1** in cell B2.

6. Use AutoFill to complete the headings in the range C2:E2. Activate cell B2 and place the mouse pointer over the Fill Handle.
7. Select the range B2:E2 and click the Format button in the Home tab of the Ribbon. Click the Column Width option, type 11.57 in the Column Width dialog box, and then click the OK button in the Column Width dialog box.
8. Enter the words **Medical Office Budget** in cell A1.
9. Insert a blank column between Columns A and B by clicking on any cell in Column B. Then, click the drop-down arrow of the Insert button in the Home tab of the Ribbon. Click the Insert Sheet Columns option.
10. Enter the words **Budget Cost** in cell B2.
11. Adjust the width of Column B to approximately 12.0 characters.
12. Select the range A1:F1 and click the Merge & Center button in the Home tab of the Ribbon to merge the cells in that range.
13. Make the following format adjustments to the range A1:F1: bold; italics; change the font size to 14 points; change the cell fill color to Aqua, Accent 5, Darker 50%; and change the font color to white.
14. Increase the height of Row 1 to approximately 24.75 points.
15. Make the following format adjustment to the range A2:F2: bold; and fill color to Tan, Background 2, Darker 10%. Center the column titles so that they are horizontally centered in each cell.
16. Select B2 and choose the Wrap Text button in the Home tab of the Ribbon. Increase the height of Row 2 to approximately 30 points.
17. Copy cell C3 and paste the contents into the range D3:F3.
18. Copy the contents in the range C6:C8 by highlighting the range and clicking the Copy button in the Home tab of the Ribbon. Then, highlight the range D6:F8 and click the Paste button in the Home tab of the Ribbon.
19. Calculate the total budget for all four quarters for the salaries. Click cell B3 and click the down arrow on the AutoSum button in the Formulas tab of the Ribbon. Click the Sum option from the drop-down list. Then, highlight the range C3:F3 and press the ENTER key on your keyboard.
20. Copy the formula in cell B3 and paste them into the range B4:B8.
21. Format the range B3:F8 with Accounting format and zero decimal places.
    If any of the cells display pound symbols (######), simply widen the column to display the values again.
22. Select the range A1:F8 and click the down arrow next to the Borders button in the Home tab of the Ribbon. Select the All Borders option from the drop-down list.
23. Double click the Sheet1 worksheet tab to change the name of Sheet1 to the word **Budget**, and press the ENTER key. Delete any unnecessary worksheets.
24. Change the orientation of the Budget worksheet so it prints landscape instead of portrait.
25. Use Fit to 1 page so the Budget worksheet prints on one piece of paper, if it does not already.
26. Add a header to the Budget worksheet that shows the date in the upper left corner and your name in the

center.

27. Add a footer to the Budget worksheet that shows the page number in the lower right corner.
28. Check the spelling on the worksheet and make any necessary changes. Save **PR1 Medical Office Budget** workbook.
29. Compare your work to the screenshot below and then submit the **PR1 Medical Office Budget** workbook as directed by your instructor.



3/19/2020 Student Name

Medical Office Budget					
Item	Budget Cost	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Salaries	$ 120,000	$ 30,000	$ 30,000	$ 30,000	$ 30,000
Medical Supplies	$ 60,000	$ 18,000	$ 13,800	$ 9,000	$ 19,200
Medications	$ 90,000	$ 27,000	$ 20,700	$ 13,500	$ 28,800
Rent	$ 30,000	$ 7,500	$ 7,500	$ 7,500	$ 7,500
Office Supplies	$ 12,000	$ 3,000	$ 3,000	$ 3,000	$ 3,000
Phone & Utilities	$ 8,400	$ 2,100	$ 2,100	$ 2,100	$ 2,100

1

PR 1 Medical Office Budget Solution Screenshot

Attribution

Adapted by Barbara Lave from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

2.XLSX.6 SCORED ASSESSMENT

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Sales and Inventory Items

Download Data File: SC1 Data

A key activity for marketing professionals is to analyze projected sales and inventory information. This is especially important for retail environments. This exercise utilizes the skills covered in this chapter to analyze sales and inventory data.

1. Open the file named **SC1 Data** and then Save As **SC1 Sales and Inventory**
2. In the Sheet1 worksheet, enter the word Totals in cell C14.
3. Format all the cells in Sheet1 to Century font style and a 12-point font size.
4. Set the column width for Columns A through G to 13.5.
5. Edit the entry in cell B2 to read "**Item Number**."
6. Use AutoFill to fill the Item Numbers from B3 into the range B4:B13. The item numbers should increase by one as they are filled through the range.
7. Copy the contents of cell A3 and paste them into the range A4:A8.
8. Delete Column F.
9. Format the range A1:F2 so the text is Bold.
10. Set the alignment in the range A2:F2 to Wrap Text.
11. Prepare A1:F1 for the title text by changing the fill color of the cells in the range A1:F1 to Red, Accent 2, Darker 25%.
12. Make the following font changes to the range A1:F1: set the font color to white, add italics, and set the font size to 14.
13. Merge and center the cells in the range A1:F1.
14. Enter the title for this worksheet in the range A1:F1. The title should appear on two lines. The first line should read **Status Report**. The second line should read **Sales and Inventory by Item**.
15. Increase the height of Row 1 so the entire title is visible.
16. Format the values in the range C3:C13 with dollar signs and two decimal places.
17. Format the values in the range E3:F13 with comma style, zero decimal places.
18. In cell E14, use AutoSum to calculate the sum of the values in the range E3:E13.
19. In cell F14, use AutoSum to calculate the sum of the values in the range F3:F13.

20. Apply All Borders to the range A1:F14.
21. Add a thick bottom border to row 2; add a thick bottom border to row 13.
22. Add a thick line border around the perimeter of the range A1:F14.
23. Insert a new blank worksheet in the workbook (this will be Sheet4).
24. Delete Sheet3.
25. Move Sheet4 ahead of Sheet2 so the order of the worksheets is Sheet1, Sheet4, and Sheet2.
26. Rename the Sheet1 worksheet tab to "**Status Report**."
27. Change the orientation of the Status Report worksheet so it prints landscape instead of portrait.
28. Add a header to the Status Report worksheet that shows the date (needs to update) in the upper left corner and your name in the center.
29. Add a footer to the Status Report worksheet that shows the page number in the lower right corner with the word "**Page**" before the number.
30. Center the worksheet both horizontally and vertically on the sheet.
31. Check the spelling of the worksheet and make any necessary changes. Save the **SC1 Sales and Inventory** workbook.
32. Submit the **SC1 Sales and Inventory** workbook as directed by your instructor.

Attribution

Adapted by Barbara Lave from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

SECTION IV

3. VISUALIZING DATA

3.1 THE HISTOGRAM

3.1: The Histogram

3.1.1: Cross Tabulation

Cross tabulation (or crosstabs for short) is a statistical process that summarizes categorical data to create a contingency table.

Learning Objective

Demonstrate how cross tabulation provides a basic picture of the interrelation between two variables and helps to find interactions between them.

Key Takeaways

Key Points

- Crosstabs are heavily used in survey research, business intelligence, engineering, and scientific research.
- Crosstabs provide a basic picture of the interrelation between two variables and can help find interactions between them.
- Most general-purpose statistical software programs are able to produce simple crosstabs.

Key Term

cross tabulation
: a presentation of data in a tabular form to aid in identifying a relationship between variables

Cross tabulation (or crosstabs for short) is a statistical process that summarizes categorical data to create a contingency table. It is used heavily in survey research, business intelligence, engineering, and scientific research. Moreover, it provides a basic picture of the interrelation between two variables and can help find interactions between them.

In survey research (e.g., polling, market research), a "crosstab" is any table showing summary statistics. Commonly, crosstabs in survey research are combinations of multiple different tables. For example, combines multiple contingency tables and tables of averages.

		BANNER							
		AGE				GENDER			
Q9. Coca Cola		Under 30	35 to 49	50 or more	NET	Male	Female	NET	TOTAL
Once every 3 months or less	% within column	21% ↓	32%	38%	30%	23% ↓	37% ↑	30%	**30%**
	% within row	24% ↓	32%	44%	100%	37% ↓	63% ↑	100%	**100%**
	Count	24	31	43	98	36	62	98	**98**
Once a month	% within column	12%	14%	14%	13%	16%	11%	13%	**13%**
	% within row	32%	32%	36%	100%	59%	41%	100%	**100%**
	Count	14	14	16	44	26	18	44	**44**
Once every 2 weeks	% within column	11%	7%	9%	9%	10%	8%	9%	**9%**
	% within row	43%	23%	33%	100%	53%	47%	100%	**100%**
	Count	13	7	10	30	16	14	30	**30**
LESS THAN WEEKLY	% within column	44%	53%	61%	53%	49%	58%	53%	**53%**
	% within row	30%	30%	40%	100%	45%	55%	100%	**100%**
	Count	51	52	69	172	78	94	172	**172**
Once a week	% within column	15%	20%	16%	17%	20%	14%	17%	**17%**
	% within row	31%	36%	33%	100%	58%	42%	100%	**100%**
	Count	17	20	18	55	32	23	55	**55**
2 to 5 days a week	% within column	25%	15%	18%	20%	23%	17%	20%	**20%**
	% within row	45%	23%	31%	100%	56%	44%	100%	**100%**
	Count	29	15	20	64	36	28	64	**64**
Every or nearly every day	% within column	16%	11%	5%	11%	9%	13%	11%	**11%**
	% within row	53%	31%	17%	100%	39%	61%	100%	**100%**
	Count	19	11	6	36	14	22	36	**36**
WEEKLY+	% within column	56%	47%	39%	47%	51%	44%	47%	**47%**
	% within row	42%	30%	28%	100%	53%	47%	100%	**100%**
	Count	65	46	44	155	82	73	155	**155**
TOTAL	% within column	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
	% within row	**35%**	**30%**	**35%**	**100%**	**49%**	**51%**	**100%**	**100%**
	Count	**116**	**98**	**113**	**327**	**160**	**167**	**327**	**327**
Average/year		107.21 ↑	77.2	60.34	82.0	79.3	84.6	82.0	**82.0**

Crosstab of Cola Preference by Age and Gender

A crosstab is a combination of various tables showing summary statistics.

Contingency Tables

A contingency table is a type of table in a matrix format that displays the (multivariate) frequency distribution of the variables. A crucial problem of multivariate statistics is finding the direct dependence structure underlying the variables contained in high dimensional contingency tables. If some of the conditional independences are revealed, then even the storage of the data can be done in a smarter way. In order to do this, one can use information theory concepts, which gain the information only from the distribution of probability. Probability can be expressed easily from the contingency table by the relative frequencies.

As an example, suppose that we have two variables, sex (male or female) and handedness (right- or left-handed). Further suppose that 100 individuals are randomly sampled from a very large population as part of a study of sex differences in handedness. A contingency table can be created to display the numbers of individuals who are male and right-handed, male and left-handed, female and right-handed, and female and left-handed .

The numbers of the males, females, and right-and-left-handed individuals are called marginal totals. The grand total–i.e., the total number of individuals represented in the contingency table– is the number in the bottom right corner.

The table allows us to see at a glance that the proportion of men who are right-handed is about the same as the proportion of women who are right-handed, although the proportions are not identical. If the proportions of individuals in the different columns vary significantly between rows (or vice versa), we say that there is a contingency between the two variables. In other words, the two variables are not independent. If there is no contingency, we say that the two variables are independent.

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Totals	87	13	100

Contingency Table

Contingency table created to display the numbers of individuals who are male and right-handed, male and left-handed, female and right-handed, and female and left-handed.

Standard Components of a Crosstab

- Multiple columns – each column refers to a specific sub-group in the population (e.g., men). The columns are sometimes referred to as banner points or cuts (and the rows are sometimes referred to as stubs).
- Significance tests – typically, either column comparisons–which test for differences between columns and display these results using letters– or cell comparisons–which use color or arrows to identify a cell in a table that stands out in some way (as in the example above).
- Nets or netts – which are sub-totals.
- One or more of the following: percentages, row percentages, column percentages, indexes, or averages.
- Unweighted sample sizes (i.e., counts).

Most general-purpose statistical software programs are able to produce simple crosstabs. Creation of the standard crosstabs used in survey research, as shown above, is typically done using specialist crosstab software packages, such as:

1. New Age Media Systems (EzTab)
2. SAS
3. Quantum
4. Quanvert
5. SPSS Custom Tables
6. IBM SPSS Data Collection Model programs
7. Uncle
8. WinCross
9. Q
10. SurveyCraft
11. BIRT

3.1.2: Drawing a Histogram

To draw a histogram, one must decide how many intervals represent the data, the width of the intervals, and the starting point for the first interval.

Learning Objective

Outline the steps involved in creating a histogram.

Key Takeaways

Key Points

- There is no "best" number of bars, and different bar sizes may reveal different features of the data.
- A convenient starting point for the first interval is a lower value carried out to one more decimal place than the value with the most decimal places.
- To calculate the width of the intervals, subtract the starting point from the ending value and divide by the number of bars.

Key Term

histogram
: a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval

To construct a histogram, one must first decide how many bars or intervals (also called classes) are needed to represent the data. Many histograms consist of between 5 and 15 bars, or classes. One must choose a starting point for the first interval, which must be less than the smallest data value. A convenient starting point is a lower value carried out to one more decimal place than the value with the most decimal places.

For example, if the value with the most decimal places is 6.1, and this is the smallest value, a convenient starting point is 6.05 (6.1−0.05=6.05″>6.1–0.05=6.05). We say that 6.05 has more precision. If the value with the most decimal places is 2.23 and the lowest value is 1.5, a convenient starting point is 1.495 (1.5−0.005=1.495″>1.5–0.005=1.495). If the value with the most decimal places is 3.234 and the lowest value is 1.0, a convenient starting point is 0.9995 (1.0−0.0005=0.9995″>1.0–0.0005=0.9995). If all the data happen to be integers and the smallest value is 2, then a convenient starting point is 1.5 (2−0.5=1.5″>2–0.5=1.5). Also, when the starting point and other boundaries are carried to one additional decimal place, no data value will fall on a boundary.

Consider the following data, which are the heights (in inches to the nearest half inch) of 100 male semiprofessional soccer players. The heights are continuous data since height is measured.

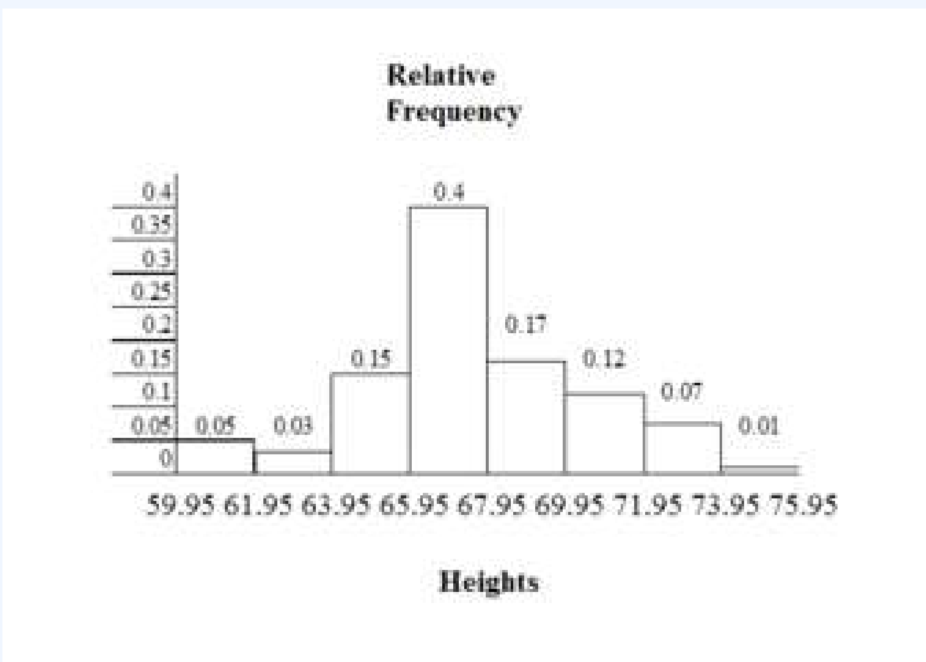
60; 60.5; 61; 61; 61.5; 63.5; 63.5; 63.5; 64; 64; 64; 64; 64; 64; 64; 64.5; 64.5; 64.5; 64.5; 64.5; 64.5; 64.5; 64.5; 66; 66; 66; 66; 66; 66; 66; 66; 66; 66; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 67; 67; 67;

67; 67; 67; 67; 67; 67; 67; 67; 67; 67.5; 67.5; 67.5; 67.5; 67.5; 67.5; 67.5; 68; 68; 69; 69; 69; 69; 69; 69; 69; 69; 69; 69; 69.5; 69.5; 69.5; 69.5; 69.5; 70; 70; 70; 70; 70; 70; 70.5; 70.5; 70.5; 71; 71; 71; 72; 72; 72; 72.5; 72.5; 73; 73.5; 74

The smallest data value is 60. Since the data with the most decimal places has one decimal (for instance, 61.5), we want our starting point to have two decimal places. Since the numbers 0.5, 0.05, 0.005, and so on are convenient numbers, use 0.05 and subtract it from 60, the smallest value, for the convenient starting point. The starting point, then, is 59.95.

The largest value is 74. 74+0.05=74.79″>74+0.05=74.79 is the ending value.

Next, calculate the width of each bar or class interval. To calculate this width, subtract the starting point from the ending value and divide by the number of bars (you must choose the number of bars you desire). Note that there is no "best" number of bars, and different bar sizes can reveal different features of the data. Some theoreticians have attempted to determine an optimal number of bars, but these methods generally make strong assumptions about the shape of the distribution. Depending on the actual data distribution and the goals of the analysis, different bar widths may be appropriate, so experimentation is usually needed to determine an appropriate width.



Histogram Example

This histogram depicts the relative frequency of heights for 100 semiprofessional soccer players.

Note the roughly normal distribution, with the center of the curve around 66 inches. The chart displays the heights on the x-axis and relative frequency on the y-axis.

Suppose, in our example, we choose 8 bars. The bar width will be as follows:

74.05−59.958=1.76″>(74.05−59.95)/8=1.76

We will round up to 2 and make each bar or class interval 2 units wide. Rounding up to 2 is one way to prevent a value from falling on a boundary. The boundaries are:

59.95, 61.95, 63.95, 65.95, 67.95, 69.95, 71.95, 73.95, 75.95

So that there are 2 units between each boundary.

The heights 60 through 61.5 inches are in the interval 59.95 – 61.95. The heights that are 63.5 are in the interval 61.95 – 63.95. The heights that are 64 through 64.5 are in the interval 63.95 – 65.95. The heights 66 through 67.5 are in the interval 65.95 – 67.95. The heights 68 through 69.5 are in the interval 67.95 – 69.95. The heights 70 through 71 are in the interval 69.95 – 71.95. The heights 72 through 73.5 are in the interval 71.95 – 73.95. The height 74 is in the interval 73.95 – 75.95.

3.1.3: Recognizing and Using a Histogram

A histogram is a graphical representation of the distribution of data.

Learning Objectives

Indicate how frequency and probability distributions are represented by histograms.

Key Takeaways

Key Points

- First introduced by Karl Pearson, a histogram is an estimate of the probability distribution of a continuous variable.
- If the distribution of
  is continuous, then
  is called a continuous random variable and, therefore, has a continuous probability distribution.
- An advantage of a histogram is that it can readily display large data sets (a rule of thumb is to use a histogram when the data set consists of 100 values or more).

Key Terms

frequency
: number of times an event occurred in an experiment (absolute frequency)

histogram
: a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval

probability distribution
: A function of a discrete random variable yielding the probability that the variable will have a given value.

A histogram is a graphical representation of the distribution of data. More specifically, a histogram is a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval. First introduced by Karl Pearson, it is an estimate of the probability distribution of a continuous variable.

A histogram has both a horizontal axis and a vertical axis. The horizontal axis is labeled with what the data represents (for instance, distance from your home to school). The vertical axis is labeled either frequency or relative frequency. The graph will have the same shape with either label. An advantage of a histogram is that it can readily display large data sets (a rule of thumb is to use a histogram when the data set consists of 100 values or more). The histogram can also give you the shape, the center, and the spread of the data.

The categories of a histogram are usually specified as consecutive, non-overlapping intervals of a variable. The categories (intervals) must be adjacent and often are chosen to be of the same size. The rectangles of a histogram are drawn so that they touch each other to indicate that the original variable is continuous.

Frequency and Probability Distributions

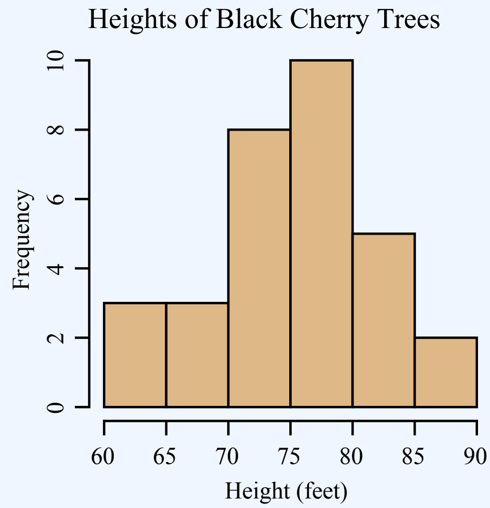
In statistical terms, the frequency of an event is the number of times the event occurred in an experiment or study. The relative frequency (or empirical probability) of an event refers to the absolute frequency normalized by the total number of events:

Put more simply, the relative frequency is equal to the frequency for an observed value of the data divided by the total number of data values in the sample.

The height of a rectangle in a histogram is equal to the frequency density of the interval, i.e., the frequency divided by the width of the interval. A histogram may also be normalized displaying relative frequencies. It then shows the proportion of cases that fall into each of several categories, with the total area equaling one.

As mentioned, a histogram is an estimate of the probability distribution of a continuous variable. To define probability distributions for the simplest cases, one needs to distinguish between discrete and continuous random variables. In the discrete case, one can easily assign a probability to each possible value. For example, when throwing a die, each of the six values 1 to 6 has the probability 1/6. In contrast, when a random variable takes values from a continuum, probabilities are nonzero only if they refer to finite intervals. For example, in quality control one might demand that the probability of a "500 g" package containing between 490 g and 510 g should be no less than 98%.

Intuitively, a continuous random variable is the one which can take a continuous range of values — as opposed to a discrete distribution, where the set of possible values for the random variable is, at most, countable. If the distribution of X is continuous, then X is called a continuous random variable and, therefore, has a continuous probability distribution. There are many examples of continuous probability distributions: normal, uniform, chi-squared, and others.



The Histogram: This is an example of a histogram, depicting graphically the distribution of heights for 31 Black Cherry trees.

3.1.4: The Density Scale

Density estimation is the construction of an estimate based on observed data of an unobservable, underlying probability density function.

Learning Objective

Describe how density estimation is used as a tool in the construction of a histogram.

Key Takeaways

Key Points

- The unobservable density function is thought of as the density according to which a large population is distributed. The data are usually thought of as a random sample from that population.
- A probability density function, or density of a continuous random variable, is a function that describes the relative likelihood for a random variable to take on a given value.
- Kernel density estimates are closely related to histograms, but can be endowed with properties such as smoothness or continuity by using a suitable kernel.

Key Terms

quartile
: any of the three points that divide an ordered distribution into four parts, each containing a quarter of the population

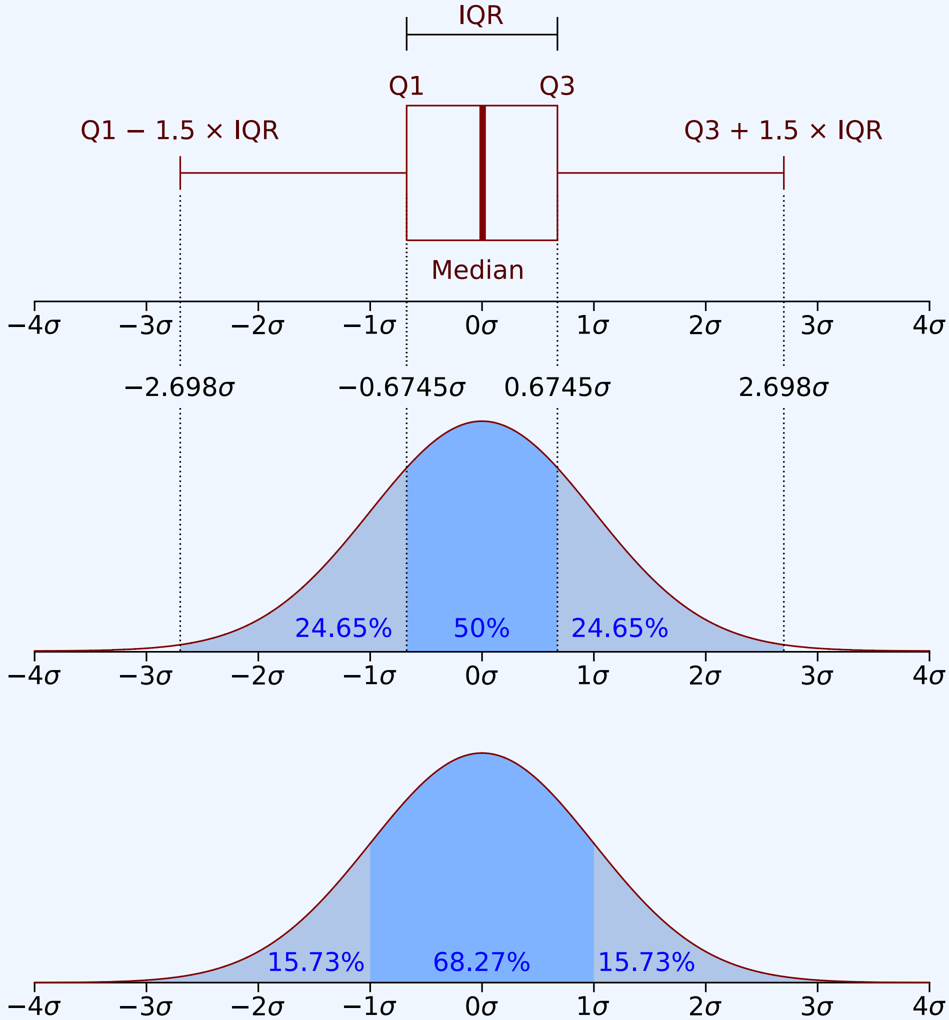
density
: the probability that an event will occur, as a function of some observed variable

interquartile range
: The difference between the first and third quartiles; a robust measure of sample dispersion.

Density Estimation

Histograms are used to plot the density of data, and are often a useful tool for density estimation. Density estimation is the construction of an estimate based on observed data of an unobservable, underlying probability density function. The unobservable density function is thought of as the density according to which a large population is distributed. The data are usually thought of as a random sample from that population.

A probability density function, or density of a continuous random variable, is a function that describes the relative likelihood for this random variable to take on a given value. The probability for the random variable to fall within a particular region is given by the integral of this variable's density over the region .



Boxplot Versus Probability Density Function

This image shows a boxplot and probability density function of a normal distribution.

The above image depicts a probability density function graph against a box plot. A box plot is a convenient way of graphically depicting groups of numerical data through their quartiles. The spacings between the different parts of the box help indicate the degree of dispersion (spread) and skewness in the data and to identify outliers. In addition to the points themselves, box plots allow one to visually estimate the interquartile range.

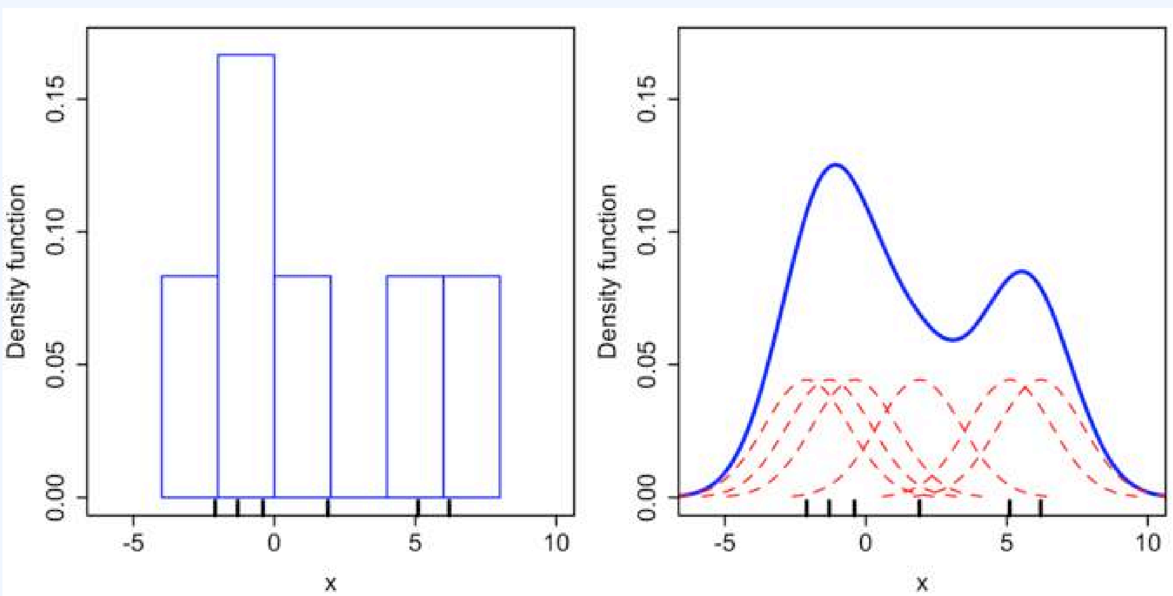
A range of data clustering techniques are used as approaches to density estimation, with the most basic form being a rescaled histogram.

Kernel Density Estimation

Kernel density estimates are closely related to histograms, but can be endowed with properties such as smoothness or continuity by using a suitable kernel. To see this, we compare the construction of histogram and kernel density estimators using these 6 data points:

x1=−2.1">x1=−2.1, x2=−1.3">x2=−1.3, x3=−0.4">x3=−0.4, x4=1.9">x4=1.9, x5=5.1">x5=5.1, x6=6.2">x6=6.2

For the histogram, first the horizontal axis is divided into sub-intervals, or bins, which cover the range of the data. In this case, we have 6 bins, each having a width of 2. Whenever a data point falls inside this interval, we place a box of height 112">1/12. If more than one data point falls inside the same bin, we stack the boxes on top of each other.



Histogram Versus Kernel Density Estimation

Comparison of the histogram (left) and kernel density estimate (right) constructed using the same data. The 6 individual kernels are the red dashed curves, the kernel density estimate the blue curves. The data points are the rug plot on the horizontal axis.

For the kernel density estimate, we place a normal kernel with variance 2.25 (indicated by the red dashed lines) on each of the data points xi">xi. The kernels are summed to make the kernel density estimate (the solid blue curve). Kernel density estimates converge faster to the true underlying density for continuous random variables thus accounting for their smoothness compared to the discreteness of the histogram.

3.1.5: Types of Variables

A variable is any characteristic, number, or quantity that can be measured or counted.

Learning Objective

Distinguish between quantitative and categorical, continuous and discrete, and ordinal and nominal variables.

Key Takeaways

Key Points

- Numeric (quantitative) variables have values that describe a measurable quantity as a number, like "how many" or "how much".
- A continuous variable is an observation that can take any value between a certain set of real numbers.
- A discrete variable is an observation that can take a value based on a count from a set of distinct whole values.
- Categorical variables have values that describe a "quality" or "characteristic" of a data unit, like "what type" or "which category".
- An ordinal variable is an observation that can take a value that can be logically ordered or ranked.
- A nominal variable is an observation that can take a value that is not able to be organized in a logical sequence.

Key Terms

continuous variable
: a variable that has a continuous distribution function, such as temperature

discrete variable
: a variable that takes values from a finite or countable set, such as the number of legs of an animal

variable
: a quantity that may assume any one of a set of values

What Is a Variable?

A variable is any characteristic, number, or quantity that can be measured or counted. A variable may also be called a data item. Age, sex, business income and expenses, country of birth, capital expenditure, class grades, eye colour and vehicle type are examples of variables. Variables are so-named because their value may vary between data units in a population and may change in value over time.

What Are the Types of Variables?

There are different ways variables can be described according to the ways they can be studied, measured, and presented. Numeric variables have values that describe a measurable quantity as a number, like “how many” or “how much. ” Therefore, numeric variables are quantitative variables.

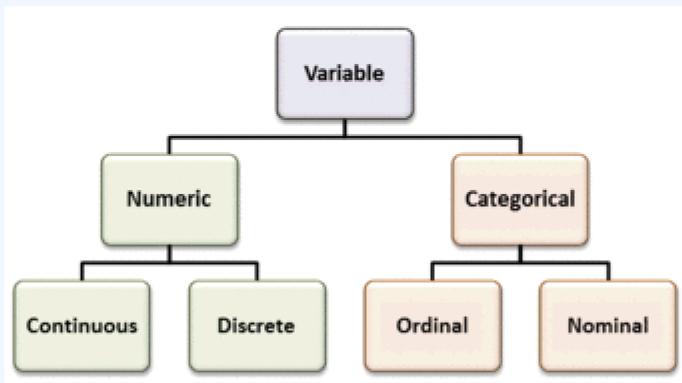
Numeric variables may be further described as either continuous or discrete. A *continuous variable* is a numeric variable. Observations can take any value between a certain set of real numbers. The value given to an observation for a continuous variable can include values as small as the instrument of measurement allows. Examples of continuous variables include height, time, age, and temperature.

A *discrete variable* is a numeric variable. Observations can take a value based on a count from a set of distinct whole values. A discrete variable cannot take the value of a fraction between one value and the next closest value. Examples of discrete variables include the number of registered cars, number of business locations, and number of children in a family, all of of which measured as whole units (i.e., 1, 2, 3 cars).

Categorical variables have values that describe a “quality” or “characteristic” of a data unit, like “what type” or “which category. ” Categorical variables fall into mutually exclusive (in one category or in another) and exhaustive (include all possible options) categories. Therefore, categorical variables are qualitative variables and tend to be represented by a non-numeric value.

Categorical variables may be further described as ordinal or nominal. An *ordinal variable* is a categorical variable. Observations can take a value that can be logically ordered or ranked. The categories associated with ordinal variables can be ranked higher or lower than another, but do not necessarily establish a numeric difference between each category. Examples of ordinal categorical variables include academic grades (i.e., A, B, C), clothing size (i.e., small, medium, large, extra large) and attitudes (i.e., strongly agree, agree, disagree, strongly disagree).

A *nominal variable* is a categorical variable. Observations can take a value that is not able to be organized in a logical sequence. Examples of nominal categorical variables include sex, business type, eye colour, religion and brand.



Types of Variables

Variables can be numeric or categorial, being further broken down in continuous and discrete, and nominal and ordinal variables.

3.1.6: Controlling for a Variable

Controlling for a variable is a method to reduce the effect of extraneous variations that may also affect the value of the dependent variable.

Learning Objective

Discuss how controlling for a variable leads to more reliable visualizations of probability distributions.

Key Takeaways

Key Points

- Variables refer to measurable attributes, as these typically vary over time or between individuals.
- Temperature is an example of a continuous variable, while the number of legs of an animal is an example of a discrete variable.
- In causal models, a distinction is made between "independent variables" and "dependent variables," the latter being expected to vary in value in response to changes in the former.
- While independent variables can refer to quantities and qualities that are under experimental control, they can also include extraneous factors that influence results in a confusing or undesired manner.
- The essence of controlling is to ensure that comparisons between the control group and the experimental group are only made for groups or subgroups for which the variable to be controlled has the same statistical distribution.

Key Terms

correlation
: One of the several measures of the linear statistical relationship between two random variables, indicating both the strength and direction of the relationship.

control
: a separate group or subject in an experiment against which the results are compared where the primary variable is low or nonexistence

variable
: a quantity that may assume any one of a set of values

Histograms help us to visualize the distribution of data and estimate the probability distribution of a continuous variable. In order for us to create reliable visualizations of these distributions, we must be able to procure reliable results for the data during experimentation. A method that significantly contributes to our success in this matter is the controlling of variables.

Defining Variables

In statistics, variables refer to measurable attributes, as these typically vary over time or between individuals. Variables can be discrete (taking values from a finite or countable set), continuous (having a continuous distribution function), or neither. For instance, temperature is a continuous variable, while the number of legs of an animal is a discrete variable.

In causal models, a distinction is made between "independent variables" and "dependent variables," the latter being expected to vary in value in response to changes in the former. In other words, an independent variable is presumed to potentially affect a dependent one. In experiments, independent variables include factors that can be altered or chosen by the researcher independent of other factors.

There are also quasi-independent variables, which are used by researchers to group things without affecting the variable itself. For example, to separate people into groups by their sex does not change whether they are male or female. Also, a researcher may separate people, arbitrarily, on the amount of coffee they drank before beginning an experiment.

While independent variables can refer to quantities and qualities that are under experimental control, they can also include extraneous factors that influence results in a confusing or undesired manner. In statistics the technique to work this out is called correlation.

Controlling Variables

In a scientific experiment measuring the effect of one or more independent variables on a dependent variable, controlling for a variable is a method of reducing the confounding effect of variations in a third variable that may also affect the value of the dependent variable. For example, in an experiment to determine the effect of nutrition (the independent variable) on organism growth (the dependent variable), the age of the organism (the third variable) needs to be controlled for, since the effect may also depend on the age of an individual organism.

The essence of the method is to ensure that comparisons between the control group and the experimental group are only made for groups or subgroups for which the variable to be controlled has the same statistical distribution. A common way to achieve this is to partition the groups into subgroups whose members have (nearly) the same value for the controlled variable.

Controlling for a variable is also a term used in statistical data analysis when inferences may need to be made for the relationships within one set of variables, given that some of these relationships may spuriously reflect relationships to variables in another set. This is broadly equivalent to conditioning on the variables in the second set. Such analyses may be described as "controlling for variable x">x" or "controlling for the variations in x">x". Controlling, in this sense, is performed by including in the experiment not only the explanatory variables of interest but also the extraneous variables. The failure to do so results in omitted-variable bias.

Controlling for Variables

Controlling is very important in experimentation to ensure reliable results. For example, in an experiment to see which type of vinegar displays the greatest reaction to baking soda, the brand of baking soda should be controlled.

3.1.7: Selective Breeding

Selective breeding is a field concerned with testing hypotheses and theories of evolution by using controlled experiments.

Learning Objective

Illustrate how controlled experiments have allowed human beings to selectively breed domesticated plants and animals.

Key Takeaways

Key Points

- Unwittingly, humans have carried out evolution experiments for as long as they have been domesticating plants and animals.
- More recently, evolutionary biologists have realized that the key to successful experimentation lies in extensive parallel replication of evolving lineages as well as a larger number of generations of selection.
- Because of the large number of generations required for adaptation to occur, evolution experiments are typically carried out with microorganisms such as bacteria, yeast, or viruses.

Key Terms

breeding
: the process through which propagation, growth, or development occurs

evolution
: a gradual directional change, especially one leading to a more advanced or complex form; growth; development

stochastic
: random; randomly determined

Experimental Evolution and Selective Breeding

Experimental evolution is a field in evolutionary and experimental biology that is concerned with testing hypotheses and theories of evolution by using controlled experiments. Evolution may be observed in the laboratory as populations adapt to new environmental conditions and/or change by such stochastic processes as random genetic drift.

With modern molecular tools, it is possible to pinpoint the mutations that selection acts upon, what brought about the adaptations, and to find out how exactly these mutations work. Because of the large number of generations required for adaptation to occur, evolution experiments are typically carried out with microorganisms such as bacteria, yeast, or viruses.

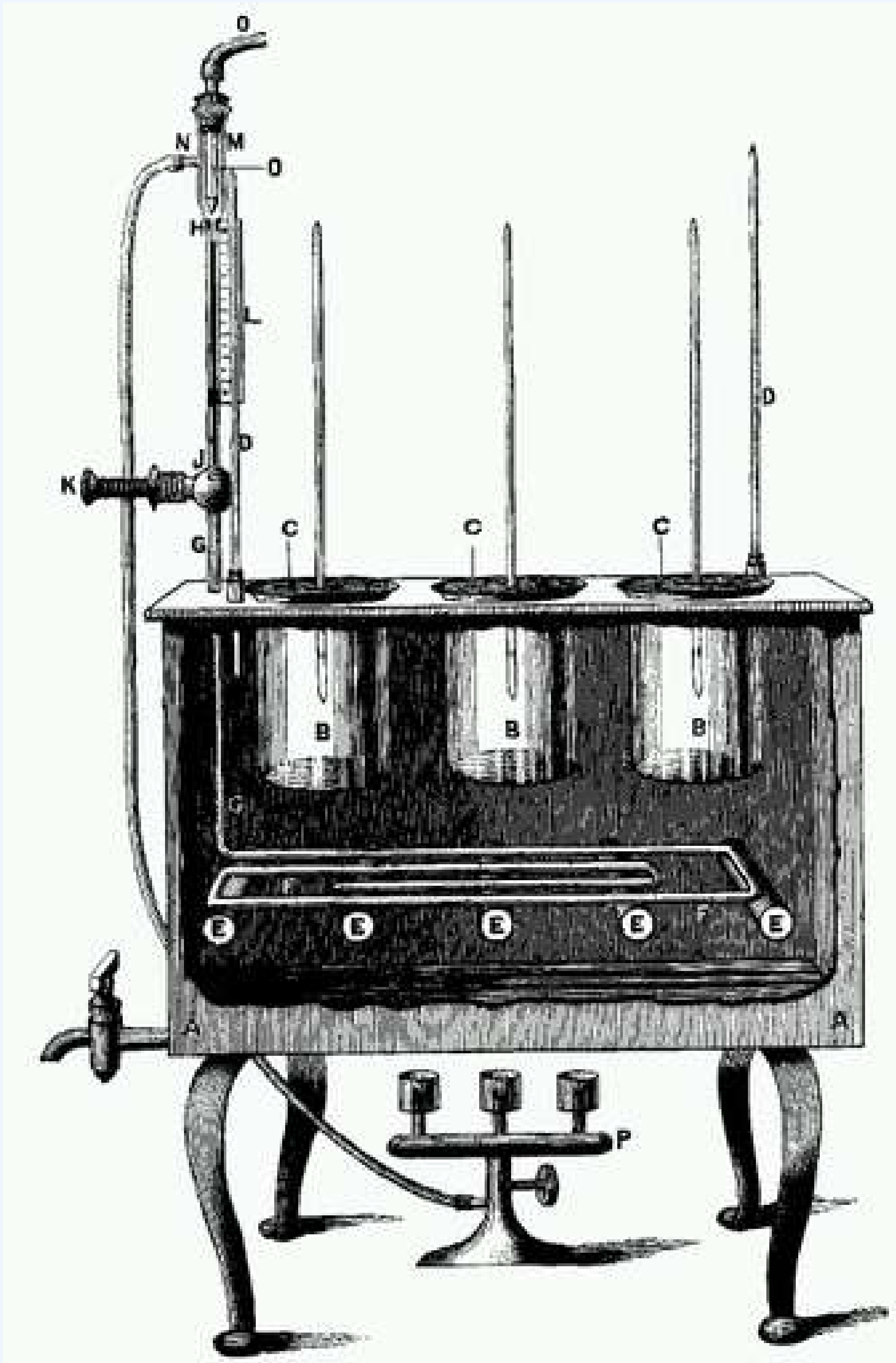
History of Selective Breeding

Unwittingly, humans have carried out evolution experiments for as long as they have been domesticating plants and animals. Selective breeding of plants and animals has led to varieties that differ dramatically from their original wild-type ancestors. Examples are the cabbage varieties, maize, or the large number of different dog breeds .

Selective Breeding

This Chihuahua mix and Great Dane show the wide range of dog breed sizes created using artificial selection, or selective breeding.

One of the first to carry out a controlled evolution experiment was William Dallinger. In the late 19th century, he cultivated small unicellular organisms in a custom-built incubator over a time period of seven years (1880–1886). Dallinger slowly increased the temperature of the incubator from an initial 60 °F up to 158 °F. The early cultures had shown clear signs of distress at a temperature of 73 °F, and were certainly not capable of surviving at 158 °F. The organisms Dallinger had in his incubator at the end of the experiment, on the other hand, were perfectly fine at 158 °F. However, these organisms would no longer grow at the initial 60 °F. Dallinger concluded that he had found evidence for Darwinian adaptation in his incubator, and that the organisms had adapted to live in a high-temperature environment .



Dallinger Incubator

Drawing of the incubator used by Dallinger in his evolution experiments.

More recently, evolutionary biologists have realized that the key to successful experimentation lies in extensive parallel replication of evolving lineages as well as a larger number of generations of selection. For example, on February 15, 1988, Richard Lenski started a long-term evolution experiment with the bacterium E. coli. The experiment continues to this day, and is by now probably the largest controlled evolution experiment ever undertaken. Since the inception of the experiment, the bacteria have grown for more than 50,000 generations.

Attributions

- Cross Tabulation
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Crosstab."
    http://en.wikipedia.org/wiki/Crosstab.
    Wikipedia
    CC BY-SA 3.0.
  - "Contingency table."
    http://en.wikipedia.org/wiki/Contingency_table.
    Wikipedia
    CC BY-SA 3.0.
  - "Cross tabulation."
    http://en.wikipedia.org/wiki/Cross_tabulation.
    Wikipedia
    CC BY-SA 3.0.
  - "cross tabulation."
    http://en.wiktionary.org/wiki/cross_tabulation.
    Wiktionary
    CC BY-SA 3.0.
  - "Contingency table."
    http://en.wikipedia.org/wiki/Contingency_table.
    Wikipedia
    GNU FDL.
  - "Crosstab of Cola Preference by Age and Gender."
    http://commons.wikimedia.org/wiki/
    File:Crosstab_of_Cola_Preference_by_Age_and_Gender.png.
    Wikimedia
    CC BY-SA.
- Drawing a Histogram

    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "Histogram."
      http://en.wikipedia.org/wiki/Histogram.
      Wikipedia
      CC BY-SA 3.0.
    - "Susan DeanBarbara Illowsky, Ph.D., Descriptive Statistics: Histogram. Mar 28, 2013."
      http://cnx.org/contents/20a79748-b312-4c07-ab87-820c5d8aec6e@14.
      OpenStax CNX
      CC BY 4.0.
- Recognizing and Using a Histogram
    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "Frequency (statistics)."
      http://en.wikipedia.org/wiki/Frequency_(statistics).
      Wikipedia
      CC BY-SA 3.0.
    - "Probability distribution."
      http://en.wikipedia.org/wiki/Probability_distribution.
      Wikipedia
      CC BY-SA 3.0.
    - "frequency."
      http://en.wiktionary.org/wiki/frequency.
      Wiktionary
      CC BY-SA 3.0.
    - "Histogram."
      http://en.wikipedia.org/wiki/Histogram.
      Wikipedia
      CC BY-SA 3.0.
    - "probability distribution."
      http://en.wiktionary.org/wiki/probability_distribution.
      Wiktionary
      CC BY-SA 3.0.

    - "Black cherry tree histogram." http://en.wikipedia.org/wiki/File:Black_cherry_tree_histogram.svg. Wikipedia CC BY-SA.
- The Density Scale
    - "Boundless." http://www.boundless.com/. Boundless Learning CC BY-SA 3.0.
    - "Box plot." http://en.wikipedia.org/wiki/Box_plot. Wikipedia CC BY-SA 3.0.
    - "Probability density function." http://en.wikipedia.org/wiki/Probability_density_function. Wikipedia CC BY-SA 3.0.
    - "interquartile range." http://en.wiktionary.org/wiki/interquartile_range. Wiktionary CC BY-SA 3.0.
    - "Density estimation." http://en.wikipedia.org/wiki/Density_estimation. Wikipedia CC BY-SA 3.0.
    - "Histogram." http://en.wikipedia.org/wiki/Histogram. Wikipedia CC BY-SA 3.0.
    - "density." http://en.wiktionary.org/wiki/density. Wiktionary CC BY-SA 3.0.
    - "Kernel density estimation." http://en.wikipedia.org/wiki/Kernel_density_estimation. Wikipedia CC BY-SA 3.0.

- "quartile."
  http://en.wiktionary.org/wiki/quartile.
  Wiktionary
  CC BY-SA 3.0.
- "Boxplot vs PDF."
  http://en.wikipedia.org/wiki/File:Boxplot_vs_PDF.svg.
  Wikipedia
  CC BY-SA.
- "Comparison of 1D histogram and KDE."
  http://en.wikipedia.org/wiki/File:Comparison_of_1D_histogram_and_KDE.png.
  Wikipedia
  CC BY-SA.

- Types of Variables
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Statistical Language – What are Variables?."
    http://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+what+are+variables.
    Austrailian Bureau of Statistics
    CC BY.
  - "continuous variable."
    http://en.wiktionary.org/wiki/continuous_variable.
    Wiktionary
    CC BY-SA 3.0.
  - "variable."
    http://en.wiktionary.org/wiki/variable.
    Wiktionary
    CC BY-SA 3.0.
  - "discrete variable."
    http://en.wiktionary.org/wiki/discrete_variable.
    Wiktionary
    CC BY-SA 3.0.
  - "Statistical Language – What are Variables?."
    http://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+what+are+variables.

Austrailian Bureau of Statistics
CC BY.

- Controlling for a Variable
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Variable (mathematics)."
    http://en.wikipedia.org/wiki/Variable_(mathematics).
    Wikipedia
    CC BY-SA 3.0.
  - "Controlling for a variable."
    http://en.wikipedia.org/wiki/Controlling_for_a_variable.
    Wikipedia
    CC BY-SA 3.0.
  - "control."
    http://en.wiktionary.org/wiki/control.
    Wiktionary
    CC BY-SA 3.0.
  - "variable."
    http://en.wiktionary.org/wiki/variable.
    Wiktionary
    CC BY-SA 3.0.
  - "correlation."
    http://en.wiktionary.org/wiki/correlation.
    Wiktionary
    CC BY-SA 3.0.
  - "vinegar experiment | Flickr – Photo Sharing!."
    http://www.flickr.com/photos/jimmiehomeschoolmom/3464814231/.
    Flickr
    CC BY.
- Selective Breeding
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "evolution."

http://en.wiktionary.org/wiki/evolution.
Wiktionary
CC BY-SA 3.0.

- "stochastic."
  http://en.wiktionary.org/wiki/stochastic.
  Wiktionary
  CC BY-SA 3.0.
- "breeding."
  http://en.wiktionary.org/wiki/breeding.
  Wiktionary
  CC BY-SA 3.0.
- "Experimental evolution."
  http://en.wikipedia.org/wiki/Experimental_evolution.
  Wikipedia
  CC BY-SA 3.0.
- "Big and little dog 1."
  http://en.wikipedia.org/wiki/File:Big_and_little_dog_1.jpg.
  Wikipedia
  CC BY-SA.
- "Dallinger Incubator J.R.Microscop.Soc.1887p193."
  http://en.wikipedia.org/wiki/File:Dallinger_Incubator_J.R.Microscop.Soc.1887p193.png.
  Wikipedia
  Public domain.

3.2 GRAPHING DATA

3.2: Graphing Data

3.2.1: Statistical Graphics

Statistical graphics allow results to be displayed in some sort of pictorial form and include scatter plots, histograms, and box plots.

Learning Objective

Recognize the techniques used in exploratory data analysis

Key Takeaways

Key Points

- Graphical statistical methods explore the content of a data set.
- Graphical statistical methods are used to find structure in data.
- Graphical statistical methods check assumptions in statistical models.
- Graphical statistical methods communicate the results of an analysis.
- Graphical statistical methods communicate the results of an analysis.

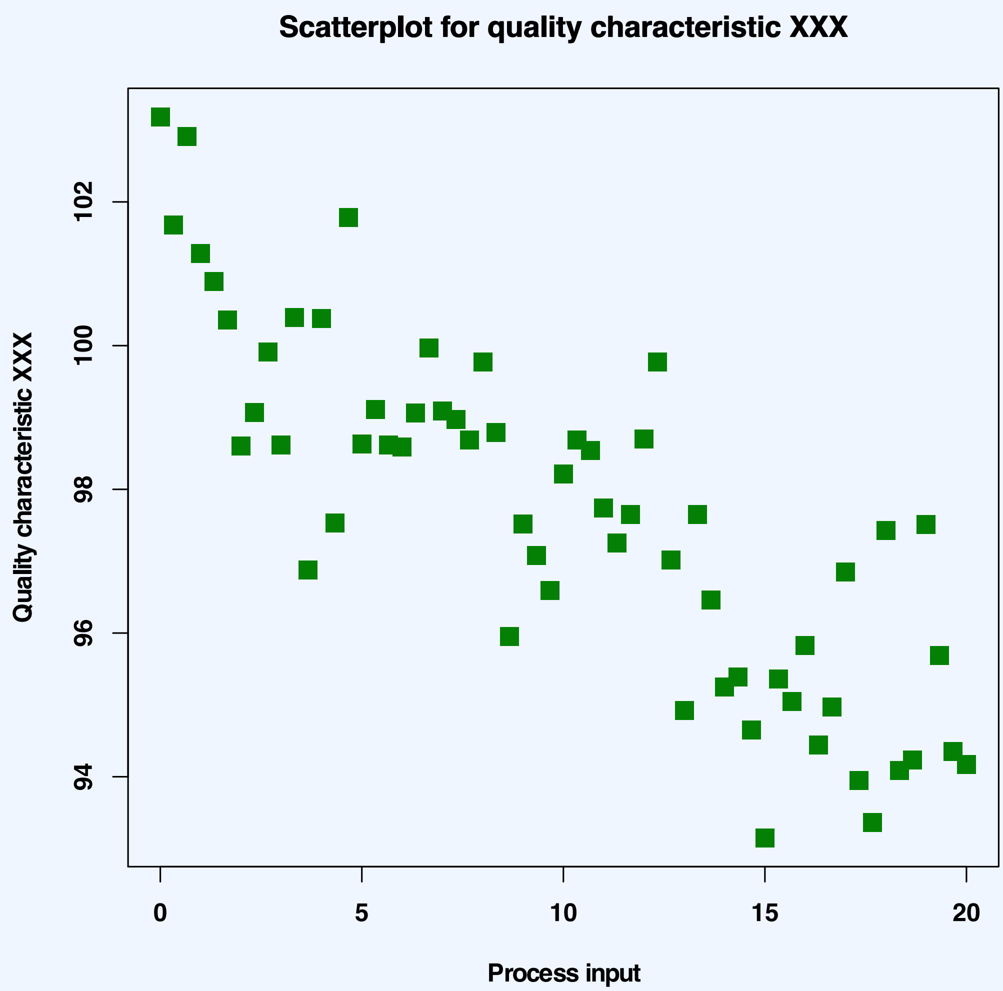
Key Terms

histogram
: a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval

scatter plot
: A type of display using Cartesian coordinates to display values for two variables for a set of data.

box plot
: A graphical summary of a numerical data sample through five statistics: median, lower quartile, upper quartile, and some indication of more extreme upper and lower values.

Statistical graphics are used to visualize quantitative data. Whereas statistics and data analysis procedures generally yield their output in numeric or tabular form, graphical techniques allow such results to be displayed in some sort of pictorial form. They include plots such as scatter plots , histograms, probability plots, residual plots, box plots, block plots and bi-plots.



An example of a scatter plot

A scatter plot helps identify the type of relationship (if any) between two variables.

Exploratory data analysis (EDA) relies heavily on such techniques. They can also provide insight into a data set to help with testing assumptions, model selection and regression model validation, estimator selection, relationship identification, factor effect determination, and outlier detection. In addition, the choice of appropriate statistical graphics can provide a convincing means of communicating the underlying message that is present in the data to others.

Graphical statistical methods have four objectives:

- The exploration of the content of a data set

- The use to find structure in data
- Checking assumptions in statistical models
- Communicate the results of an analysis.

If one is not using statistical graphics, then one is forfeiting insight into one or more aspects of the underlying structure of the data.

Statistical graphics have been central to the development of science and date to the earliest attempts to analyse data. Many familiar forms, including bivariate plots, statistical maps, bar charts, and coordinate paper were used in the 18th century. Statistical graphics developed through attention to four problems:

- Spatial organization in the 17th and 18th century
- Discrete comparison in the 18th and early 19th century
- Continuous distribution in the 19th century and
- Multivariate distribution and correlation in the late 19th and 20th century.

Since the 1970s statistical graphics have been re-emerging as an important analytic tool with the revitalization of computer graphics and related technologies.

3.2.2: Stem-and-Leaf Displays

A stem-and-leaf display presents quantitative data in a graphical format to assist in visualizing the shape of a distribution.

Learning Objective

Construct a stem-and-leaf display

Key Takeaways

Key Points

- Stem-and-leaf displays are useful for displaying the relative density and shape of the data, giving the reader a quick overview of distribution.
- They retain (most of) the raw numerical data, often with perfect integrity. They are also useful for highlighting outliers and finding the mode.
- With very small data sets, a stem-and-leaf displays can be of little use, as a reasonable number of data points are required to establish definitive distribution properties.
- With very large data sets, a stem-and-leaf display will become very cluttered, since each data point must be represented numerically.

Key Terms

outlier
: a value in a statistical sample which does not fit a pattern that describes most other data points; specifically, a value that lies 1.5 IQR beyond the upper or lower quartile

stemplot
: a means of displaying data used especially in exploratory data analysis; another name for stem-and-leaf display

histogram
: a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval

A stem-and-leaf display is a device for presenting quantitative data in a graphical format in order to assist in visualizing the shape of a distribution. This graphical technique evolved from Arthur Bowley's work in the early 1900s, and it is a useful tool in exploratory data analysis. A stem-and-leaf display is often called a stemplot (although, the latter term more specifically refers to another chart type).

Stem-and-leaf displays became more commonly used in the 1980s after the publication of John Tukey 's book on exploratory data analysis in 1977. The popularity during those years is attributable to the use of monospaced (typewriter) typestyles that allowed computer technology of the time to easily produce the graphics. However, the superior graphic capabilities of modern computers have lead to the decline of stem-and-leaf displays.

While similar to histograms, stem-and-leaf displays differ in that they retain the original data to at least two significant digits and put the data in order, thereby easing the move to order-based inference and non-parametric statistics.

Construction of Stem-and-Leaf Displays

A basic stem-and-leaf display contains two columns separated by a vertical line. The left column contains the *stems* and the right column contains the *leaves*. To construct a stem-and-leaf display, the observations must first be sorted in ascending order. This can be done most easily, if working by hand, by constructing a draft of the stem-and-leaf display with the leaves unsorted, then sorting the leaves to produce the final stem-and-leaf display. Consider the following set of data values:

{44,46,47,49,63,64,66,68,68,72,72,75,76,81,84,88,106}">{44,46,47,49,63,64,66,68,68,72,72,75,76,81,84,88,106}

It must be determined what the stems will represent and what the leaves will represent. Typically, the leaf contains the last digit of the number and the stem contains all of the other digits. In the case of very large numbers, the data values may be rounded to a particular place value (such as the hundreds place) that will be used for the leaves. The remaining digits to the left of the rounded place value are used as the stem. In this example, the leaf represents the ones place and the stem will represent the rest of the number (tens place and higher).

The stem-and-leaf display is drawn with two columns separated by a vertical line. The stems are listed to the left of the vertical line. It is important that each stem is listed only once and that no numbers are skipped, even if it means that some stems have no leaves. The leaves are listed in increasing order in a row to the right of each stem. Note that when there is a repeated number in the data (such as two values of 72">72) then the plot must reflect such. Therefore, the plot would appear as 7|2256">7|2256 when it has the numbers {72,72,75,76}">{72,72,75,76}. The display for our data would be as follows:

4	**4679**
5	
6	34688
7	2256
8	148
9	
10	6

Now, let’s consider a data set with both negative numbers and numbers that need to be rounded:

{−23.678758,−12.45,−3.4,4.43,5.5,5.678,16.87,24.7,56.8}">{−23.678758,−12.45,−3.4,4.43,5.5,5.678,16.87,24.7,56.8}

For negative numbers, a negative is placed in front of the stem unit, which is still the value X|10">X|10. Non-integers are rounded. This allows the stem-and-leaf plot to retain its shape, even for more complicated data sets:

-2	**4**
-1	2
-0	3
0	466
1	6
2	4
3	
4	
5	7

Applications of Stem-and-Leaf Displays

Stem-and-leaf displays are useful for displaying the relative density and shape of data, giving the reader a quick overview of distribution. They retain (most of) the raw numerical data, often with perfect integrity. They are also useful for highlighting outliers and finding the mode.

However, stem-and-leaf displays are only useful for moderately sized data sets (around 15 to 150 data points). With very small data sets, stem-and-leaf displays can be of little use, as a reasonable number of data points are required to establish definitive distribution properties. With very large data sets, a stem-and-leaf display will become very cluttered, since each data point must be represented numerically. A box plot or histogram may become more appropriate as the data size increases.

Stem-and-Leaf Display: MPG

```
Stem-and-leaf of MPG  N  = 100
Leaf Unit = 0.10


 1     30  0
 2     31  8
 6     32  5799
 12    33  126899
 18    34  024588
 29    35  01235667899
 49    36  01233445566777888999
(21)   37  000011122334456677899
 30    38  0122345678
 20    39  00345789
 12    40  0123557
 5     41  002
 2     42  1
 1     43
 1     44  9
```

Stem-and-Leaf Display

This is an example of a stem-and-leaf display for EPA data on miles per gallon of gasoline.

3.2.3: Reading Points on a Graph

A graph is a representation of a set of objects where some pairs of the objects are connected by links.

Learning Objective

Distinguish direct and indirect edges

Key Takeaways

Key Points

- The interconnected objects are represented by mathematical abstractions called vertices.
- The links that connect some pairs of vertices are called edges.
- Vertices are also called nodes or points, and edges are also called lines or arcs.

Key Term

graph
: A diagram displaying data; in particular one showing the relationship between two or more quantities, measurements or indicative numbers that may or may not have a specific mathematical formula relating them to each other.

In mathematics, a graph is a representation of a set of objects where some pairs of the objects are connected by links. The interconnected objects are represented by mathematical abstractions called vertices, and the links that connect some pairs of vertices are called edges .Typically, a graph is depicted in diagrammatic form as a set of dots for the vertices, joined by lines or curves for the edges. Graphs are one of the objects of study in discrete mathematics.

The edges may be directed or indirected. For example, if the vertices represent people at a party, and there is an edge between two people if they shake hands, then this is an indirected graph, because if person A shook hands with person B, then person B also shook hands with person A. In contrast, if the vertices represent people at a party, and there is an edge from person A to person B when person A knows of person B, then this graph is directed, because knowledge of someone is not necessarily a symmetric relation (that is, one person knowing another person does not necessarily imply the reverse; for example, many fans may know of a celebrity, but the celebrity is unlikely to know of all their fans). This latter type of graph is called a directed graph and the edges are called directed edges or arcs.Vertices are also called nodes or points, and edges are also called lines or arcs. Graphs are the basic subject studied by graph theory. The word “graph” was first used in this sense by J.J. Sylvester in 1878.

3.2.4: Plotting Points on a Graph

A plot is a graphical technique for representing a data set, usually as a graph showing the relationship between two or more variables.

Learning Objective

Differentiate the different tools used in quantitative and graphical techniques

Key Takeaways

Key Points

- Graphs are a visual representation of the relationship between variables, very useful because they allow us to quickly derive an understanding which would not come from lists of values.
- Quantitative techniques are the set of statistical procedures that yield numeric or tabular output.
- Examples include hypothesis testing, analysis of variance, point estimates and confidence intervals, and least squares regression.
- There are also many statistical tools generally referred to as graphical techniques, which include: scatter plots, histograms, probability plots, residual plots, box plots, and block plots.

Key Term

plot
: a graph or diagram drawn by hand or produced by a mechanical or electronic device

A plot is a graphical technique for representing a data set, usually as a graph showing the relationship between two or more variables. The plot can be drawn by hand or by a mechanical or electronic plotter. Graphs are a visual representation of the relationship between variables, very useful because they allow us to quickly derive an understanding which would not come from lists of values. Graphs can also be used to read off the value of an unknown variable plotted as a function of a known one. Graphs of functions are used in mathematics, sciences, engineering, technology, finance, and many other areas.

Plots play an important role in statistics and data analysis. The procedures here can be broadly split into two parts: quantitative and graphical. Quantitative techniques are the set of statistical procedures that yield numeric or tabular output. Examples of quantitative techniques include:

- hypothesis testing,
- analysis of variance (ANOVA),
- point estimates and confidence intervals, and
- least squares regression.

These and similar techniques are all valuable and are mainstream in terms of classical analysis. There are also many statistical tools generally referred to as graphical techniques. These include:

- scatter plots,
- histograms,
- probability plots,
- residual plots,
- box plots, and
- block plots.

Graphical procedures such as plots are a short path to gaining insight into a data set in terms of testing assumptions, model selection, model validation, estimator selection, relationship identification, factor effect determination, and outlier detection. Statistical graphics give insight into aspects of the underlying structure of the data.

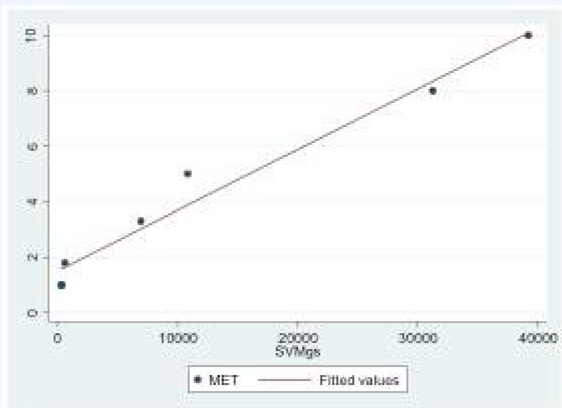
Plotting Points

As an example of plotting points on a graph, consider one of the most important visual aids available to us in the context of statistics: the scatter plot.

To display values for “lung capacity” (first variable) and how long that person could hold his breath, a researcher would choose a group of people to study, then measure each one’s lung capacity (first variable) and how long that person could hold his breath (second variable). The researcher would then plot the data in a scatter plot, assigning "lung capacity" to the horizontal axis, and "time holding breath" to the vertical axis.

A person with a lung capacity of 400 ml who held his breath for 21.7 seconds would be represented by a single dot on the scatter plot at the point

. The scatter plot of all the people in the study would enable the researcher to obtain a visual comparison of the two variables in the data set and will help to determine what kind of relationship there might be between the two variables.



Scatterplot

Scatterplot with a fitted regression line.

3.2.5: Slope and Intercept

The concepts of slope and intercept are essential to understand in the context of graphing data.

Learning Objectives

Explain the term rise over run when describing slope

Key Takeaways

Key Points

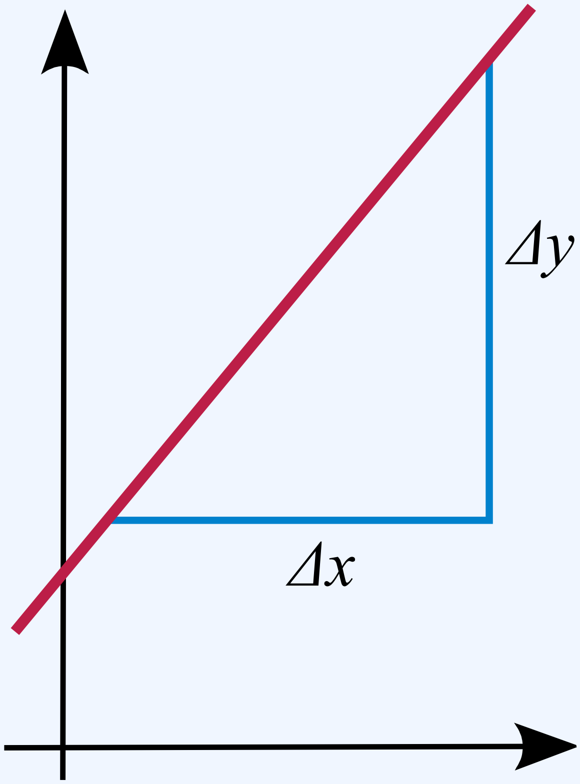
- The slope or gradient of a line describes its steepness, incline, or grade — with a higher slope value indicating a steeper incline.
- The slope of a line in the plane containing the x and y axes is generally represented by the letter m, and is defined as the change in the y coordinate divided by the corresponding change in the x coordinate, between two distinct points on the line.
- Using the common convention that the horizontal axis represents a variable x and the vertical axis represents a variable y, a y– intercept is a point where the graph of a function or relation intersects with the y-axis of the coordinate system.
- Analogously, an x-intercept is a point where the graph of a function or relation intersects with the x-axis.

Key Terms

- intercept: the coordinate of the point at which a curve intersects an axis
- slope: the ratio of the vertical and horizontal distances between two points on a line; zero if the line is horizontal, undefined if it is vertical.

Slope

The slope or gradient of a line describes its steepness, incline, or grade. A higher slope value indicates a steeper incline. Slope is normally described by the ratio of the "rise" divided by the "run" between two points on a line. The line may be practical (as for a roadway) or in a diagram.



Slope: The slope of a line in the plane is defined as the rise over the run, m=ΔyΔx">m=Δy/Δx.

The slope of a line in the plane containing the x and y axes is generally represented by the letter m, and is defined as the change in the y coordinate divided by the corresponding change in the x coordinate, between two distinct points on the line. This is described by the following equation:

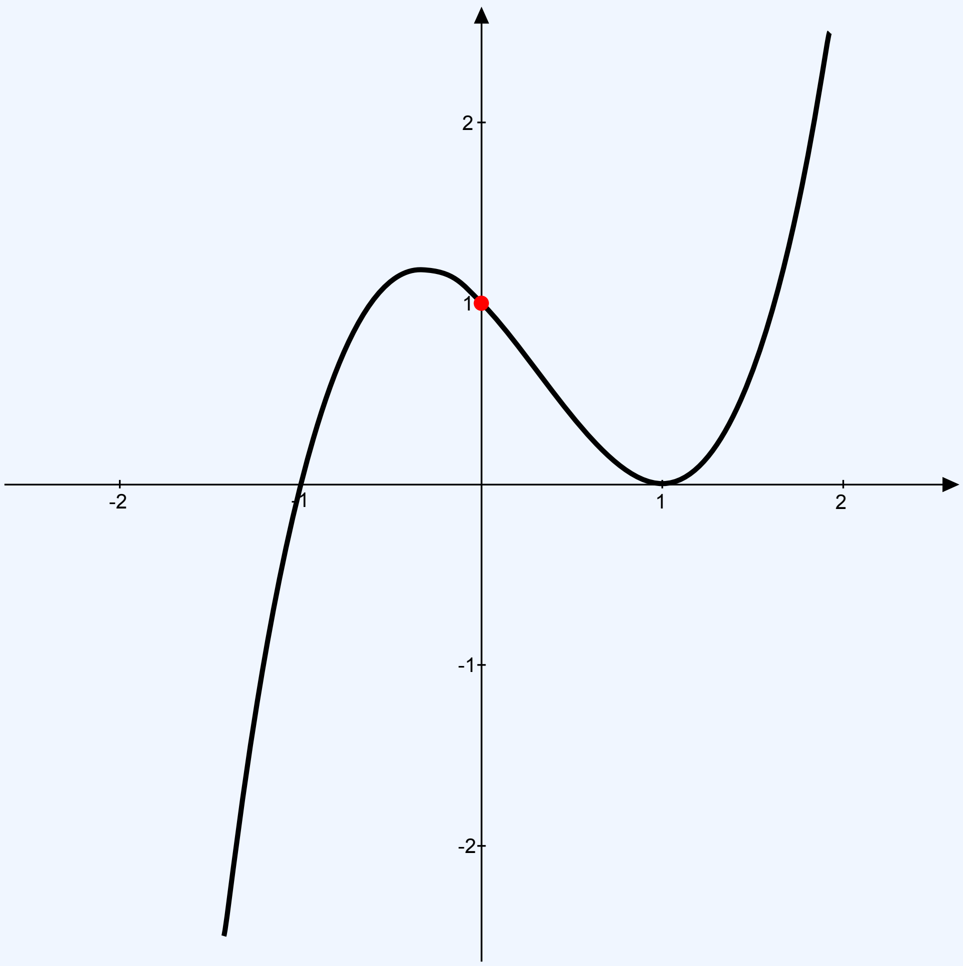
m=ΔyΔx=riserun">m=Δ/yΔx=rise/run

The Greek letter delta, Δ">Δ, is commonly used in mathematics to mean “difference” or “change”. Given two points (x1,y1)">(x1,y1) and (x2,y2)">(x2,y2), the change in x">x from one to the other is x2−x1">x2−x1 (run), while the change in y">y is y2−y1">y2−y1 (rise).

Intercept

Using the common convention that the horizontal axis represents a variable
and the vertical axis represents a variable
, a

-intercept is a point where the graph of a function or relation intersects with the
-axis of the coordinate system. It also acts as a reference point for slopes and some graphs.



Intercept

Graph with a y-intercept at (0,1).

If the curve in question is given as y=f(x)">y=f(x), the y">y-coordinate of the y">y-intercept is found by calculating f(0)">f(0). Functions which are undefined at x=0">x=0 have no y">y-intercept.

Some 2-dimensional mathematical relationships such as circles, ellipses, and hyperbolas can have more than one y">y-intercept. Because functions associate x">x values to no more than one y">y value as part of their definition, they can have at most one y">y-intercept.

Analogously, an x">x-intercept is a point where the graph of a function or relation intersects with the x">x-axis. As such, these points satisfy y=0">y=0. The zeros, or roots, of such a function or relation are the x">x-coordinates of these x">x-intercepts.

3.2.6: Plotting Lines

A line graph is a type of chart which displays information as a series of data points connected by straight line segments.

Learning Objective

Explain the principles of plotting a line graph

Key Takeaways

Key Points

- A line chart is often used to visualize a trend in data over intervals of time – a time series – thus the line is often drawn chronologically.
- A line chart is typically drawn bordered by two perpendicular lines, called axes. The horizontal axis is called the x-axis and the vertical axis is called the y-axis.
- Typically the y-axis represents the dependent variable and the x-axis (sometimes called the abscissa) represents the independent variable.
- In statistics, charts often include an overlaid mathematical function depicting the best-fit trend of the scattered data.

Key Terms

bell curve
: In mathematics, the bell-shaped curve that is typical of the normal distribution.

line
: a path through two or more points (compare 'segment'); a continuous mark, including as made by a pen; any path, curved or straight

gradient
: of a function y = f(x) or the graph of such a function, the rate of change of y with respect to x, that is, the amount by which y changes for a certain (often unit) change in x

A line graph is a type of chart which displays information as a series of data points connected by straight line segments. It is a basic type of chart common in many fields. It is similar to a scatter plot except that the measurement points are ordered (typically by their x-axis value) and joined with straight line segments. A line chart is often used to visualize a trend in data over intervals of time – a time series – thus the line is often drawn chronologically.

Plotting

A line chart is typically drawn bordered by two perpendicular lines, called axes. The horizontal axis is called the x-axis and the vertical axis is called the y-axis. To aid visual measurement, there may be additional lines drawn parallel either axis. If lines are drawn parallel to both axes, the resulting lattice is called a grid.

Each axis represents one of the data quantities to be plotted. Typically the y-axis represents the dependent variable and the x-axis (sometimes called the abscissa) represents the independent variable. The chart can then be referred to as a graph of quantity one versus quantity two, plotting quantity one up the y-axis and quantity two along the x-axis.

Example

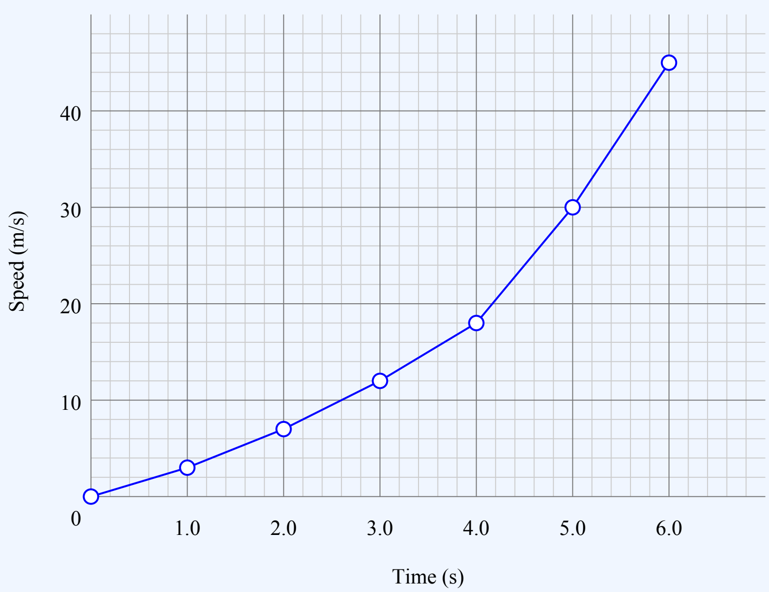
In the experimental sciences, such as statistics, data collected from experiments are often visualized by a graph. For example, if one were to collect data on the speed of a body at certain points in time, one could visualize the data to look like the graph in :

Elapsed Time (s)	Speed (m s^-1)
0	0
1	3
2	
3	12
4	20
5	30
6	45

Data Table

A data table showing elapsed time and measured speed.

The table "visualization" is a great way of displaying exact values, but can be a poor way to understand the underlying patterns that those values represent. Understanding the process described by the data in the table is aided by producing a graph or line chart of Speed versus Time:



Line chart

A graph of speed versus time

Best-Fit

In statistics, charts often include an overlaid mathematical function depicting the best-fit trend of the scattered data. This layer is referred to as a best-fit layer and the graph containing this layer is often referred to as a line graph.

It is simple to construct a "best-fit" layer consisting of a set of line segments connecting adjacent data points; however, such a "best-fit" is usually not an ideal representation of the trend of the underlying scatter data for the following reasons:

1. It is highly improbable that the discontinuities in the slope of the best-fit would correspond exactly with the positions of the measurement values.
2. It is highly unlikely that the experimental error in the data is negligible, yet the curve falls exactly through each of the data points.

In either case, the best-fit layer can reveal trends in the data. Further, measurements such as the gradient or the area under the curve can be made visually, leading to more conclusions or results from the data.

A true best-fit layer should depict a continuous mathematical function whose parameters are determined by using a suitable error-minimization scheme, which appropriately weights the error in the data values. Such curve fitting functionality is often found in graphing software or spreadsheets. Best-fit curves may vary from simple linear equations to more complex quadratic, polynomial, exponential, and periodic curves. The so-called "bell curve", or normal distribution often used in statistics, is a Gaussian function.

3.2.7: The Equation of a Line

In statistics, linear regression can be used to fit a predictive model to an observed data set of y and x values.

Learning Objective

Examine simple linear regression in terms of slope and intercept

Key Takeaways

Key Points

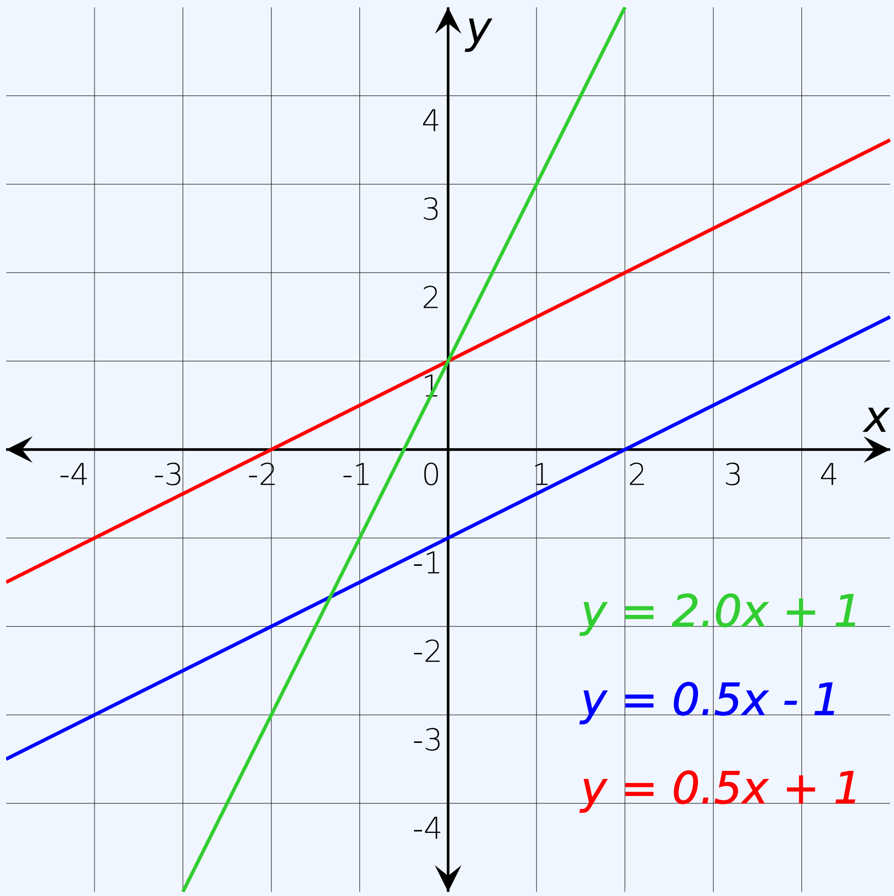
- Simple linear regression fits a straight line through a set of points that makes the vertical distances between the points of the data set and the fitted line as small as possible.
- y=mx+b">y=mx+b, where m">m and b">b designate constants is a common form of a linear equation.
- Linear regression can be used to fit a predictive model to an observed data set of y">y and x">x values.

Key Term

linear regression
: an approach to modeling the relationship between a scalar dependent variable y and one or more explanatory variables denoted x.

In statistics, simple linear regression is the least squares estimator of a linear regression model with a single explanatory variable. Simple linear regression fits a straight line through the set of n">n points in such a way that makes the sum of squared residuals of the model (that is, vertical distances between the points of the data set and the fitted line) as small as possible.

The slope of the fitted line is equal to the correlation between y">y and x">x corrected by the ratio of standard deviations of these variables. The intercept of the fitted line is such that it passes through the center of mass (x,y)">(x,y) of the data points.



The function of a line

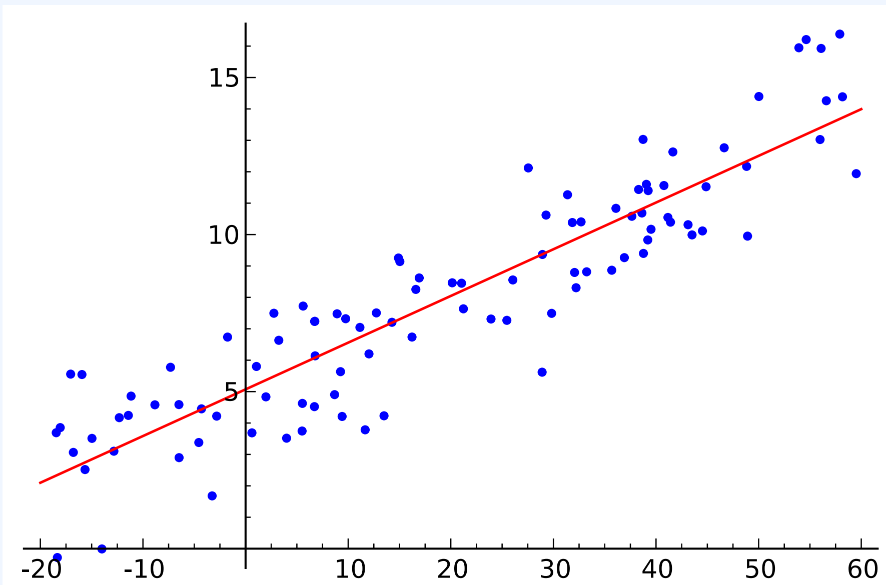
Three lines — the red and blue lines have the same slope, while the red and green ones have same y-intercept.

Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications. This is because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine.

A common form of a linear equation in the two variables x and y is: y= mx + b.

Where m">m (slope) and b">b (intercept) designate constants. The origin of the name "linear" comes from the fact that the set of solutions of such an equation forms a straight line in the plane. In this particular equation, the constant m">m determines the slope or gradient of that line, and the constant term b">b determines the point at which the line crosses the y">y-axis, otherwise known as the y">y-intercept.

If the goal is prediction, or forecasting, linear regression can be used to fit a predictive model to an observed data set of y">y and X">X values. After developing such a model, if an additional value of X">X is then given without its accompanying value of y">y, the fitted model can be used to make a prediction of the value of y">y.



Linear regression

An example of a simple linear regression analysis

Attributions

- Statistical Graphics
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "scatter plot."
    http://en.wiktionary.org/wiki/scatter_plot.
    Wiktionary
    CC BY-SA 3.0.

- "Exploratory data analysis."
  http://en.wikipedia.org/wiki/Exploratory_data_analysis.
  Wikipedia
  CC BY-SA 3.0.
- "histogram."
  http://en.wikipedia.org/wiki/histogram.
  Wikipedia
  CC BY-SA 3.0.
- "box plot."
  http://en.wiktionary.org/wiki/box_plot.
  Wiktionary
  CC BY-SA 3.0.
- "Scatter diagram for quality characteristic XXX."
  http://en.wikipedia.org/wiki/File:Scatter_diagram_for_quality_characteristic_XXX.svg.
  Wikipedia
  CC BY-SA.

- Stem-and-Leaf Displays
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Stem-and-leaf display."
    http://en.wikipedia.org/wiki/Stem-and-leaf_display.
    Wikipedia
    CC BY-SA 3.0.
  - "stemplot."
    http://en.wiktionary.org/wiki/stemplot.
    Wiktionary
    CC BY-SA 3.0.
  - "histogram."
    http://en.wikipedia.org/wiki/histogram.
    Wikipedia
    CC BY-SA 3.0.
  - "outlier."
    http://en.wiktionary.org/wiki/outlier.
    Wiktionary
    CC BY-SA 3.0.

    - “Lab 1 – Stem-and-Leaf Display.”
      http://commons.wikimedia.org/wiki/File:Lab_1_-_Stem-and-Leaf_Display.png.
      Wikimedia
      CC BY-SA.
- Reading Points on a Graph
    - “Boundless.”
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - “Points and lines.”
      http://en.wikipedia.org/wiki/Points_and_lines.
      Wikipedia
      CC BY-SA 3.0.
    - “graph.”
      http://en.wiktionary.org/wiki/graph.
      Wiktionary
      CC BY-SA 3.0.
    - “Linear regression.”
      http://en.wikipedia.org/wiki/Linear_regression.
      Wikipedia
      CC BY-SA 3.0.
- Plotting Points on a Graph
    - “Boundless.”
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - “Plot (graphics).”
      http://en.wikipedia.org/wiki/Plot_(graphics).
      Wikipedia
      CC BY-SA 3.0.
    - “Plot (graphics).”
      http://en.wikipedia.org/wiki/Plot_(graphics).
      Wikipedia
      CC BY-SA 3.0.
    - “Scatterplot.”
      http://en.wikipedia.org/wiki/Scatterplot.
      Wikipedia

CC BY-SA 3.0.
- "plot."
  http://en.wiktionary.org/wiki/plot.
  Wiktionary
  CC BY-SA 3.0.
- "Scatterplot with a fitted regression line for Mean SVMgs = accelerometer output and MET value."
  http://commons.wikimedia.org/wiki/File:Scatterplot_with_a_fitted_regression_line_for_Mean_SVMgs_=_accelerometer_output_and_MET_value.jpg.
  Wikimedia
  CC BY-SA.

- Slope and Intercept
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Y-intercept."
    http://en.wikipedia.org/wiki/Y-intercept.
    Wikipedia
    CC BY-SA 3.0.
  - "Slope."
    http://en.wikipedia.org/wiki/Slope.
    Wikipedia
    CC BY-SA 3.0.
  - "intercept."
    http://en.wiktionary.org/wiki/intercept.
    Wiktionary
    CC BY-SA 3.0.
  - "slope."
    http://en.wiktionary.org/wiki/slope.
    Wiktionary
    CC BY-SA 3.0.
  - "Y-intercept."
    http://en.wikipedia.org/wiki/File:Y-intercept.svg.
    Wikipedia
    CC BY.

  - "Slope picture." http://en.wikipedia.org/wiki/File:Slope_picture.svg. Wikipedia Public domain.
- Plotting Lines
  - "Boundless." http://www.boundless.com/. Boundless Learning CC BY-SA 3.0.
  - "Line chart." https://en.wikipedia.org/wiki/Line_chart. Wikipedia CC BY-SA 3.0.
  - "Line chart." http://en.wikipedia.org/wiki/Line_chart. Wikipedia CC BY-SA 3.0.
  - "bell curve." http://en.wiktionary.org/wiki/bell_curve. Wiktionary CC BY-SA 3.0.
  - "line." http://en.wiktionary.org/wiki/line. Wiktionary CC BY-SA 3.0.
  - "gradient." http://en.wiktionary.org/wiki/gradient. Wiktionary CC BY-SA 3.0.
  - "Line chart." https://en.wikipedia.org/wiki/Line_chart. Wikipedia CC BY-SA.
  - "Line chart." http://en.wikipedia.org/wiki/Line_chart. Wikipedia Public domain.

- The Equation of a Line
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Line equation."
    http://en.wikipedia.org/wiki/Line_equation.
    Wikipedia
    CC BY-SA 3.0.
  - "Simple linear regression."
    http://en.wikipedia.org/wiki/Simple_linear_regression.
    Wikipedia
    CC BY-SA 3.0.
  - "Correlation and dependence."
    http://en.wikipedia.org/wiki/Correlation_and_dependence.
    Wikipedia
    CC BY-SA 3.0.
  - "Linear regression."
    http://en.wikipedia.org/wiki/Linear_regression.
    Wikipedia
    CC BY-SA 3.0.
  - "FuncionLineal01."
    http://en.wikipedia.org/wiki/File:FuncionLineal01.svg.
    Wikipedia
    CC BY-SA.
  - "Linear regression."
    http://en.wikipedia.org/wiki/File:Linear_regression.svg.
    Wikipedia
    Public domain.

SECTION V

3.XLSX – EXCEL CHALLENGE - PRESENTING DATA WITH CHARTS

One of the most important things to consider when using charts in Excel is that they are intended to be used for communicating an idea to an audience. Your audience can be reading your charts in a written document or listening to you in a live presentation. In fact, Excel charts are often imported or pasted into Word documents or PowerPoint slides, which serve this very purpose of communicating ideas to an audience. Although there are no rules set in stone for using specific charts for certain data types, some chart types are designed to communicate certain messages better than others. This chapter explores numerous charts that can be used for a variety of purposes. In addition, we will examine formatting charts and using those charts in Word and PowerPoint documents.

Attribution

Adapted by Hallie Puncochar and Noreen Brown from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

3.XLSX.1 CHOOSING A CHART TYPE

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

LEARNING OBJECTIVES

1. Construct a line chart to show timeline and comparison trends.
2. Learn how to use a column chart to show one and two frequency distributions.
3. Create and format a map chart.
4. Insert a funnel chart.
5. Learn how to use a pie chart to show the percent of the total for a data set.
6. Compare the difference between a column chart and a bar chart.
7. Construct column charts to show how a percent of total changes over time.

This section reviews the most commonly used Excel chart types. To demonstrate the variety of chart types available in Excel, it is necessary to use a variety of data sets. This is necessary not only to demonstrate the construction of charts but also to explain how to choose the right type of chart given your data and the idea you intend to communicate.

Choosing a Chart Type

Before we begin, let's review a few key points you need to consider before creating any chart in Excel.

1. **The first is identifying your idea or message**. It is important to keep in mind that the primary purpose of a chart is to present quantitative information to an audience. Therefore, you must first decide what message or idea you wish to present. This is critical in helping you select specific data from a worksheet that will be used in a chart. Throughout this chapter, we will reinforce the intended message first before creating each chart.
2. **The second key point is selecting the right chart type**. The chart type you select will depend on the data you have and the message you intend to communicate.

3. **The third key point is identifying the values that should appear on the X and Y axes**. One of the ways to identify which values belong on the X and Y axes is to sketch the chart on paper first. If you can visualize what your chart is supposed to look like, you will have an easier time selecting information correctly and using Excel to construct an effective chart that accurately communicates your message. **Table 4.1** "Key Steps Before Constructing an Excel Chart" provides a brief summary of these points.

Integrity Check

Carefully Select Data When Creating a Chart

Just because you have data in a worksheet does not mean it must all be placed onto a chart. When creating a chart, it is common for only specific data points to be used. To determine what data should be used when creating a chart, you must first identify the message or idea that you want to communicate to an audience.

Table 4.1 Key Steps before Constructing an Excel Chart

Step	**Description**
Define your message.	Identify the main idea you are trying to communicate to an audience. If there is no main point or important message that can be revealed by a chart, you might want to question the necessity of creating a chart.
Identify the data you need.	Once you have a clear message, identify the data on a worksheet that you will need to construct a chart. In some cases, you may need to create formulas or consolidate items into broader categories.
Select a chart type.	The type of chart you select will depend on the message you are communicating and the data you are using.
Identify the values for the X and Y axes.	After you have selected a chart type, you may find that drawing a sketch is helpful in identifying which values should be on the X and Y axes. In Excel, the axes are: The "category" axis. Usually the horizontal axis – where the labels are found. The "value" axis. Usually the vertical axis – where the numbers are found.

Time Series Trend: Line Chart 1

The first chart we will demonstrate is a line chart. Figure 4.1 shows part of the data that will be used to create two line charts. This chart will show the trend of the NASDAQ stock index.

Read more: http://www.investopedia.com/terms/n/nasdaq.asp

This chart will be used to communicate a simple message: to show how the index has performed over a two-year period. We can use this chart in a presentation to show whether stock prices have been increasing, decreasing, or remaining constant over the designated period of time.

Before we create the line chart, it is important to identify why it is an appropriate chart type given the message we wish to communicate and the data we have. When presenting the trend for any data over a designated period of time, the most commonly used chart types are the line chart and the column chart. With the column chart, you are limited to a certain number of bars or data points. As shown below in **Figure 4.1**, as the number of bars increases on a column chart, it becomes increasingly difficult to read. In our first example, there are 24 points of data used to construct the chart. This is generally too many data points to put on a column chart, which is why we are using a line chart.

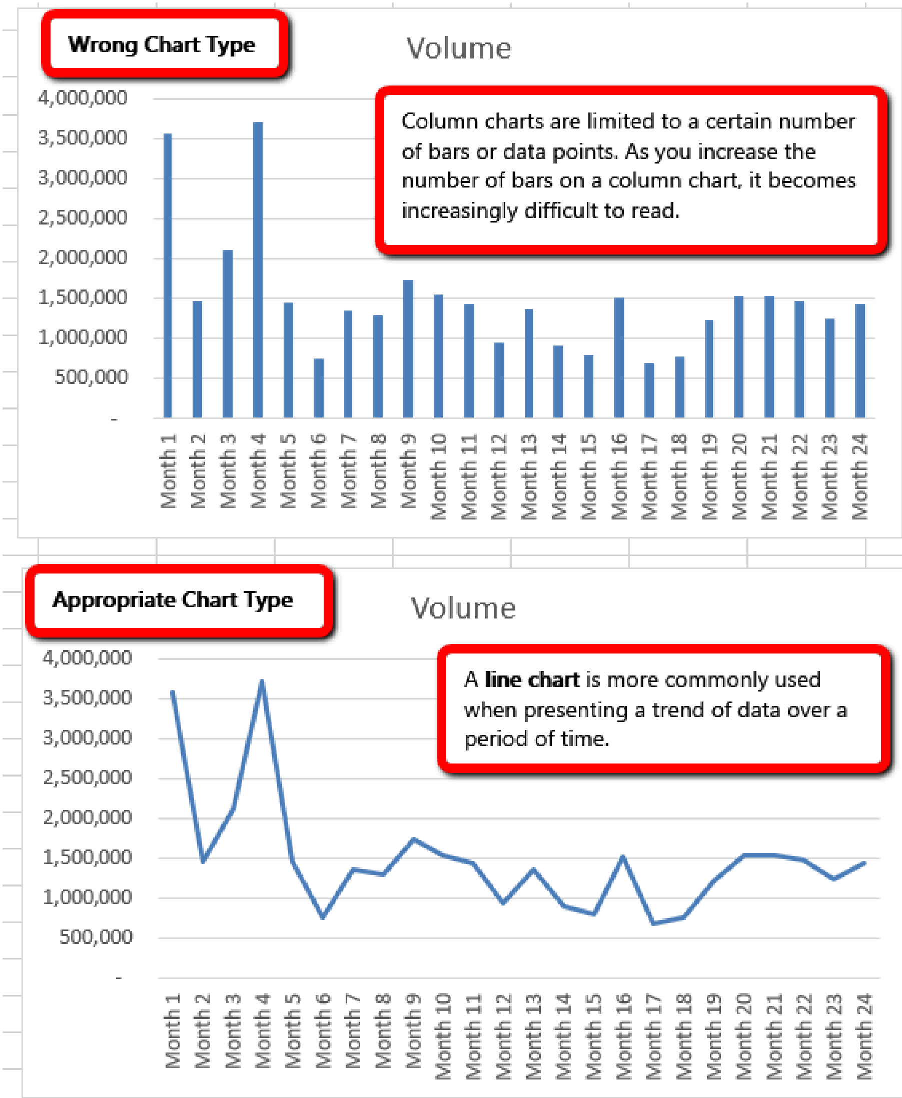


Figure 4.1 Chart Types

Our line chart will show the **Volume** of sales for the NASDAQ on the Y-axis and the **Month** number on the X-axis. Notice when we select the information we are including the column headings/labels. Excel will use the column headings/labels to identify each axis. Excel named the horizontal axis labels Month and the Vertical axis series Volume.

The following steps explain how to construct this chart:

Download Data file: CH4 Data

1. Open data file *CH4 Data* and save a file to your computer as **CH4 Charting**.
2. Navigate to the **Stock Trend** worksheet.
3. Highlight the range **B4:C28** on the Stock Trend worksheet. (Note – you have selected a label in the first row and more labels in column B. Watch where they show up in your completed chart.)
4. Click the Insert tab of the ribbon.

5. Click the **Line** button in the **Charts** group of commands. Click the first option from the list, which is a basic 2D Line Chart (see **Figure 4.2**). Notice Excel adds, or embeds, the line chart into the worksheet.

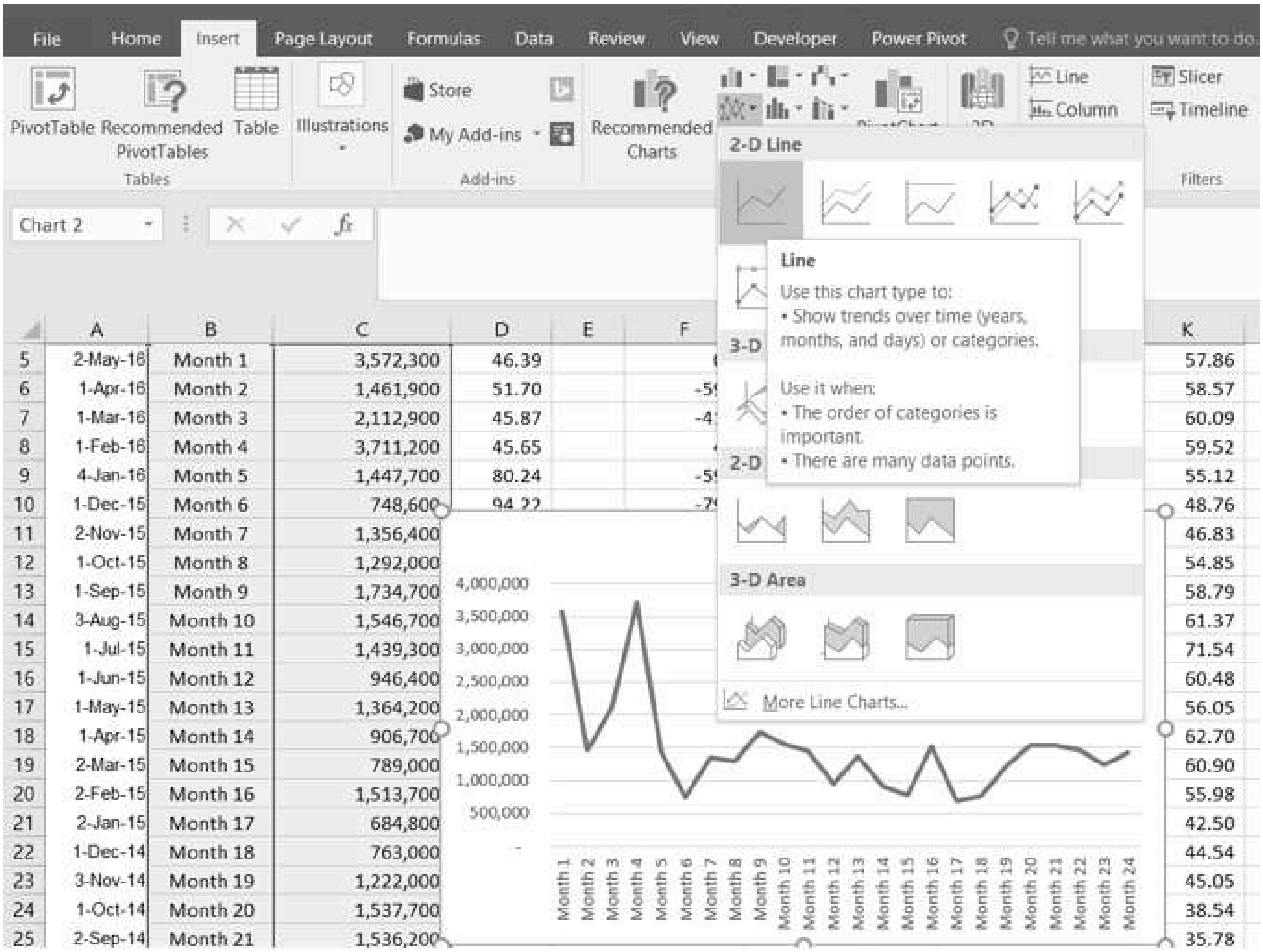


Figure 4.2 Selecting the Basic Line Chart

Why?

Line Chart vs. Column Chart

We can use both a line chart and a column chart to illustrate a trend over time. However, a line chart is far more effective when there are many periods of time being measured. For example, if we are measuring fifty-two weeks, a column chart would require fifty-two bars. A general rule of thumb is to use a column chart when twenty bars or less are required. A column chart becomes difficult to read as the number of bars exceeds twenty.

Figure 4.3 shows the embedded line chart in the Stock Trend worksheet. Do you see where your labels showed up on the chart?

Notice that additional tabs, or contextual tabs, are added to the ribbon. We will demonstrate the commands in these tabs throughout this chapter. These tabs appear only when the chart is activated.

Note: Excel 2010 uses three contextual tabs for charts. Later versions use only two. Each has all the same tools. They are just organized a little differently.

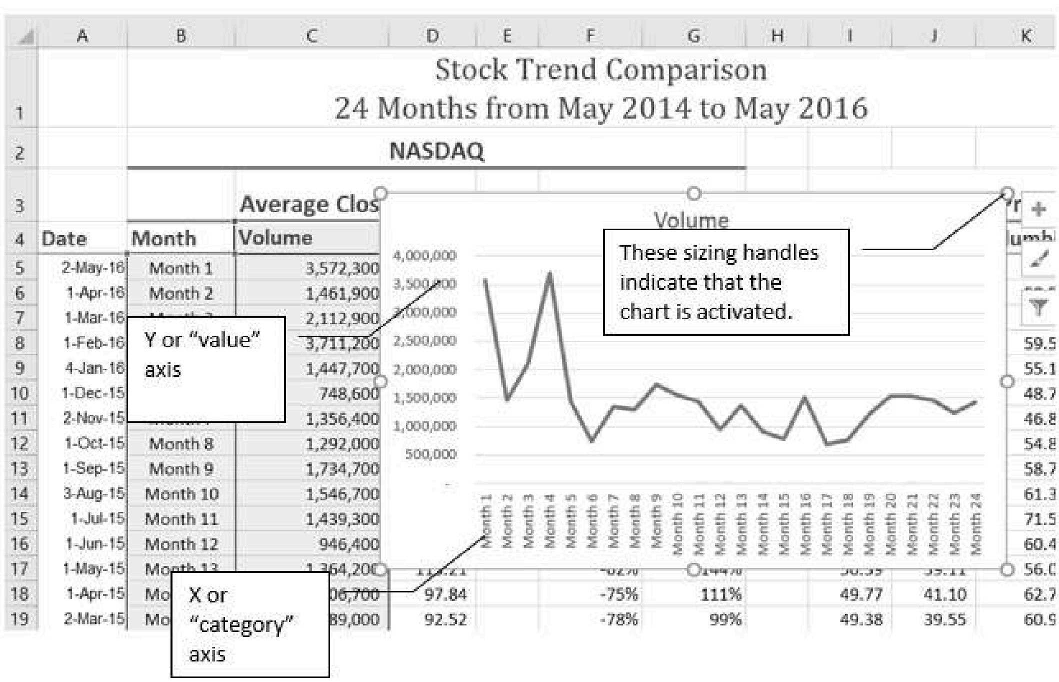
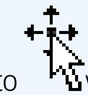


Figure 4.3 Embedded Line Chart in the Stock Trend Worksheet

As shown in **Figure 4.3**, the embedded chart is not placed in an ideal location on the worksheet since it is covering several cell locations that contain data. The following steps demonstrate common adjustments that are made when working with embedded charts:

1. **Moving a chart**: Click and drag the upper left corner of the chart to the corner of cell B30.

Note: Keep an eye on your pointer. It will change into  when you are in the right place to move your chart.

2. **Resizing a chart**: Place the mouse pointer over the bottom lower corner sizing handle, drag and drop to approximately the end of Column I, and Row 45.

Note: keep an eye on your pointer. It will change into  when you are in the right place to resize your chart

3. **Adjusting the chart title**: Click the chart title once. Then click in front of the first letter. You should see a blinking cursor in front of the letter. This allows you to modify the title of the chart.
4. Type the following in front of the first letter in the chart title: **May 2014-2016 Trend for NASDAQ Sales**.
5. Click anywhere outside of the chart to deactivate it.
6. Save your work.

Figure 4.4 shows the line chart after it is moved and resized. Notice that the sizing handles do not appear around the perimeter of the chart. This is because the chart has been deactivated. To activate the chart, click anywhere inside the chart perimeter.

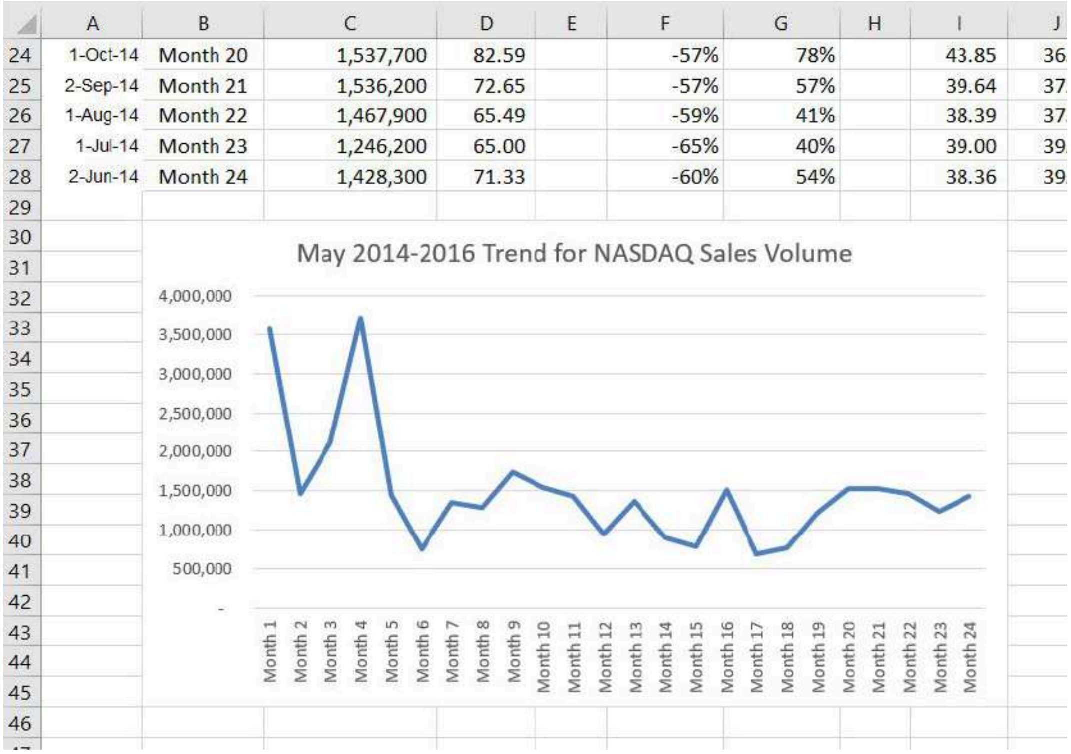


Figure 4.4 Line Chart Moved and Resized

Integrity Check

When using line charts in Excel, keep in mind that anything placed on the X-axis is considered a descriptive label, not a numeric value. This is an example of a category axis. This is important because there will never be a change in the spacing of any items placed on the X-axis of a line chart. If you need to create a chart using numeric data on the category axis, you will have to modify the chart. We will do that later in the chapter.

Skill Refresher

Inserting a Line Chart

1. Highlight a range of cells that contain data that will be used to create the chart. Be sure to include labels in your selection.
2. Click the Insert tab of the ribbon.
3. Click the Line button in the Charts group.
4. Select a format option from the Line Chart drop-down menu.

Adjusting the Y-Axis Scale

After creating an Excel chart, you may find it necessary to adjust the scale of the Y-axis. Excel automatically sets the maximum value for the Y-axis based on the data used to create the chart. The minimum value is usually set to zero. That is usually a good thing. However, depending on the data you are using to create the chart, setting the minimum value to zero can substantially minimize the graphical presentation of a trend. For example, the trend shown in Figure 4.4 appears to be increasing slightly in recent months. The presentation of this trend can be improved if the minimum value started at 500,000. The following steps explain how to make this adjustment to the Y-axis:

1. Click anywhere on the Y (value or vertical) axis on the **May 2014-2016 Trend for NASDAQ Sales Volume** line chart (Stock Trend worksheet).
2. Right Click and select **Format Axis.** The Format Axis Pane should appear, as shown in **Figure 4.5**.

Mac Users: Hold down the Control key and click the Y axis. Then choose **Format Axis.**

Note: If you do not see "Format Axis . . . on your menu, you have not right-clicked in the correct spot. Press "Escape" to turn the menu off and try again

3. In the Format Axis Pane, click the input box for the "**Minimum**" axis option and **delete** the zero. Then type the number **500000** and hit Enter. As soon as you make this change, the Y axis on the chart adjusts.

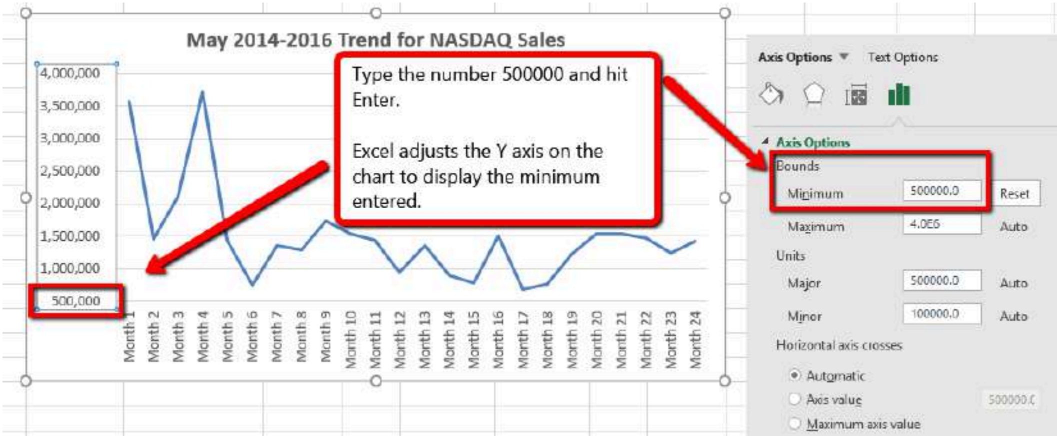


Figure 4.5 Format Axis Pane

4. Click the X in the upper right corner of the Format Axis pane to close it.
5. Save your work.

Figure 4.6 shows the change in the presentation of the trend line. Notice that with the Y axis starting at 500,000, the trend for the NASDAQ is more pronounced. This adjustment makes it easier for the audience to see the magnitude of the trend.

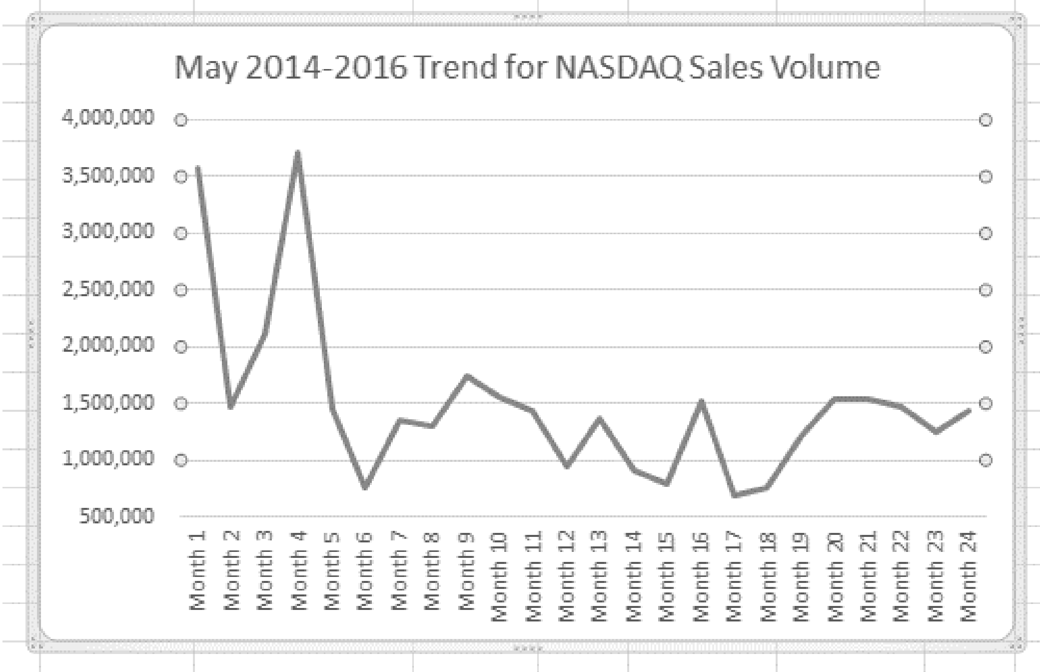


Figure 4.6 Adjusted Y-Axis for the S&P 500 Chart

Skill Refresher

Adjusting the Y-Axis Scale

1. Click anywhere along the **Y-axis** to activate it.
2. Right Click.
   (Note, you can also select the Format tab in the Chart Tools section of the ribbon.)
3. Select **Format Axis** . . .
4. In the **Format Axis pane**, make your changes to the **Axis Options**.
5. Click in the input box next to the desired axis option and then type the new scale value.
6. Click the **Close** button at the top right of the Format Axis pane to close it.

Trend Comparisons: Line Chart 2

We will now create a second line chart using the data in the Stock Trend worksheet. The purpose of this chart is to compare two trends: the change in volume for the NASDAQ and the change in the Closing price.

Before creating the chart to compare the NASDAQ volume and sales price, it is important to review the data in the range B4:D28 on the Stock Trend worksheet. We cannot use the volume of sales and the closing price because the values are not comparable. That is, the closing price is in a range of $45.00 to $115.00, but the data for the volume of Sales is in a range of 684,000 to 3,711,000. If we used these values – without making changes to the chart — we would not be able to see the closing price at all.

The construction of this second line chart will be similar to the first line chart. The X axis will be the months in the range B4:D28.

1. Highlight the range **B4:D28** on the Stock Trend worksheet.
2. Click the **Insert** tab of the ribbon.
3. Click the **Line** button in the **Charts** group of commands.
4. Click the first option from the list, which is a basic line chart.

Figure 4.6.5 shows the appearance of the line chart comparing both the volume and the closing price before it is moved and resized. Notice that the line for the closing price (Close) appears as a straight line at the bottom of the chart.

Figure 4.6.5 Trend Comparison Line Chart

The line representing the closing values is flat along the bottom of the chart. This is hard to see and not very useful as is. Fear not. We will fix that.

1. Move the chart so the upper left corner is in the middle of cell M1.
2. Resize the chart, using the resizing handle so the graph is approximately in the area of M1:U13.
3. Click in the text box that says "**Chart Title**." Delete the text and replace it with the following: **24 Month Trend Comparison.**
4. Adjust the Closing Price axis, by double-clicking the red line across the bottom of the chart that represents the Closing Price.
5. The Format Data Series dialogue box opens. In the Series Options, select Secondary Axis.

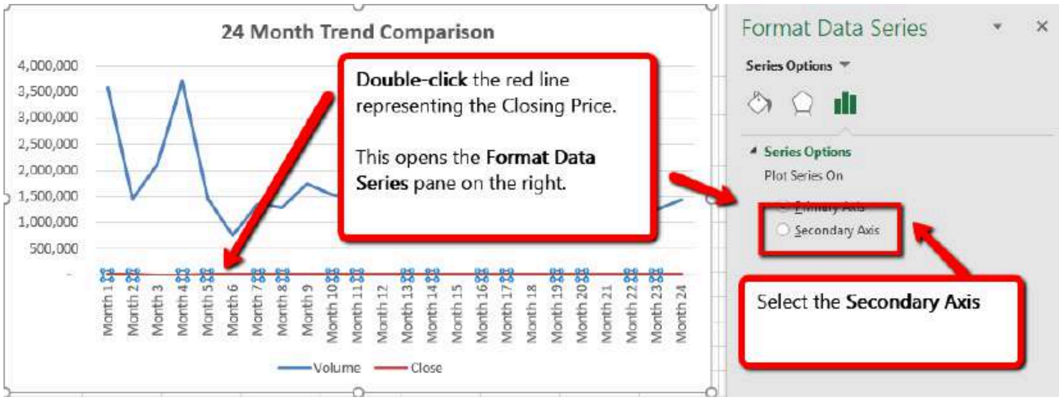


Figure 4.7 Adding a Secondary Axis

Excel adds the secondary axis. Format the values on the secondary axis to represent prices.

1. Double click the Secondary Vertical Axis. (The vertical axis on the right that goes from 0 to 140.)
2. In axis options, scroll down to the **Number** section.

   Mac Users: If needed, click the Number "expand arrow" 

3. Use the Symbol list box to add the $.
4. Press the Close button to close the Format Axis pane.
5. Save your work.

Figure 4.8 Modifying the Secondary Axis

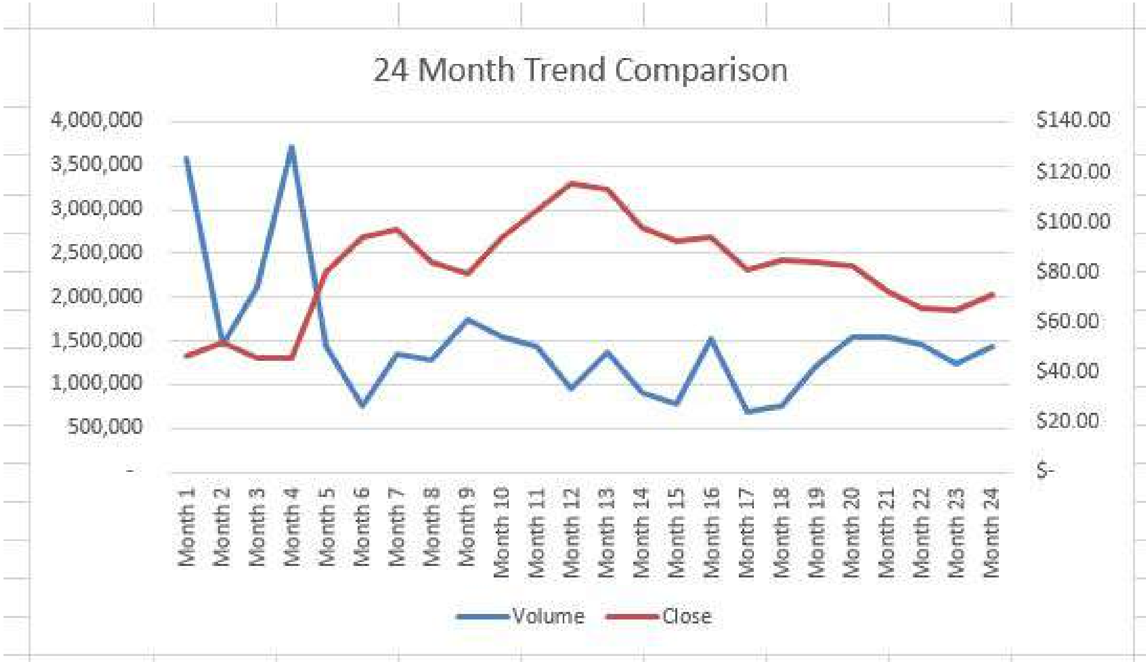


Figure 4.9 Final Comparison Line Chart

Skill Refresher

X and Y-Axis Number Formats

1. Double click anywhere along the X or Y axis to activate it.
2. Click Number from the list of options on the left side of the Format Axis dialog box.
3. Select a number format and set decimal places on the right side of the Format Axis dialog box.
4. Click the Close button in the Format Axis pane.

Frequency Distribution: Column Chart 1

A column chart is commonly used to show trends over time, as long as the data are limited to approximately twenty points or less. A common use for column charts is frequency distributions. A frequency distribution shows the number of occurrences by established categories.

For example, a common frequency distribution used in most academic institutions is a grade distribution. A grade distribution shows the number of students that achieve each level of a typical grading scale (A, A–, B+, B, etc.). The Grade Distribution worksheet contains final grades for some hypothetical Excel classes.

To show the grade frequency distribution for all the Excel classes in that year, the **Numbers of Students** appear on the Y-axis and the **Grade Categories** appear on the X-axis. In this situation, notice we do not select the **Total** row. The totals are a representation of all data and would skew the graph. Essentially you would be graphing the information twice. If you want to display the totals in a chart, the best approach is to create a separate chart that only displays the total values.

The following steps to create the column chart:

1. Select the Grade Distribution worksheet.
2. In Row3, replace the red text at states **[Insert Current Year]** and replace it with the actual current academic term and year.
3. Select two non-adjacent columns by selecting **A3:A8**.
4. Press, and hold down the **Crtl key**.

   Mac Users: Hold down the **Command key** instead.
5. Without letting go of the **Ctrl key**, select **C3:C8**
6. From the ribbon click the **Insert** tab. Choose the **Column** button.
7. Select the **Clustered Column** format. (First option listed.)
8. Click and drag the chart so the upper left corner is in the middle of cell **H2**. Resize the graph to fit in the area of H2: O13.
9. Click any cell location on the **Grade Distribution** worksheet to deactivate the chart.
10. Save your work.

Figure 4.10 shows the completed grade frequency distribution chart. By looking at the chart, you can immediately see that the greatest number of students earned a final grade in the B+ to B– range.

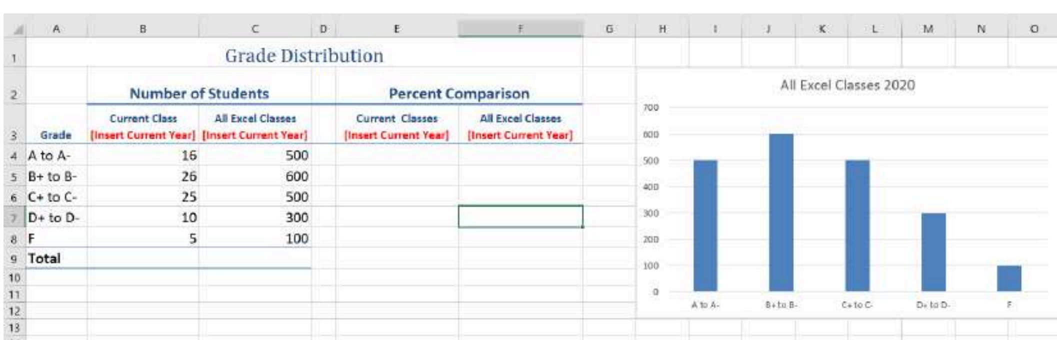


Figure 4.10 Grade Frequency Distribution Chart

Why?

Column Chart vs. Bar Chart

When using charts to show frequency distributions, the difference between a column chart and a bar chart is really a matter of preference. Both are very effective in showing frequency distributions. However, if you are showing a trend over a period of time, a column chart is preferred over a bar chart. This is because a period of time is typically shown horizontally, with the oldest date on the far left and the newest date on the far right. Therefore, the descriptive categories for the chart would have to fall on the horizontal – or category axis, which is the configuration of a column chart. On a bar chart, the descriptive categories are displayed on the vertical axis.

Creating a Chart Sheet

The charts we have created up to this point have been added, or embedded in, an existing worksheet. Charts can also be placed in a dedicated worksheet called a chart sheet. It is called a chart sheet because it can only contain an Excel chart. Chart sheets are useful if you need to create several charts using the data in a single worksheet. If you embed several charts in one worksheet, it can be cumbersome to navigate and browse through the charts. It is easier to browse through charts when they are moved to a chart sheet because a separate sheet tab is added to the workbook for each chart. The following steps explain how to move the grade frequency distribution chart to a dedicated chart sheet:

1. Click anywhere on the **Final Grades for All Excel Classes** chart on the Grade Distribution worksheet.
2. From the Chart Tools Design tab. Select **Move Chart .** This opens the Move Chart Dialog box.
3. Click the **New sheet** option on the Move Chart dialog box.
4. The entry in the input box for assigning a name to the chart sheet tab should automatically be highlighted once you click the New sheet option. Type **All Excel Classes.** This replaces the generic name in the input box (see **Figure 4.11**).
5. Click the **OK** button at the bottom of the Move Chart dialog box. This adds a new chart sheet to the

workbook with the name **All Excel Classes**.

6. Save your work.

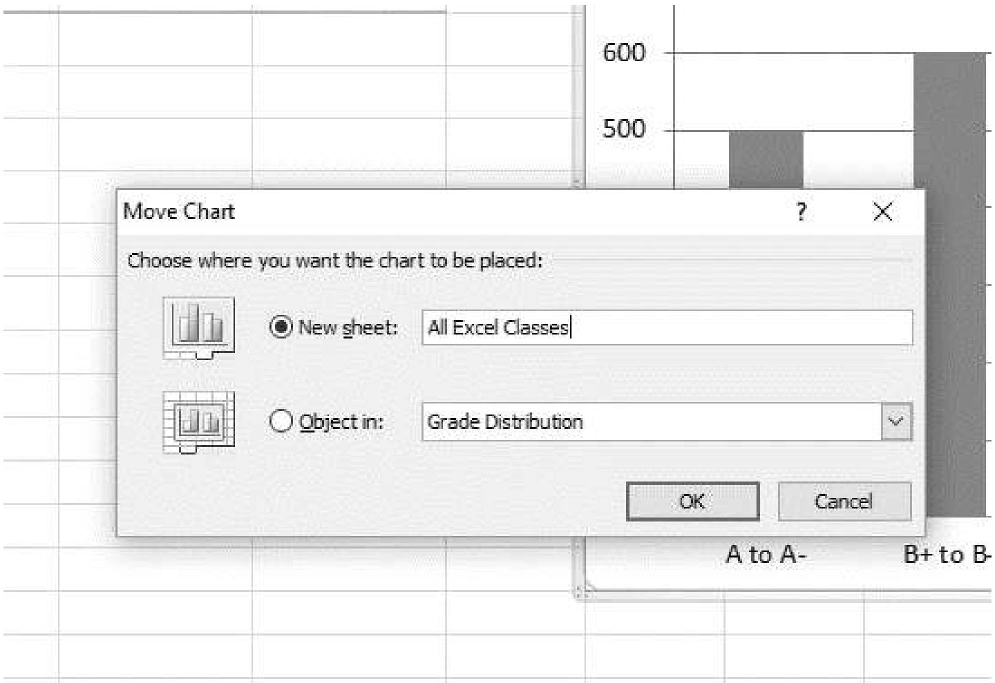


Figure 4.11 Move Chart

Figure 4.12 shows the Final Grades for all the Excel Classes column chart is in a separate chart sheet. Notice the new worksheet tab added to the workbook matches the New sheet name entered into the Move Chart dialog box. Since the chart is moved to a separate chart sheet, it no longer is displayed in the Grade Distribution worksheet.

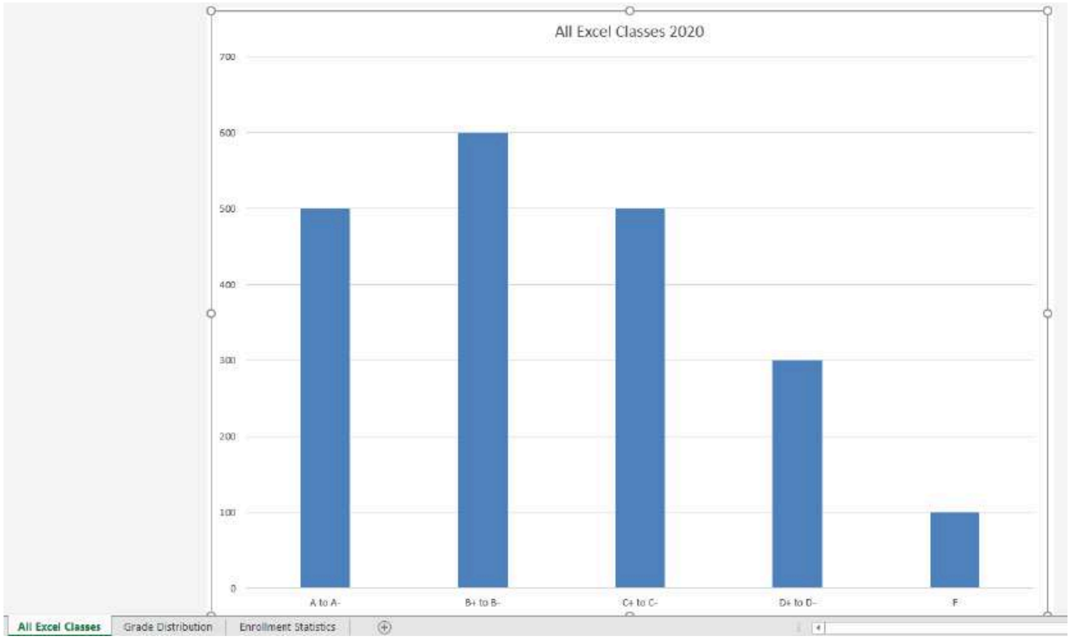


Figure 4.12 Chart Sheet Added to the Workbook

Frequency Comparison: Column Chart 2

We will create a second column chart to show a comparison between two frequency distributions. Column B on the Grade Distribution worksheet contains data showing the number of students who received grades within each category for the Current Excel Class Class. We will use a column chart to compare the grade distribution for the current class (Column B) with the overall grade distribution for Excel courses for the whole year (Column C).

However, since the number of students in the term is significantly different from the total number of students in the year, we must calculate percentages in order to make an effective comparison. The following steps explain how to calculate the percentages:

1. Highlight the range **B4:C9** on the **Grade Distribution** worksheet.
2. Click the **AutoSum** button in the Editing group of commands on the Home tab of the ribbon. This automatically sums the values in the selected range.
3. Select cell **E4**. Enter a formula that divides the value in cell **B4** by the total in cell **B9**. Add an absolute reference to cell **B9** in the formula =**B4/$B$9**. Autofill the formula down to cell **E8.**
4. Select cell **F4** . Enter a formula that divides the value in cell **C4** by the total in cell **C9**. Add an absolute reference to cell **C9** in the formula =**C4/$C$9**.
5. Autofill the down to **F8**.
6. Select **A3:A8,** press and hold down the **Ctrl** key and select **E3:F8.**

   Mac Users: Hold down the **Command key**
7. Click the **Insert** tab of the ribbon.

8. Select the **Column** button. Select the first option from the drop-down list of chart formats, which is the **Clustered Column**.
9. Click and drag the chart so the upper left corner is in the middle of cell **H2**.
10. Resize the chart to the approximate area of H2:N12.
11. Change the chart title to **Grade Distribution Comparison**. If you do not have a chart title, you can add one. On the **Design** tab, select **Add Chart Element**. Find the **Chart Title**. Select the **Above Chart** option from the drop-down list.
12. Save your work.

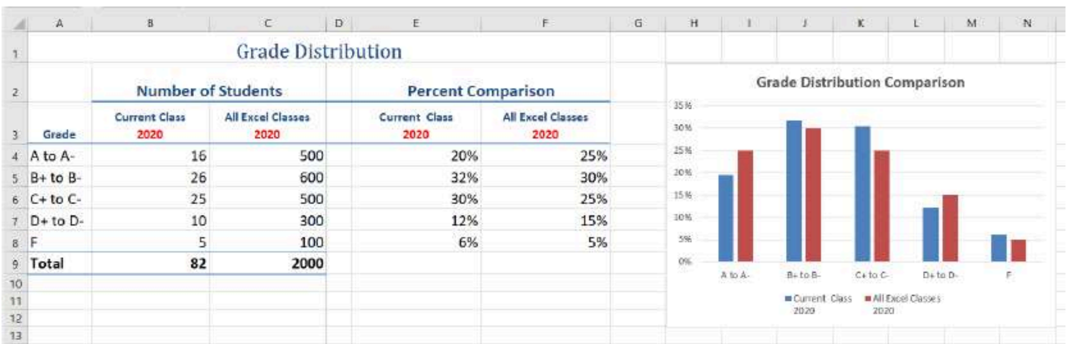


Figure 4.13 Completed Data Series for the Class Grade Distribution

Figure 4.13 shows the final appearance of the column chart. The column chart is an appropriate type for this data as there are fewer than twenty data points and we can easily see the comparison for each category. An audience can quickly see that the class issued fewer As compared to the college. However, the class had more Bs and Cs compared with the college population.

Integrity Check

Too Many Bars on a Column Chart?

Although there is no specific limit for the number of bars you should use on a column chart, a general rule of thumb is twenty bars or less.

Map Charts

Data visualization brings more depth in how information, in this case geographically, connects. You can use a map chart to compare values and show categories across geographical regions like countries/regions, states, counties or postal codes. Excel will automatically convert data to geographical locations and will display values on a map. As shown below, in **Figure 4.14,** in the next steps we will compare West Coast Community College enrollments for Fall of 2019 using a map chart.

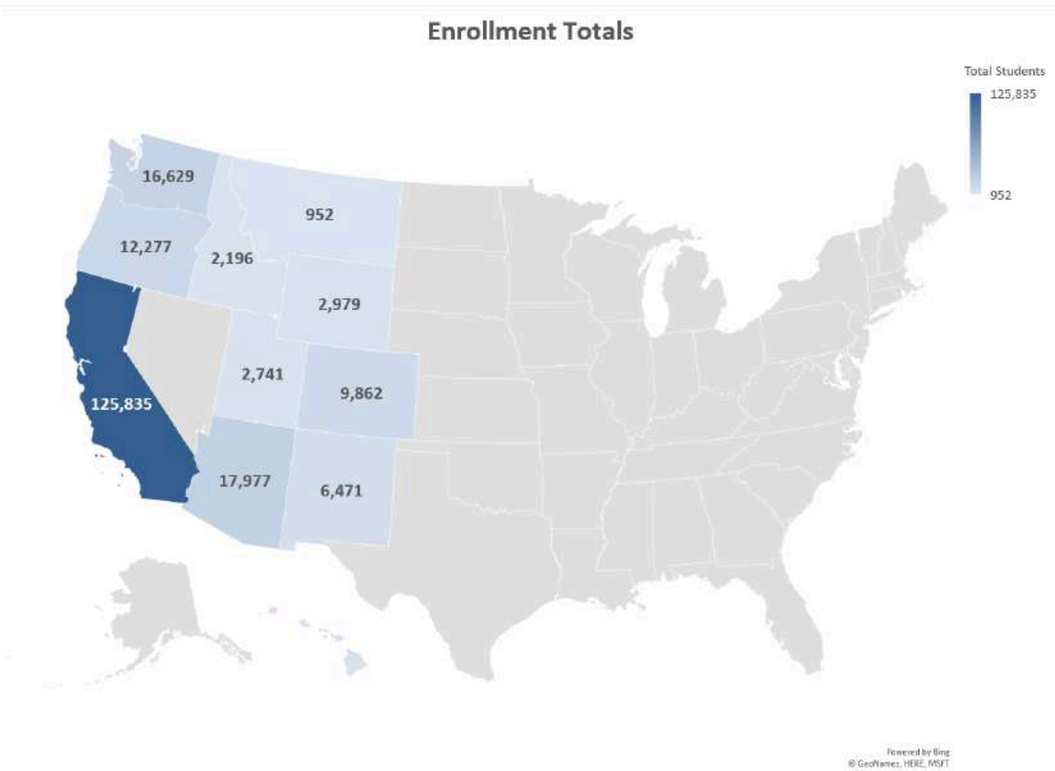


Figure 4.14 Map Chart Solution

1. From the **Enrollment Statistics** worksheet, select **A3:A13.** Next, press and hold the **CTRL** key and select **C3:C13**.
2. Click the Insert tab on the Ribbon.
3. Click **Maps**, and choose the **Filled Map** option.
4. From the Charts Design Tab, choose **Move Chart**, and select the **New Sheet** option. In the name box type **Map**. Click **Ok**.
5. To make sure the map and data are vibrant and will stand out in a presentation make the following changes:

a) Select the **Title.** Type **Enrollment Totals.** Change the font to **bold**, **size 18**.
b) From the top right corner of the Chart area, choose the Charts Elements plus sign.

c) Select the Data Labels checkbox. Notice the values appear on each State.

Mac Users: there is no "Charts Element plus sign". Follow the alternate steps below.

Click the **"Chart Design"** tab on the Ribbon

Click the **"Add Chart Element"** button on the Ribbon

Point to **"Data Labels"** option and click **"Show"**

d) Save your work.

FUNNEL CHARTS

Another graph to visualize data is a Funnel chart. Funnel charts provide a visual snapshot of a process. From our data, we will create a Funnel Chart to show how many students we have in the admissions process. You can quickly review the funnel chart to see admissions predicts to have 932 new enrolled students for Winter Term 2020.

Insert a Funnel chart by following the below steps.

1. From the **Admissions** sheet, select **A3:B7**.
2. From the Insert tab, choose **Recommended Charts**.
3. Scroll down the list and select the **Funnel Chart**. Click **OK**.
4. Move and resize the graph to approximately fit in the range of **D1:K15.**
5. From the Design Tab, click the **6th style option** provided in the styles gallery. (The background is black.) It is okay if your text color is different than the figure below.
6. Change the Chart Title to **Admissions Pipeline Winter 2020.**
7. From the Charts Elements, turn off the **Legend**.

   Mac Users: Click the "**Add Chart Element**" button and change the "**Legend**" option to "**None**"
8. Save your work.

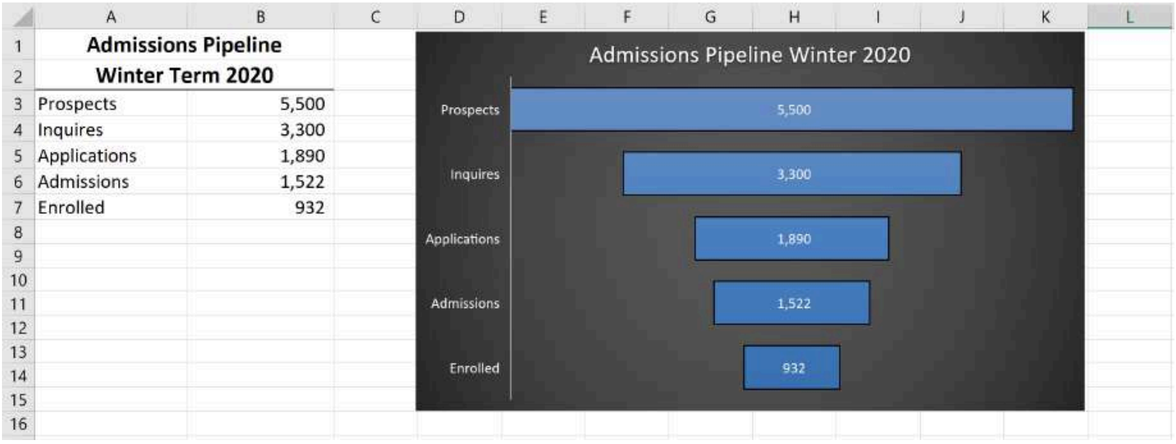


Figure 4.15 Funnel Chart

PERCENT OF TOTAL: DOUGHNUT PIE CHART

The next chart we will demonstrate is a pie chart. A pie chart is used to show a percent of the total for a data set at a specific point in time. Using the Doughnut Pie Chart, show the percentage of students enrolled at a full-time status. As in the last example, the data is located on the Enrollment Statistics sheet.

1. From the **Enrollment Statistics** worksheet, select **A3:A13.** Next, press and hold the **CTRL** key and select **D3:D13**.
2. Click the **Insert** tab on the ribbon.
3. Click the **Pie** button in the **Charts** group of commands.
4. Select the "**Doughnut**" option from the drop-down list of options.
5. From the Charts Design Tab, choose Move Chart, and select the New Sheet option. In the name box type **Full-Time Students**. Click **Ok**.
6. From the Chart Tools Design Tab, click the Chart Styles gallery and apply **Style 7**.
7. On the Chart, expand the Charts Elements tools. Select the **Legend** Choose to display the legend on the **Right**.

   Mac Users: Click "Add Chart Element" button, point to **Legend** and choose **"Right"**
8. From the Charts Elements, select the Data Labels checkbox. Then, drop down the Data Labels menu. Choose, **More Options**.

   Mac Users: Click "Add Chart Element" button, point to **Data Labels** and choose **"More Data Label Options"**

Figure 4.15
Doughnut Pie
Chart Elements

9. From the Format Data Label Options menu, select **Percentages**, and **Deselect Values** to show the percent of total students that are enrolled at a full-time status.
10. Close the Format Data Labels menu.

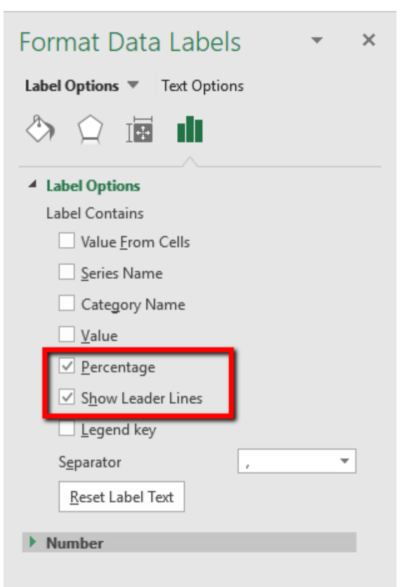


Figure 4.16
Doughnut Pie
Format Data
Labels

Notice the font is small compared to the graph size. Adjust the font size of the Title, Legend, and Data Label by following the below steps:

1. Select the Title. Change the font to **bold**, **size 18**.
2. Select the Legend. Change the font to **bold, size 14.**
3. Select one of the data labels in a doughnut wedge. Notice now all labels are selected. Change the font **bold, size 12.**

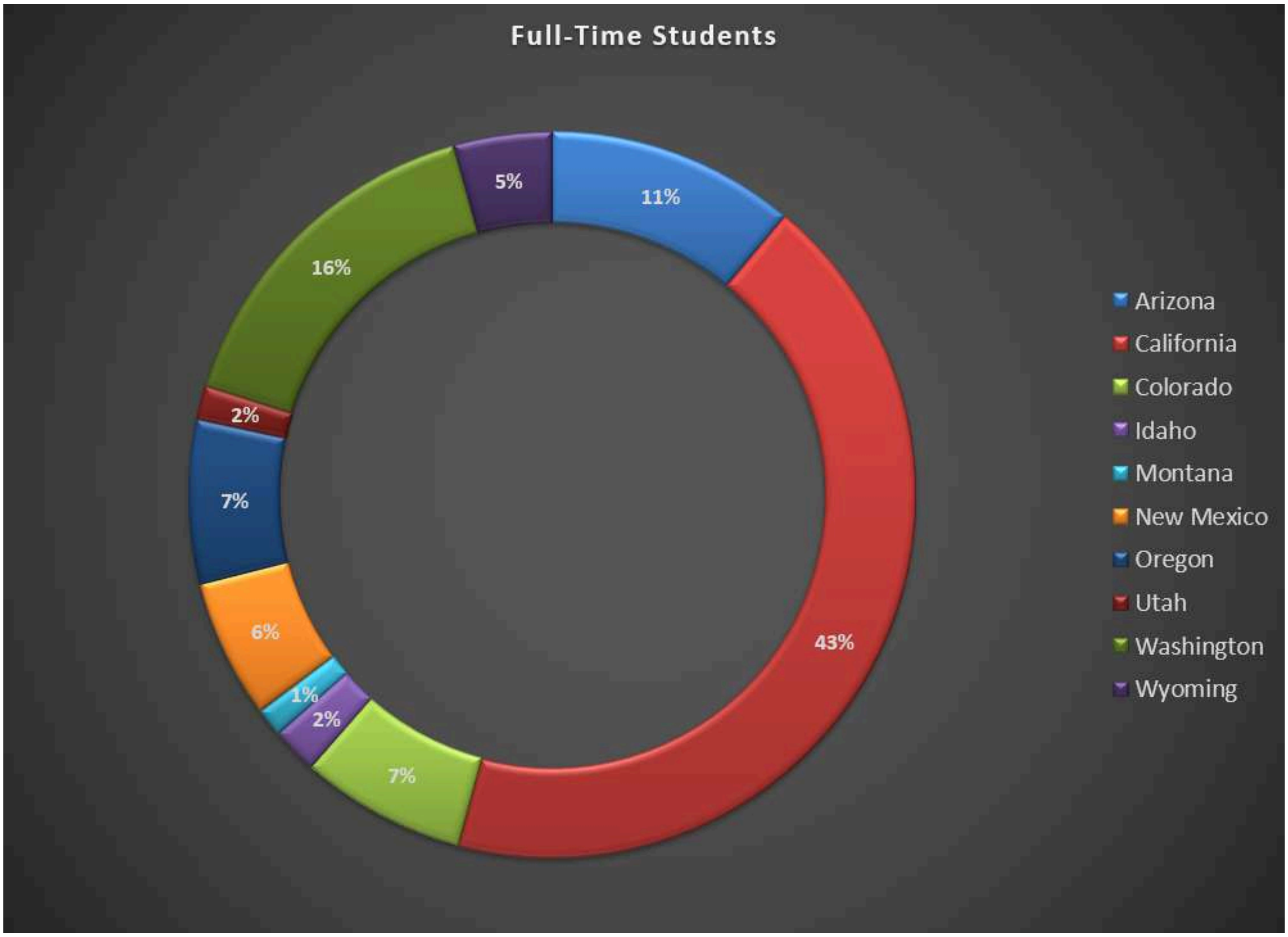


Figure 4.18 Doughnut Pie Solution

Skill Refresher

Inserting a Pie Chart

1. Highlight a range of cells that contain the data you will use to create the chart.
2. Click the Insert tab of the ribbon.
3. Click the Pie button in the Charts group.

4. Select a format option from the Pie Chart drop-down menu.

Bar chart vs Column chart

We will statistical data to compare a bar and column chart. Both the Bar and the Column chart display data using rectangular bars where the length of the bar is proportional to the data value. Both charts are used to compare two or more values. However, the difference lies in their orientation. A bar chart is oriented horizontally whereas the column chart is oriented vertically. Although alike, they cannot be always used interchangeably. The difference in their orientation, meaning typically the more data values the harder it is to read in a column format. This is where visually a bar chart would be a better choice. Complete the below steps to insert both a bar and column chart comparing not only the gender and age differences of enrolled students but the type of graphs you are viewing the data in.

1. From the Enrollment Statistics sheet, select **A3:A13**. Press and hold the CTRL key and select **G3:H13**.
2. From the Insert tab, choose **Recommended Charts**. Scroll down and select the **Stacked Bar** chart option. Click **Ok**.
3. Move and resize the graph so it fits approximately in **L1:U13**.
4. Add a chart title. Type **Age Comparison**.
5. Notice the age difference. Currently, per State, the majority of students are under 20 years old. **Save your work**.

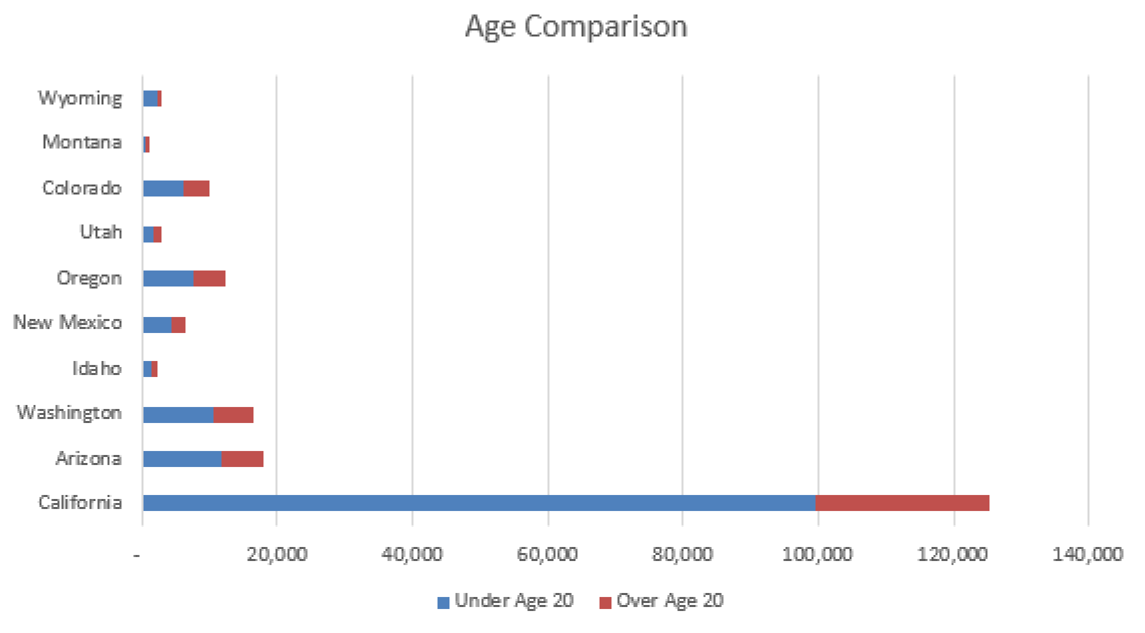


Figure 4.19 Stacked Bar chart solutions

Next, insert a column chart comparing gender.

1. From the Enrollment Statistics sheet, select **A3:A13**. Press and hold the CTRL key and select I**3:J13**.
2. From the Insert tab, choose **Recommended Charts**. Choose the first option, Clustered Column chart.
3. Move and resize the graph so it fits approximately in **L15:U32**.
4. Add a chart title. Type **Gender Comparison**.
5. Notice the ratio of women and men enrolled are pretty equal per State. **Save your work**.

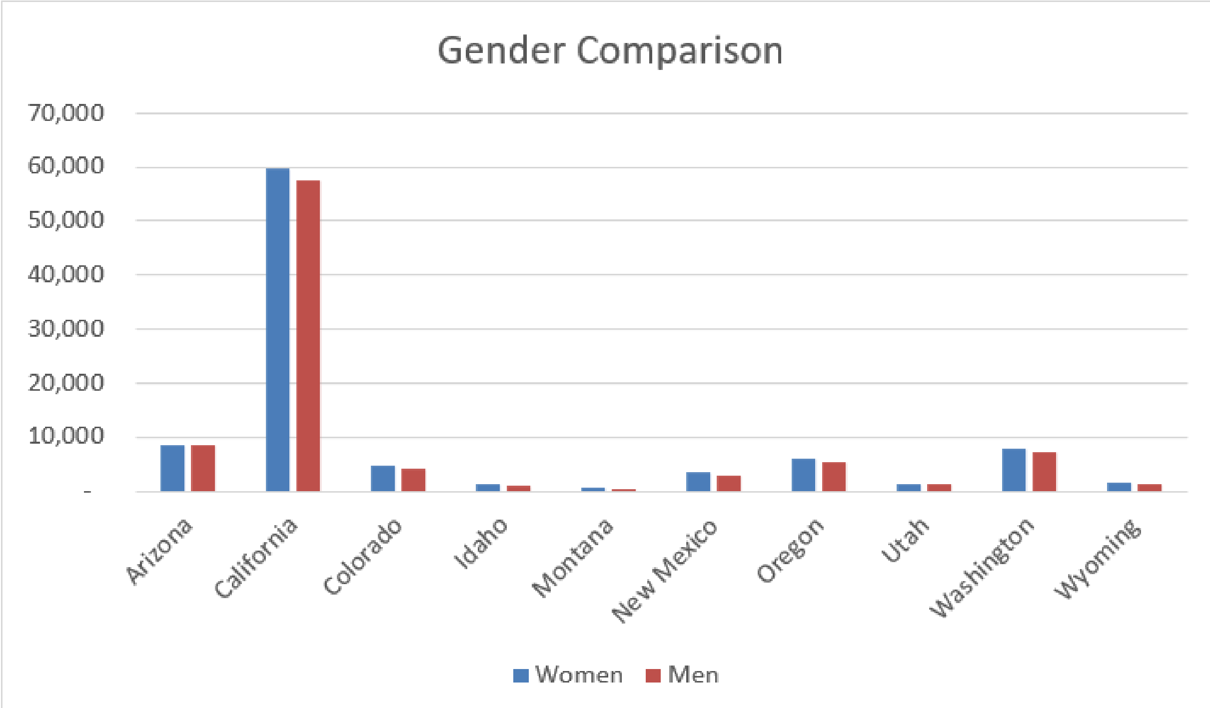


Figure 4.20 Column Chart Age Comparison

Stacked Column Chart

The last chart types we will demonstrate is the stacked column chart and a bar chart. You will use a stacked column chart to show differences in budgeted expense accounts for the admissions department and a bar chart for age comparisons of enrolled students at the college.

Follow the below steps to insert a stacked column chart.

1. Click the Expenses sheet. Select the range **A4:G9.**
2. Click the **Insert** tab of the ribbon.
3. Click the **Column** button in the **Charts** group of commands. Select the **3D– Stacked Column** format.
4. Change the Chart Title to **Expenses**.
5. Move the Chart to a New Sheet. In the name box type Budget.
6. Save your work.

Figure 4.21 shows the final stacked column chart.

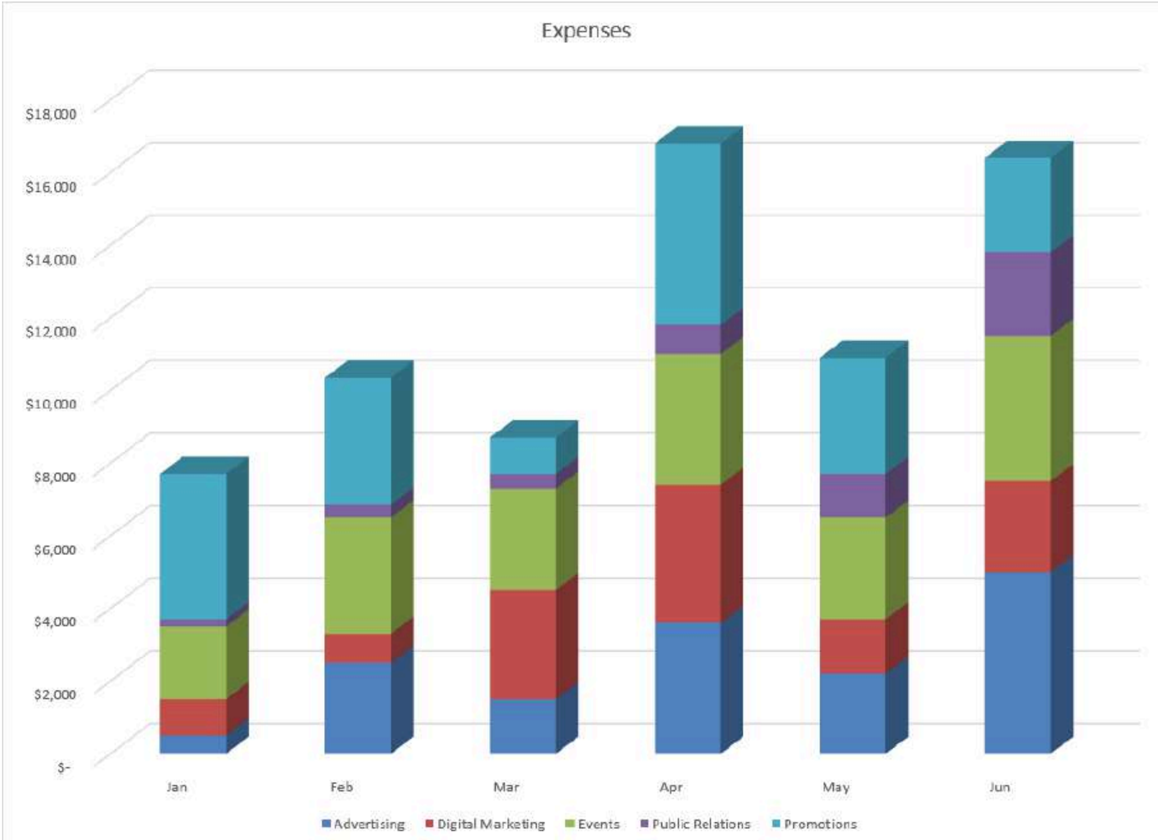


Figure 4.21 Stacked Column Chart

Skill Refresher:

Inserting a Stacked Column Chart

1. Highlight a range of cells that contain data that will be used to create the chart.
2. Click the Insert tab of the ribbon.
3. Click the Column button in the Charts group.
4. Select the Stacked Column format option from the Column Chart drop-down menu to show the values of each category on the Y-axis. Select the Stacked Column option to show each category on the Y-axis.

Key Takeaways

- Identifying the message you wish to convey to an audience is a critical first step in creating an Excel chart.
- Both a column chart and a line chart can be used to present a trend over a period of time. However, a line chart is preferred over a column chart when presenting data over long periods of time.
- The number of bars on a column chart should be limited to approximately twenty bars or less.
- When creating a chart to compare trends, the values for each data series must be within a reasonable range. If there is a wide variance between the values in the two data series (two times or more), the percent change should be calculated with respect to the first data point for each series.
- When working with frequency distributions, the use of a column chart or a bar chart is a matter of preference. However, a column chart is preferred when working with a trend over a period of time.
- A pie chart is used to present the percent of total for a data set.
- A stacked column chart is used to show how a percent total changes over time.

Attribution

Adapted by Noreen Brown from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

3.XLSX.2 FORMATTING CHARTS

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Learning Objectives

1. Apply formatting commands to the X and Y axes.
2. Assign titles to the X and Y axes that clarify labels and numeric values for the reader.
3. Apply labels and formatting techniques to the data series in the plot area of a chart.
4. Apply formatting commands to the chart area and the plot area of a chart.

You can use a variety of formatting techniques to enhance the appearance of a chart once you have created it. Formatting commands are applied to a chart for the same reason they are applied to a worksheet: they make the chart easier to read. However, formatting techniques also help you qualify and explain the data in a chart. For example, you can add footnotes explaining the data source as well as notes that clarify the type of numbers being presented (i.e., if the numbers in a chart are truncated, you can state whether they are in thousands, millions, etc.). These notes are also helpful in answering questions if you are using charts in a live presentation.

X and Y-Axis Formats

There are numerous formatting commands we can apply to the X and Y axes of a chart. Although adjusting the font size, style, and color are common, many more options are available through the Format Axis pane. The following steps demonstrate a few of these formatting techniques on the **Grade Distribution Comparison** chart. Follow the below steps to make some changes to the percentage numbers on the Y (vertical) axis.

1. In the **Grade Distribution** worksheet, click on the **Grade Distribution Comparison** chart. Double-click the vertical (value) axis. This opens the **Format Axis** pane.
2. Select **Axis Options**. Change the **Minimum Bound** to **.05** to make the differences in the columns more dramatic.
3. Click the **Close** button at the top of the Format Axis pane.

4. Save your work.

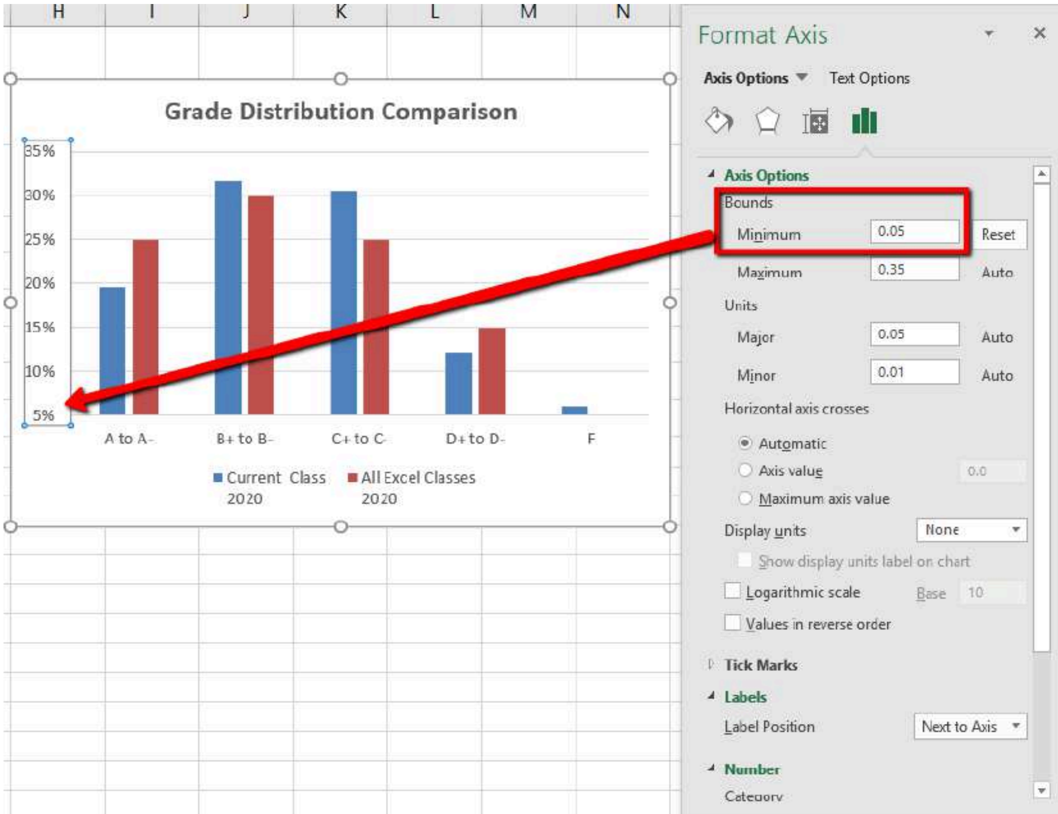


Figure 4.22 Format Axis Pane Changes

X and Y-Axis Titles

Titles for the X and Y axes are necessary for defining the numbers and categories presented on a chart. For example, by looking at the Grade Distribution Comparison chart, it is not clear what the percentages along the Y-axis represent. The following steps explain how to add titles to the X and Y axes to define these numbers and categories:

1. Click anywhere on the Grade Distribution Comparison chart in the Grade Distribution worksheet to activate it.
2. In the upper right corner of the graph, choose the Charts Element plus sign. Select the Axis Titles, then Primary Horizontal and Primary Vertical. This inserts the place holders that you will type text in.
   Mac Users click the "Add Chart Element" button in the Design tab, point to "Axis Titles" and click on "Primary Horizontal". Do this one more time and click on "Primary Vertical".
3. Click at the beginning of the Y-axis title and delete the generic title. Type **Percent of Enrolled Excel Students**.

4. Click at the beginning of the X-axis title and delete the generic title. Type **Final Course Grade**.

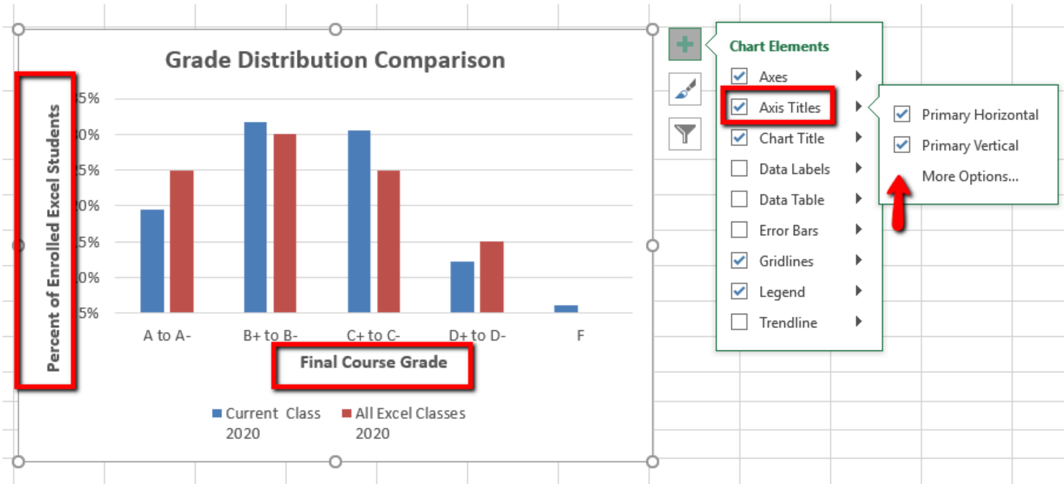


Figure 4.23 Selecting a Title for the X & Y-Axis

Skill Refresher

X and Y Axis Titles

1. Click anywhere on the chart to activate it. Choose to open the Charts Element menu.
2. Select one of the options from the second drop-down list.
3. Click in the axis title to remove the generic title and type a new title.

Data Series Labels and Formats

Adding labels to the data series of a chart is a key formatting feature. A data series is an item that is being displayed graphically on a chart. For example, the blue bars on the Grade Distribution Comparison chart represent one data series. We can add labels at the end of each bar to show the exact percentage the bar represents. In addition, we can add other formatting enhancements to the data series, such as changing the color of the bars or adding an effect. The following steps explain how to add these labels and formats to the chart:

1. Click on any of the **red columns** representing the **All Excel Classes** data series, then **Right-Click** to open the menu.

   Mac Users should hold down the **CTRL key** and click on any of the red columns.
2. From the menu, select **Format Data Series.**
3. From the **Format Data Series** pane, click the **Fill and Line** (paint bucket) button to bring up the Fill and Border group of commands.
4. Click the word **Fill** (if needed) to expand the list of Fill options.
5. Select **Pattern Fill**. Then select **40%** (last option in the top row). Change the Foreground to white, and the Background to Red.
6. Close the **Format Data Series** pane.

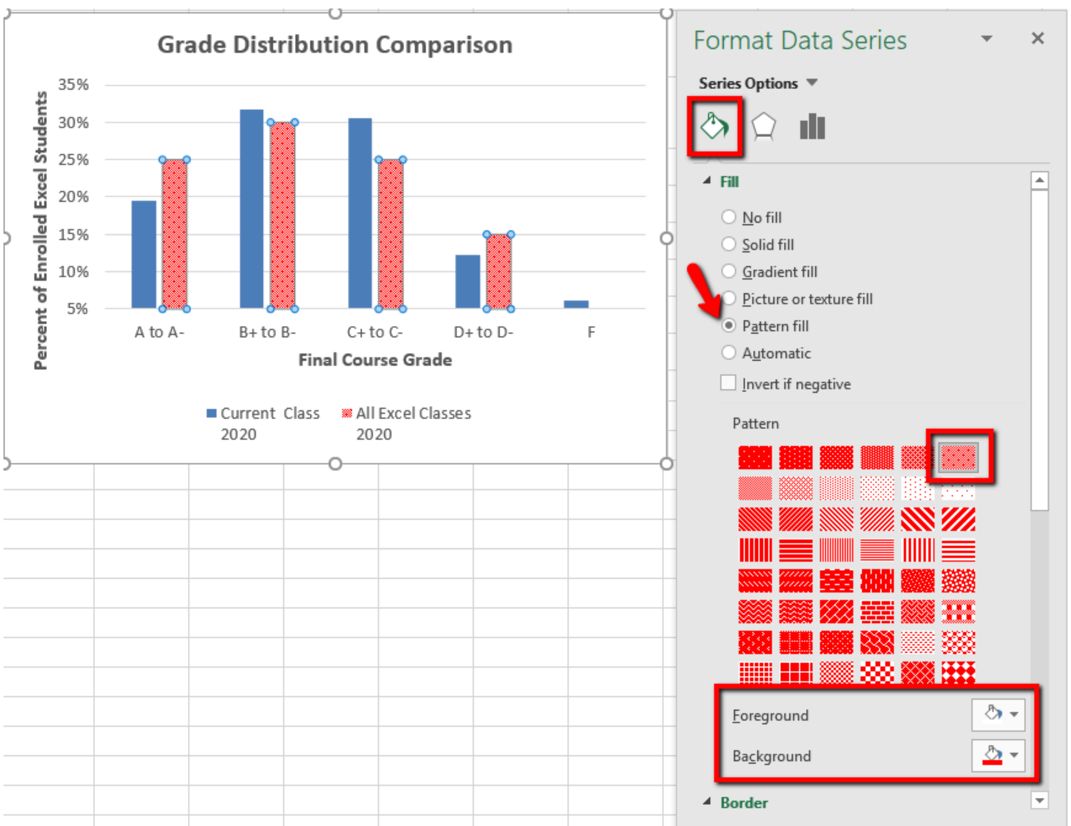


Figure 4.24 Changing the Fill of a Data Series

Now we are going to add the Data Labels at the end of the columns.

1. Be sure that your entire chart is selected, not just one of the data series. Click the **Design** tab in the **Chart Tools** section of the ribbon.
2. On the **Design** tab select the **Add Chart Element** button, then **Data Labels**, then **Outside End** (see

Figure 4.25.)

3. Click on one of the **Data Labels**. Note that all of the data labels for that data series are selected.
4. Check the spelling on all of the worksheets and make any necessary changes. Save your work.

Figure 4.25 shows the Grade Distribution Comparison chart with the completed formatting adjustments and labels added to the data series. Note that we can move each individual data label. This might be necessary if two data labels overlap or if a data label falls in the middle of a grid line. To move an individual data label, click it twice, then click and drag.

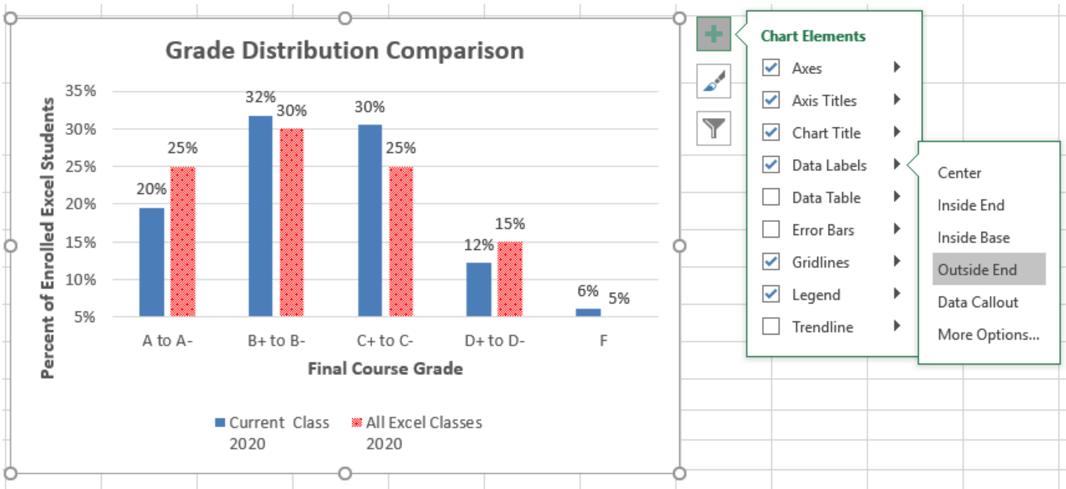


Figure 4.25 Data Labels Outside End

Skill Refresher:

Adding Data Labels

1. Click anywhere on the chart to activate it.
2. Open the **Add Chart Element** group.
3. Then, select **Data Labels**
4. Select one of the preset positions from the drop-down list.

Key Takeaways

- Applying appropriate formatting techniques is critical for making a chart easier to read.
- Many formatting commands in the Home tab of the ribbon can be applied to a chart.
- To change the number format for an axis or data label, you must use the Number section in the Format Data Labels dialog box. You cannot use the Number format commands in the Home tab of the ribbon.
- Axis titles help the reader sees the most accurate representation of the information presented on a chart.

Attribution

Adapted by Hallie Puncochar and Noreen Brown from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

3.XLSX.3 USING CHARTS WITH MICROSOFT® WORD® AND MICROSOFT® POWERPOINT®

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Learning Objectives

1. Learn how to paste an image of an Excel chart into a Word document.
2. Learn how to paste a link to an Excel chart into a PowerPoint slide.

Charts that are created in Excel are commonly used in Microsoft Word documents or for presentations that use Microsoft PowerPoint slides. Excel provides options for pasting an image of a chart into either a Word document or a PowerPoint slide. You can also establish a link to your Excel charts so that if you change the data in your Excel file, it is automatically reflected in your Word or PowerPoint files. We will demonstrate both methods in this section.

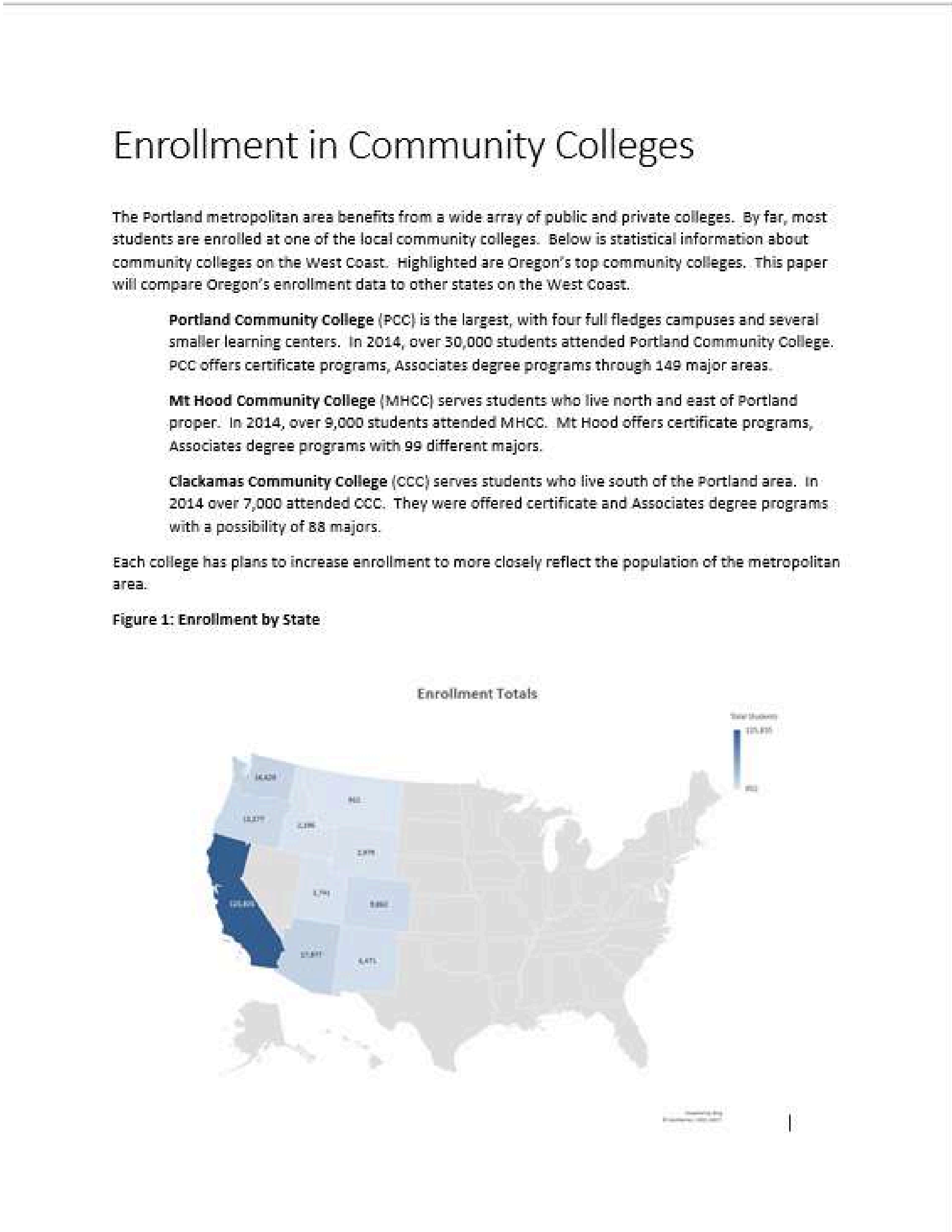
Pasting a Chart Image into Word

For this exercise you will need **two** files:

- *The Excel spreadsheet you have been working on this chapter — **CH4 Charting**.*
- *A Word document data file — CH4 CC Enrollment*

Excel charts can be valuable tools for explaining quantitative data in a written report. Reports that address business plans, public policies, budgets, and so. For this example, we will assume that the total enrollment per state from the Enrollment Statistics Map chart is being used in a student's written report. (see **Figure 4.26**). The following steps demonstrate how to paste an image, or picture, of this chart into a Word document:

1. Open ***CH4 CC Enrollment***. Save it as ***CH4 Enrollment Totals Per State.***
2. Move your cursor to the bottom of the document by clicking below the heading that reads: **Figure 1: Enrollment by State**. The image of the Map chart will be placed below this heading.
3. If needed, open the Excel file you have been working with (*CH4 Charting)*. Activate the **Map** chart.
4. On the **Home** tab of the ribbon. Click the **Copy button** dropdown arrow and select **Copy as Picture.**
5. Select **OK** — Accepting the Copy Pictures defaults:
    - As shown on Screen
    - Picture
6. Go back to the ***CH4 Enrollment Totals Per State*** Word document.
7. Confirm that the insertion point is below **Figure 1: Enrollment by State** heading, click the **Paste** button in the Home tab of the ribbon (or press **Crtl-V**).
8. Note the picture of the Map chart will need to be resized. Resize the image by slowly dragging one of the corner sizing handles so it's large enough to fill space below the text on the first page. Make sure the image does not spill over to the next page. This document should be one page only.
9. Save your work.



Enrollment in Community Colleges

The Portland metropolitan area benefits from a wide array of public and private colleges. By far, most students are enrolled at one of the local community colleges. Below is statistical information about community colleges on the West Coast. Highlighted are Oregon's top community colleges. This paper will compare Oregon's enrollment data to other states on the West Coast.

> **Portland Community College** (PCC) is the largest, with four full fledges campuses and several smaller learning centers. In 2014, over 30,000 students attended Portland Community College. PCC offers certificate programs, Associates degree programs through 149 major areas.
>
> **Mt Hood Community College** (MHCC) serves students who live north and east of Portland proper. In 2014, over 9,000 students attended MHCC. Mt Hood offers certificate programs, Associates degree programs with 99 different majors.
>
> **Clackamas Community College** (CCC) serves students who live south of the Portland area. In 2014 over 7,000 attended CCC. They were offered certificate and Associates degree programs with a possibility of 88 majors.

Each college has plans to increase enrollment to more closely reflect the population of the metropolitan area.

Figure 1: Enrollment by State

Figure 4.26 CH4 Enrollment Totals Per State

Skill Refresher

Pasting a Chart Image into Word

1. Activate an Excel chart and click the Copy button in the Home tab of the ribbon.
2. Click on the location in the Word document where the Excel chart will be pasted.
3. Click the down arrow of the Paste button in the Home tab of the ribbon.
4. Click the Picture option from the drop-down list.
5. Click the Format tab in the Picture Tools section of the ribbon.
6. Resize the picture by clicking the up or down arrow on the Shape Width or Shape Height buttons.

Pasting a Linked Chart Image into PowerPoint

For this exercise you will need two files:

- *The Excel spreadsheet you have been working on in this chapter — **CH4 Charting**.*
- *A PowerPoint data file – CH4 PowerPoint CC*
- Microsoft PowerPoint is perhaps the most commonly used tool for delivering live presentations. The charts used in a live presentation are critical for efficiently delivering your ideas to an audience. Similar to written documents, a wide range of presentations may require an explanation of quantitative data. This demonstration includes a PowerPoint slide that could be used in a presentation. We will paste the linked Budget chart into the PowerPoint slide. As a result, if we change the chart in the Excel file, the change will be reflected in the PowerPoint file.

1. Open ***CH4 PowerPoint CC .pptx***. Save it as **CH4 PowerPoint CC Enrollment**.
2. Navigate to **Slide 6** – **Budget To Increase Enrollment**. This is the slide where you will place the linked chart.
3. If needed, open the Excel file you have been working with (*CH4 Charting)*. Activate the **Budget** chart. Click copy, (not Copy as Picture.)
4. Go back to the *CH4 PowerPoint CC Enrollment* presentation.
5. Make sure you are still on Slide 6. Click into the empty prompt box on the right.
6. Click the **Paste** button dropdown arrow in the **Home** tab of the ribbon in the PowerPoint file, choose the Paste Option **Use Destination Theme and Link Data (L).**

Mac Users should choose "**Use Destination Theme**"

This pastes an image of the Excel chart into the PowerPoint slide yet changing the appearance to match the current theme of the PowerPoint slide.

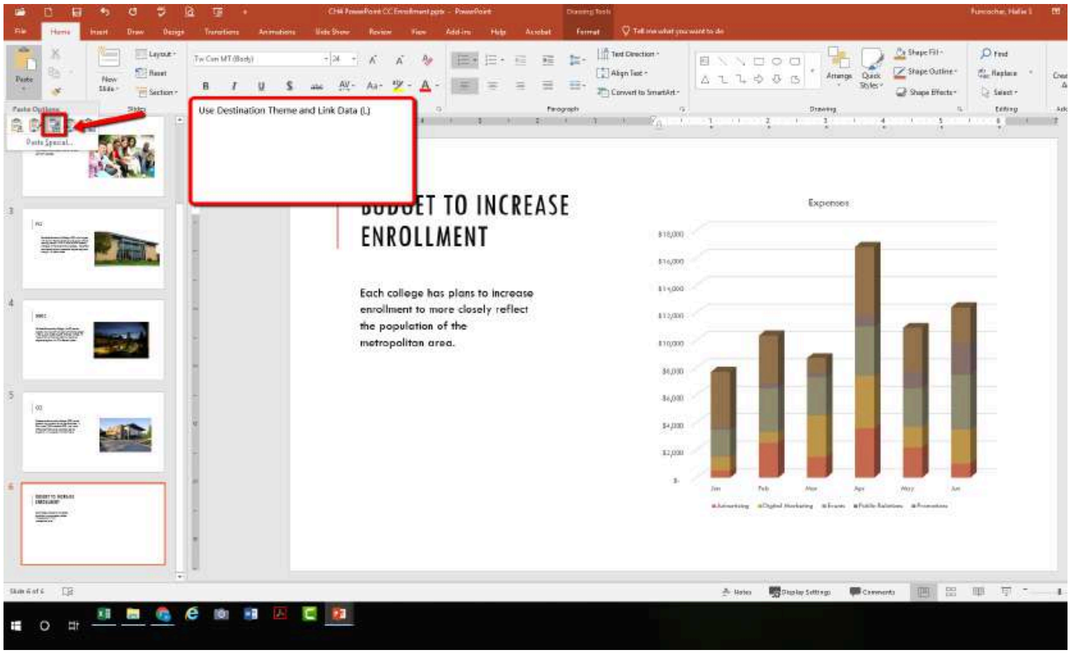


Figure 4.27 Use Destination Theme & Link_Data (L)

The benefit of adding this chart to the presentation as a link is that it will automatically update when you change the data in the linked spreadsheet file.

1. Return to your *CH4 Charting* Excel file.
2. Select the **Expenses** worksheet. The Advertising cost for June was cut. Update the spreadsheet to change. Change the value in cell G5 to **1000**.
3. Select the **Budget** worksheet. Notice how the chart has changed.
4. Return to the PowerPoint file. On Slide6, you should see the updated chart.
5. Save your work. You will submit both the **Word** and **PowerPoint** files, along with the **Excel** file, at the end of the next section.

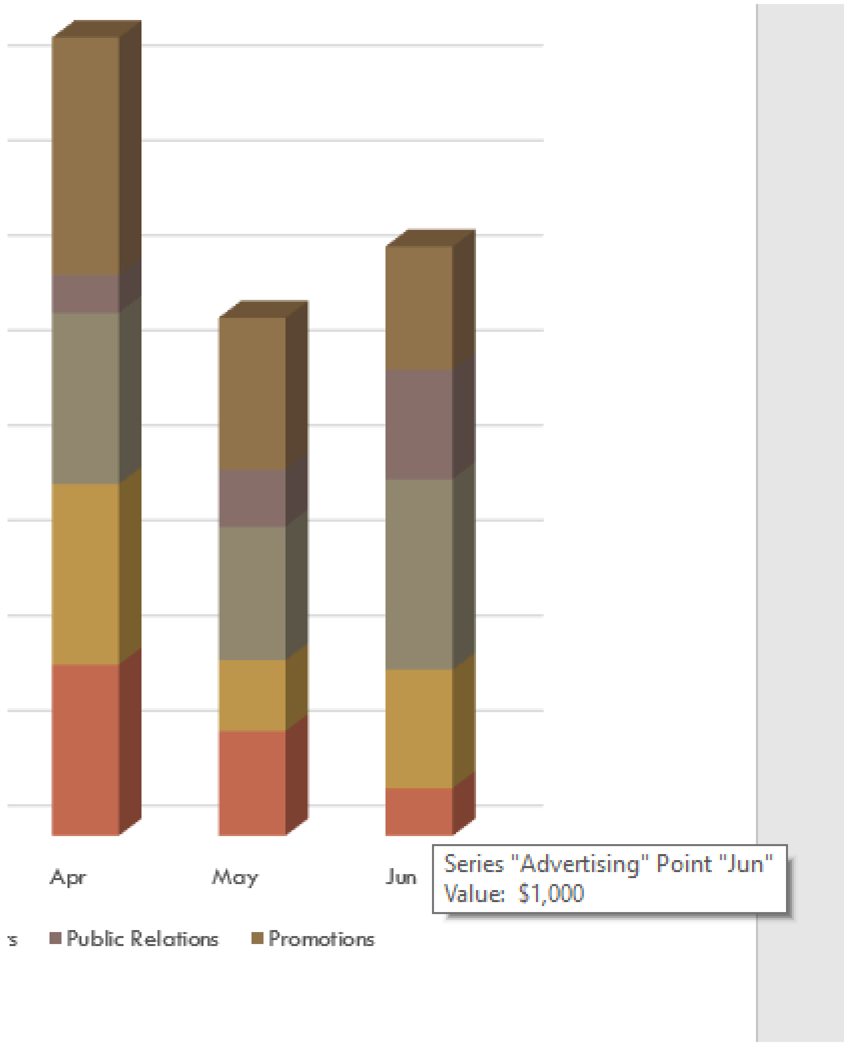


Figure 4.28
Updated June Advertising Cost

Integrity Check

Refreshing Linked Charts in PowerPoint and Word

When creating a link to a chart in Word or PowerPoint, you must refresh the data if you make any changes in the Excel workbook. This is especially true if you make changes in the Excel file prior to opening the Word or PowerPoint file that contains a link to a chart. To refresh the chart, make sure it is activated, then click the Refresh Data button in the Design tab of the ribbon. Forgetting this step can result in old or erroneous data being displayed on the chart.

Integrity Check

Severed Link?

When creating a link to an Excel chart in Word or PowerPoint, you must keep the Excel workbook in its original location on your computer or network. If you move or delete the Excel workbook, you will get an error message when you try to update the link in your Word or PowerPoint file. You will also get an error if the Excel workbook is saved on a network drive that your computer cannot access. These errors occur because the link to the Excel workbook has been severed. Therefore, if you know in advance that you will be using a USB drive to pull up your documents or presentation, move the Excel workbook to your USB drive before you establish the link in your Word or PowerPoint file.

Skill Refresher:

Pasting a Linked Chart Image into PowerPoint

1. Activate an Excel chart and click the Copy button in the Home tab of the ribbon.
2. Click in the PowerPoint slide where the Excel chart will be pasted.
3. Click the down arrow of the Paste button in the Home tab of the ribbon.
4. Click the Keep Source Formatting & Link Data option from the drop-down list.
5. Click the Refresh Data button in the Design tab of the ribbon to ensure any changes in the Excel file are reflected in the chart.

Key Takeaways

- When pasting an image of an Excel chart into a Word document or PowerPoint file, use the **Picture** option from the Paste drop-down list of options – if you want the image to act as an image. You will not be able to make any changes to the content of the picture.
- When creating a link to a chart in Word or PowerPoint, you may need to refresh the data if you make any changes in the originating spreadsheet. You should not use the **Picture** option.

Attribution

Adapted by Hallie Puncochar, and Noreen Brown from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

3.XLSX.4 PREPARING TO PRINT

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Learning Objectives

1. Review each worksheet in a workbook in Print Preview.
2. Modify worksheets as needed to professionally print data and charts.

In this section, we will take a look at each of the worksheets created in the previous sections. Since these worksheets contain a combination of data and charts, there are specific things to watch for if you will be printing the sheets.

We will start by looking at each worksheet in Print Preview in Backstage View. We will then make any changes necessary, such as changing the orientation and scaling or moving charts around on the worksheet. To make sure we don't miss any worksheets, we are going to review the worksheets in the order they appear in the tabs.

Previewing Chart Sheets for Printing

Data file: Continue with CH4 Charting.

The **All Excel Classes** is a chart sheet. This means that it does not contain any data; remember that chart sheets just contain charts. We still need to review it in Print Preview.

1. Click on the **All Excel Classes** worksheet tab.
2. Go to Print Preview by clicking Print in Backstage View.
3. Notice that the chart will print on the entire page, in Landscape orientation.
4. There is nothing to change. Exit Backstage View.

Printing Worksheets with Data and Charts

The Stock Trend worksheet has a lot of data and multiple embedded charts. We need to print the data and the charts, which will require modifications to the page setup.

1. Click on the **Stock Trend** worksheet tab.
2. Go to Print Preview by clicking Print in Backstage View.

   Mac Users choose **“File/Print…”** from the Excel File menu option.
3. Notice that this worksheet is currently printing on seven pages.
4. As you click through each page you should make the following observations:

   - The data is split between the first and third pages.
   - The line chart starts on the first page, but part of it is also on the second page.
   - The double-line chart starts on the third page and then finishes on the fifth page.
   - The fourth and sixth pages are blank.
   - The last page (page 7) has a column of seemingly random numbers.

5. Exit Backstage View.
6. The first thing we are going to do is hide the numbers that are appearing on **page 7**. We are going to hide the column, instead of deleting the numbers, in case the numbers are being utilized somewhere else in the workbook.
7. Scroll to the right on the worksheet until you find the numbers in **column AH**.
8. Click anywhere in column AH.
9. On the Home ribbon, click the **Format** button in the Cells group.
10. In the Visibility section, select **Hide & Unhide** then select **Hide Columns**.

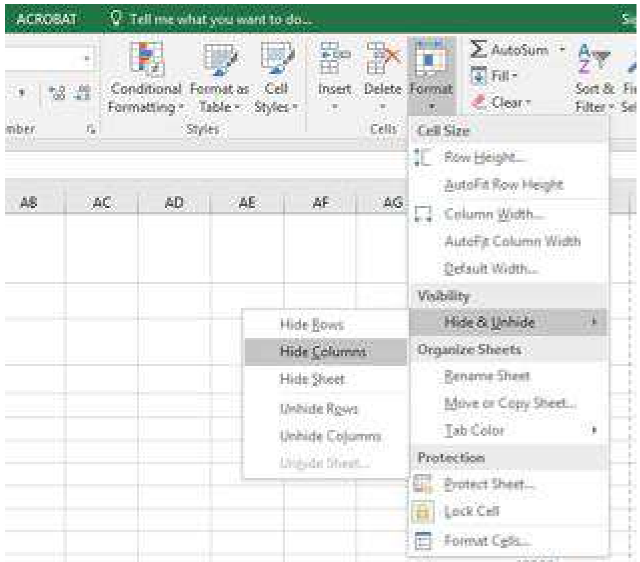


Figure 4.29 Hide Columns in Format Menu

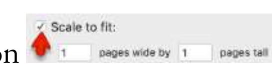
11. The visible column headings should now go from AG to AI.
12. Return to Print Preview in Backstage View to see the changes to the printed worksheet.
13. Notice that there are now five pages. The data and charts are still splitting across multiple pages, but the numbers in column AH are no longer going to print.
14. Remain in Backstage View for the next steps.

The data is still split between pages 2 and 3, and the charts are splitting oddly as well. The first step we will try to fix these issues is to change the page orientation and scaling.

1. While still in Backstage View, change the **page orientation** to Landscape (use the Orientation drop-down menu in the Settings section).

   Mac Users click the **Landscape Orientation** button 

2. This puts all of the data on one sheet, but the charts are still split between multiple pages.
3. Change the **page scaling** to Fit Sheet on One Page (use the Scaling drop-down menu in the Settings section).

   Mac Users click the **Scale to Fit** option 

4. This fits everything on one page, but it is too small to be able to read.
5. Change the page scaling back to No Scaling.

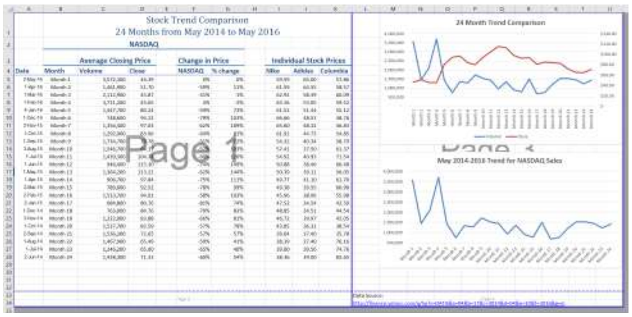
   Mac Users: uncheck the **Scale to Fit** option

The next thing we will try is moving one, or both, of the charts. In order to move the charts, we need to exit out of Backstage View.

1. Exit Backstage View.
2. Switch to the View ribbon and then select Page Break Preview. Your screen should look similar to **Figure 4.30**. (Remember that the dotted blue lines indicate automatic page breaks.)

3. Move the 24 Month Comparison (double-line) chart closer to the top of its page.
4. Move the May 2014-2015 Trend for NASDAQ Sales Volume (line chart) so that it is under the 24 Month Comparison chart.
5. The link to the data source is still at the bottom of page 2 (in A50:A51) so you need to move it as well. Using your preferred method, move the text from A50:A51 to M31:M32.

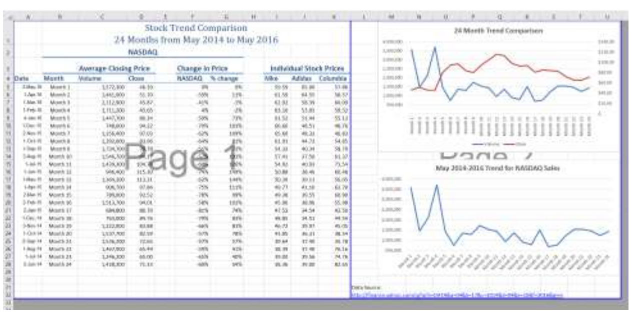
Now your screen should look similar to **Figure 4.30**.



We don't want the data source link text to print on its own page, but there is no room to move it onto the same page as the charts. To fix this, we are going to remove the automatic page break between the charts and the text in M31:M32.

1. Place your pointer on the horizontal blue dashed line (automatic page break) between the line chart and the Data Source link text.
2. When your pointer changes to the double arrow (pointing up and down), drag the page break down into the gray area. This removes the page break.
3. If your vertical automatic page break between columns K and L moves, drag it back between columns K and L. This will make it a solid blue line, which will no longer adjust automatically.

Note: you may need to slightly re-size the two charts in order to make your screen look like Figure 4.31. Your "goal" is to only have two pages.



Now you need to do one final check of this worksheet in Print Preview.

1. Go to Print Preview and look at both pages. Page 1 should contain just the data and page 2 should have both charts and the Data Source link text.
2. Exit Backstage View and save the file.

Preview Remaining Worksheets for Printing

The remaining worksheets need to be reviewed. Some of them will need minor changes and some will not need any changes. You will need to preview each one and then make the specified changes. In the following steps, you will preview and modify all other worksheets.

1. **Grade Distribution, Enrollment Statistics, and Admissions** sheets – the charts split

across two pages. Fix this by changing the orientation (Landscape) and scaling (Fit Sheet on One Page).

2. **The remaining chart sheets** should not need any changes.

Printing a Chart Only

Sometimes you might have a worksheet that has data and a chart, but you only want the chart to print. That is the case with the **Enrollment Statistics** worksheet.

1. Switch to the **Enrollment Statistics** worksheet.
2. Select the **Gender Comparision** chart.

   Mac Users: Steps 3-5 will not work in Excel for Mac. See alternate steps below step 5.

3. Go to Print Preview. Only the chart is printing. *(If it shows the data printing along with the chart, exit Backstage View and be sure to select just the chart on the worksheet.)*
4. If needed, change the orientation to Landscape. This orientation looks better when printing just a chart.
5. Exit Backstage View.

   Mac Users: the **only** way to print a Chart separately is to click on the chart you want to print move it to a new sheet by clicking on the chart, click the **Move Chart** button on the **Chart Design** tab, click New Sheet then choose **File/Print** from the Excel menu and switch to Landscape Orientation if necessary.

Hiding a Worksheet

You have actually decided that you do not want the E**xpenses** sheet to be visible at all, but you do not want to delete it. We are going to hide it from anyone looking at the workbook.

1. Right-click on the **Expenses** tab.

   Mac Users should hold down the **CTRL** key and click on the **Expenses** tab

2. Select **Hide** from the menu that appears. The sheet should no longer be visible.
3. Save the **CH4 Charting** workbook.
4. Submit all three files from this chapter: **CH4 Charting.xlsx, CH4 CC Enrollment.docx**, and **CH4 PowerPoint CC Enrollment.pptx** as directed by your instructor.

Attribution

"4.4 Preparing to Print" by Hallie Puncochar, and Julie Romey, Portland Community College is licensed under CC BY 4.0

3.XLSX.5 CHAPTER PRACTICE

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

To assess your understanding of the material covered in the chapter, please complete the following assignments.

Although Excel is primarily used in business and scientific applications, you will find it useful in other areas of study as well. In these exercises, we will use Excel to create charts using historical, and health data.

Charting Historical Data (Comprehensive Review)

Download Data File: PR4 Data

Excel is an excellent tool for helping display historical data. In this exercise, we will be examining ways to display information on minimum mage data and life expectancy.

Task1 – National Minimum Wages Changes in The United States 2019-2020

Since the beginning of the previous century, the United States has set a minimum wage, in order to set a "floor" beneath which wages cannot fall. Most states have set their own minimum wages, but none are lower than the national minimum wage. Follow the below steps to insert a Map Chart outlining what the current minimum wage is per state.

1. Open the file named **PR4 Data** and then Save As **PR4 Historical Data.**
2. On the **Minimum Wage** worksheet, select the range **B4:B55.** Press and hold the **CTRL** key and select **D4:D55**.

   Mac Users: hold down the "**Command**" key **not** the CTRL key
3. Select the **Insert** tab, then the **Map Chart** tool in the **Charts** group.
4. Move the Chart as a **New Sheet**. Rename the sheet **Map**.
5. Update the Chart Title to **US Minimum Wage 2020.**
6. From the **Charts Element** menu choose to display the **Data Labels**.
7. From the Charts Element menu, turn off the **Legend**.
8. Prepare the **Minimum Wage** worksheet for printing by changing the scaling to **Fit Sheet on One Page**.

9. Save your work.

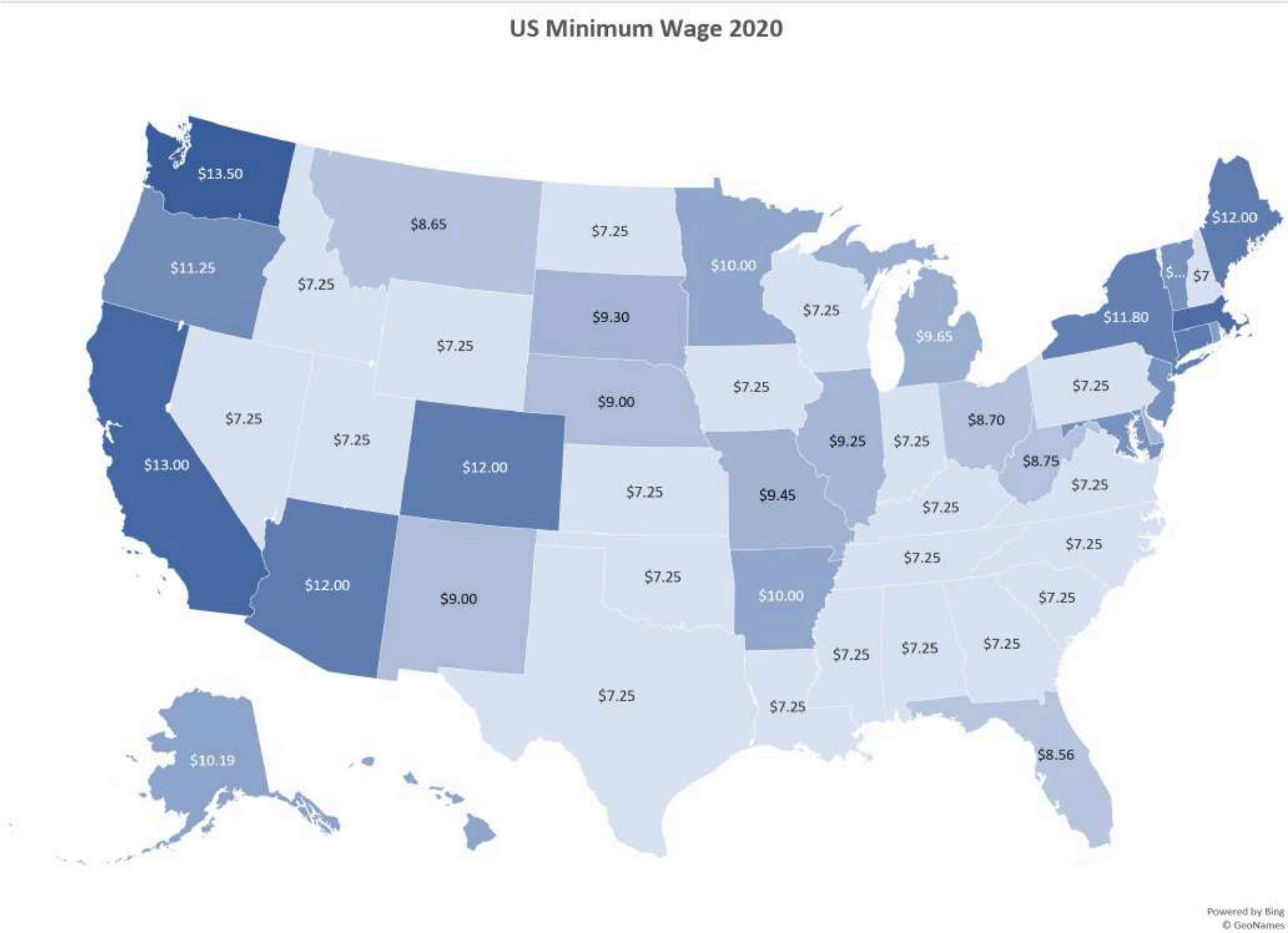


Figure 4.32 Map Chart

TASK 2 – OREGON: PROJECTED LIFE EXPECTANCY AT BIRTH

In the past 40 years, between 1970 and 2010, life expectancy for Oregon men improved by 8.7 years and for women by 5.5 years. Oregon's life expectancy has remained slightly higher than the U.S. average. The life expectancy will continue to improve for both men and women. However, the gain for men has been outpacing the gain for women. Consequently, the difference between men's and women's life expectancies has continued to shrink.

https://www.oregon.gov/das/OEA/Documents/OR_pop_trend2012.pdf

1. On the **Life Expectancy** sheet, select **A5:B11.**
2. From the Insert tab choose Recommend Charts. Select the second option, Clustered Column chart.
3. Move the chart to a new sheet. Name the sheet **Men.**

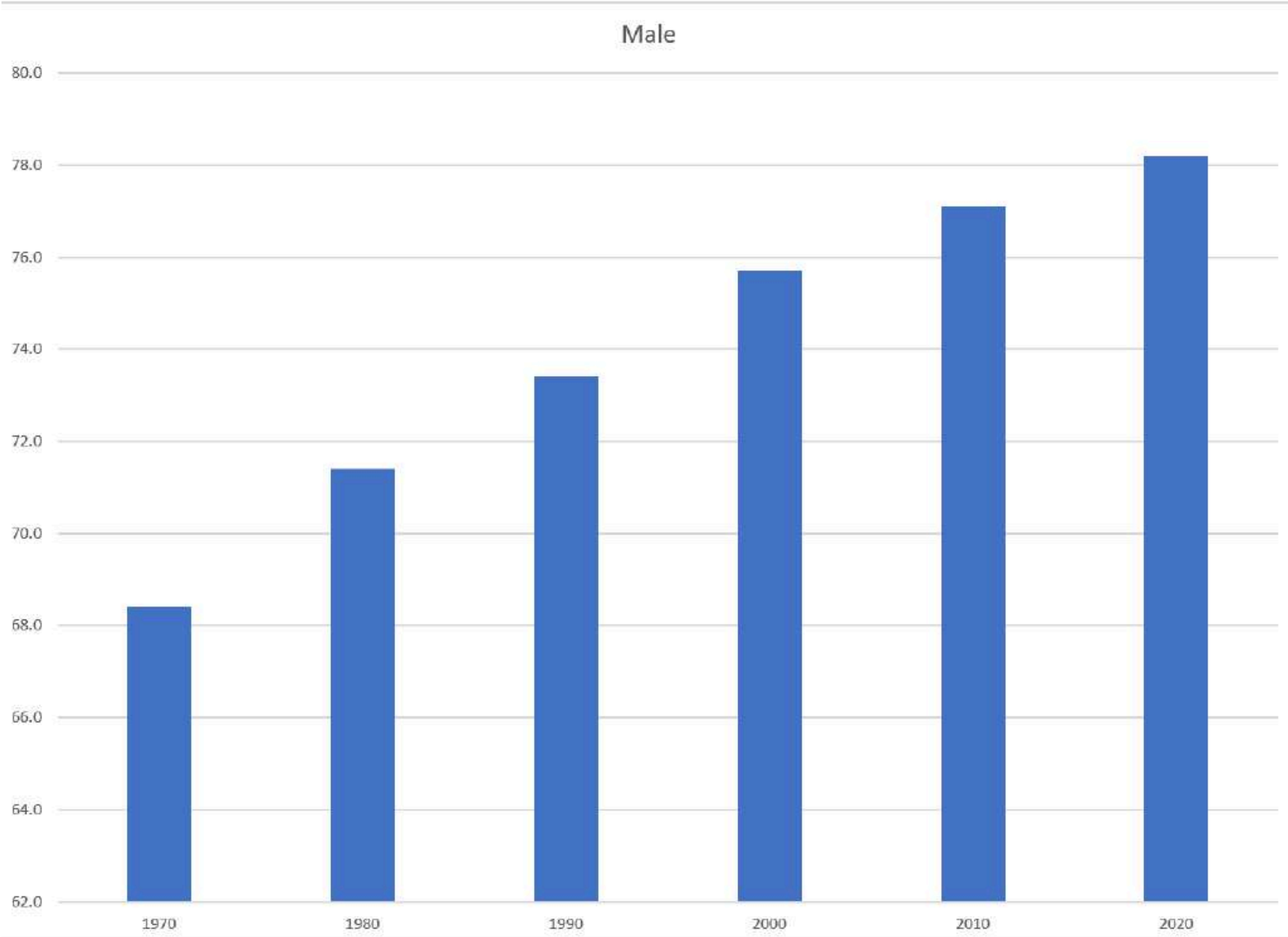


Figure 4.33 Male Bar Chart Solution

4. Repeat steps above to create a matching chart for **Life Expectancy for Oregon Women,** by selecting **A5:A11**. Press and hold the CTRL key and select **C5:C11**.

Mac Users hold down the **Command** key

5. Use the Recommended Charts and select the Clustered Column chart.
6. Move the chart to a new sheet. Name the sheet **Women**.

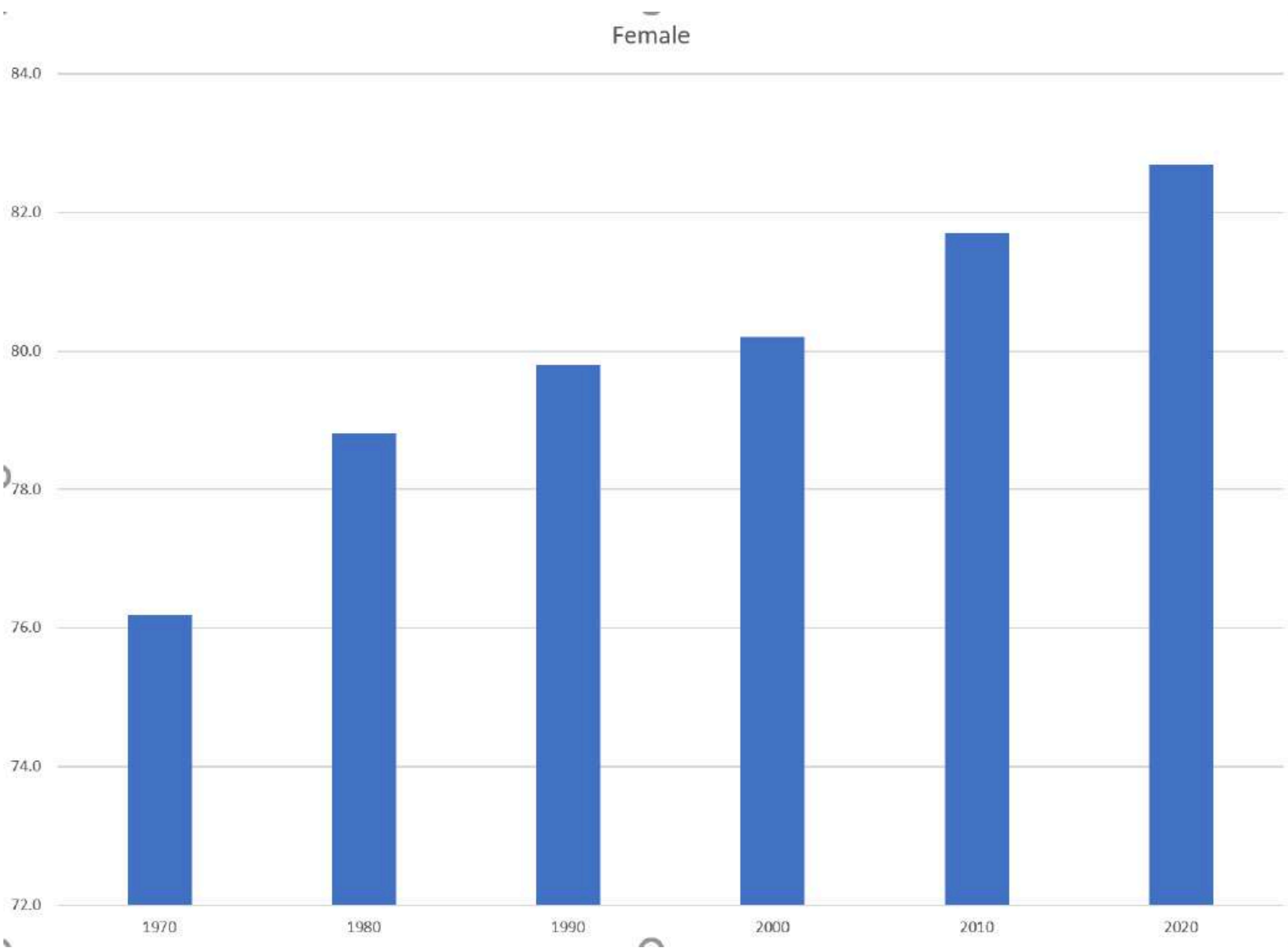


Figure 4.34 Female Bar Chart

7. Notice on the men's and women's vertical axis the min and maximum bounds do not match. To ensure data is comparable, adjust the min and max bounds of both the Mens and Womens chart to chart to match:

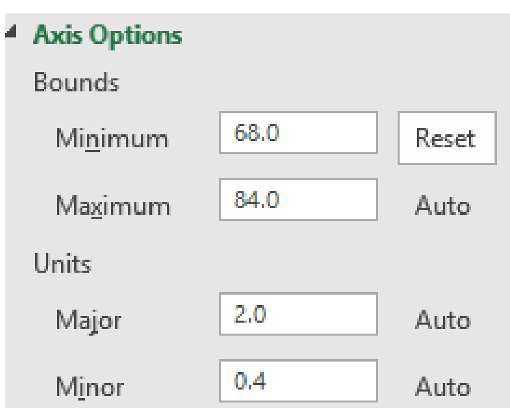
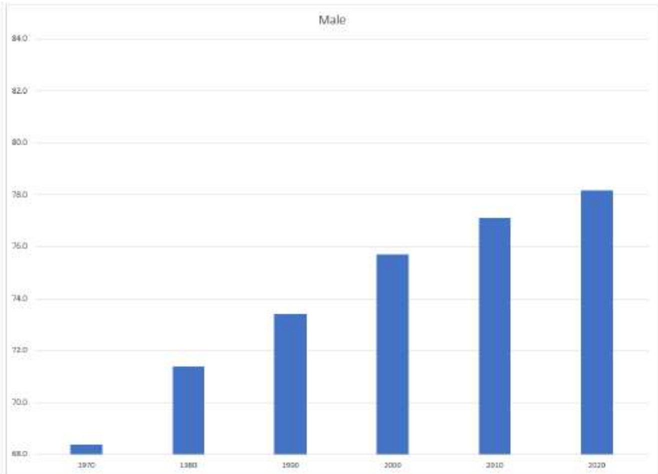
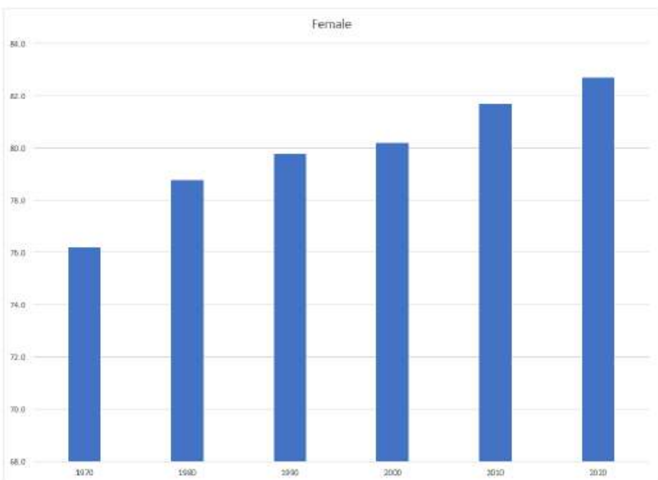


Figure 4.35 Axis Bounds





8. Return to the **Life Expectancy** tab, select **A5:D11.**

9. Use the **Recommended Charts** tool to create a simple line chart.

10. Change the Chart Title to **Oregon: Projected Life Expectancy at Birth**.

11. Leave the chart embedded in the worksheet. Move and resize it accordingly.

12. The line across the bottom of the chart represents the difference between men's and women's life expectancy. It is not very helpful as it is. Right-click on the line to open the pop-up menu. Select **Format Data Series**. In the **Format Data Series** pane, under the **Series Options** tab, select the radio button in front of **Secondary Axis**.

Mac Users should hold down the **CTRL key** and click the line at the bottom.
Select **Format Data Series**. In the **Format Data Series** pane, under the **Series Options** tab, select the radio button in front of **Secondary Axis**.

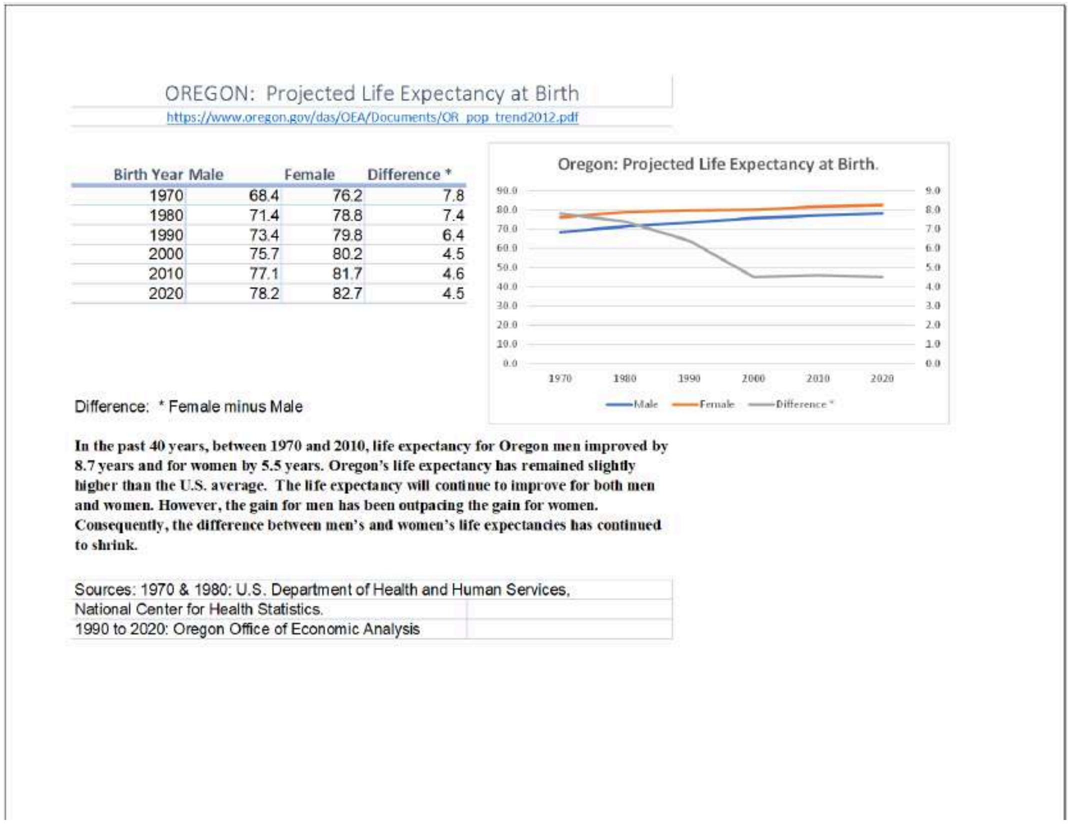
13. Close the **Format Data Series** pane.

14. Use the **Chart Style**s tools to change your chart to something a bit more dramatic.

15. Preview the **Life Expectancy** worksheet in Print Preview and make any necessary changes. The solutions are shown in below in Figure 4.35.

16. Check the spelling on all of the worksheets and make any necessary changes. Save the **PR4 Historical Data** workbook.

17. Submit the **PR4 Historical Data** workbook as directed by your instructor.


OREGON: Projected Life Expectancy at Birth
https://www.oregon.gov/das/OEA/Documents/OR_pop_trend2012.pdf

Birth Year	Male	Female	Difference *
1970	68.4	76.2	7.8
1980	71.4	78.8	7.4
1990	73.4	79.8	6.4
2000	75.7	80.2	4.5
2010	77.1	81.7	4.6
2020	78.2	82.7	4.5

Difference: * Female minus Male

In the past 40 years, between 1970 and 2010, life expectancy for Oregon men improved by 8.7 years and for women by 5.5 years. Oregon's life expectancy has remained slightly higher than the U.S. average. The life expectancy will continue to improve for both men and women. However, the gain for men has been outpacing the gain for women. Consequently, the difference between men's and women's life expectancies has continued to shrink.

Sources: 1970 & 1980: U.S. Department of Health and Human Services,
National Center for Health Statistics.
1990 to 2020: Oregon Office of Economic Analysis

Figure 4.36 Projected Life Expectancy Line Chart

Attribution

"4.5 Chapter Practice" by Hallie Puncochar and Noreen Brown, Portland Community College is licensed under CC BY 4.0

3.XLSX.6 SCORED ASSESSMENT

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Charting Sales Data

Download Data File: SC4 Data

An effective way to communicate findings in Excel is by graphing data. Graph the attached sales information, visualizing profit margin trends, and production patterns, and the sales pipeline; enabling decision-makers to grasp concepts of current business practices.

Task 1 Revenue & Profit Margin

Following the directions below create the below charts. Note to **match each chart exactly**, including the chart styles, and axis bounds.

1. Open the file named **SC4 Data** and then Save As **SC4 Sales.**
2. From the **Revenue & Profit Margin** sheet, create the below Combination Chart.
3. Move the chart to a new sheet. Name the sheet **Combo**.
4. Save your work.

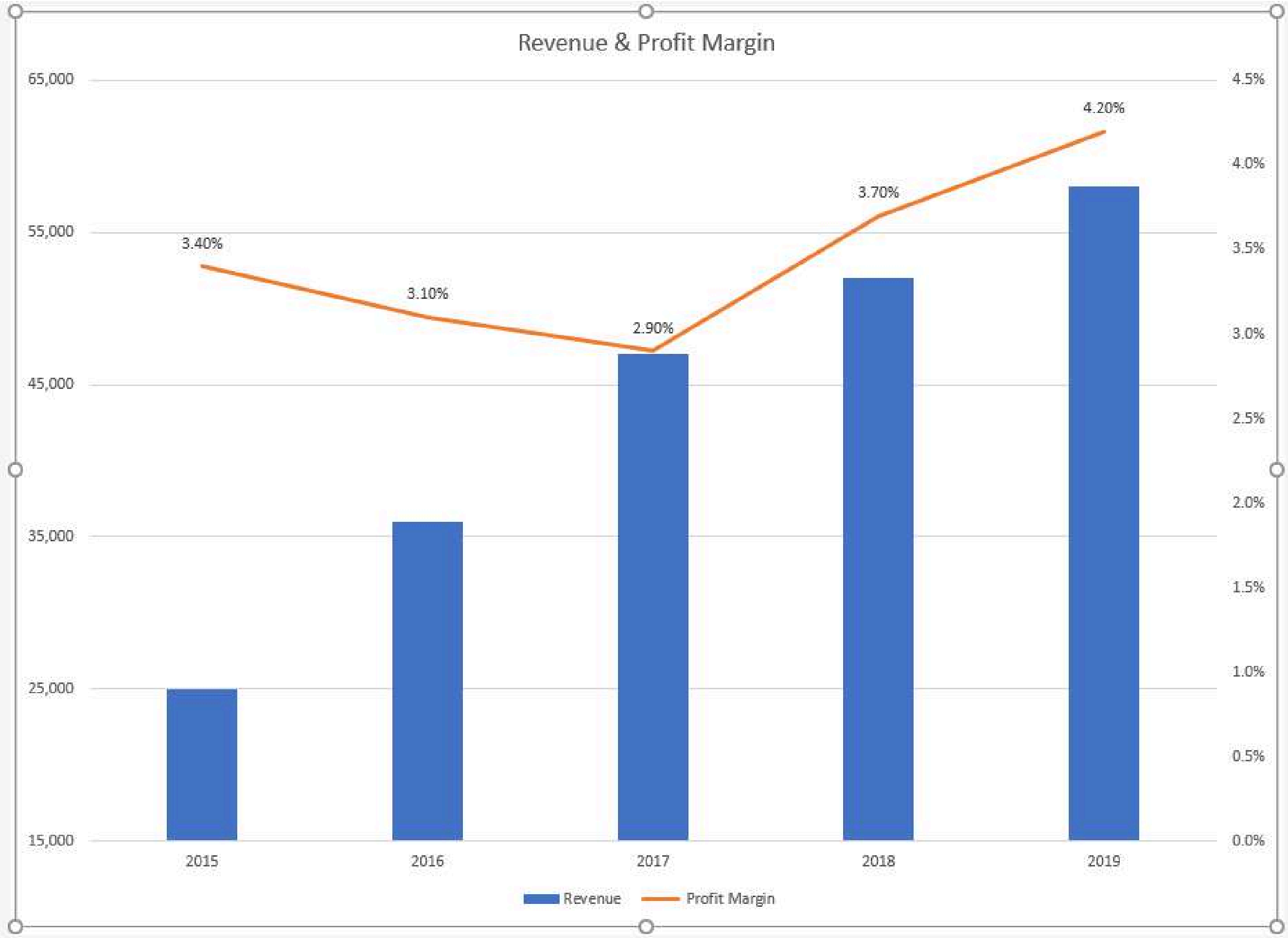


Figure 4.36 Combination Chart

Task 2 Quartly Production Data

Create the below Pie Chart showing how many locations will be producing products in the North, South, East and West Regions.

1. From the **Products sheet** create the below chart.
2. Move the chart to a new sheet. Name the sheet **Doughnut**.
3. Save your work.

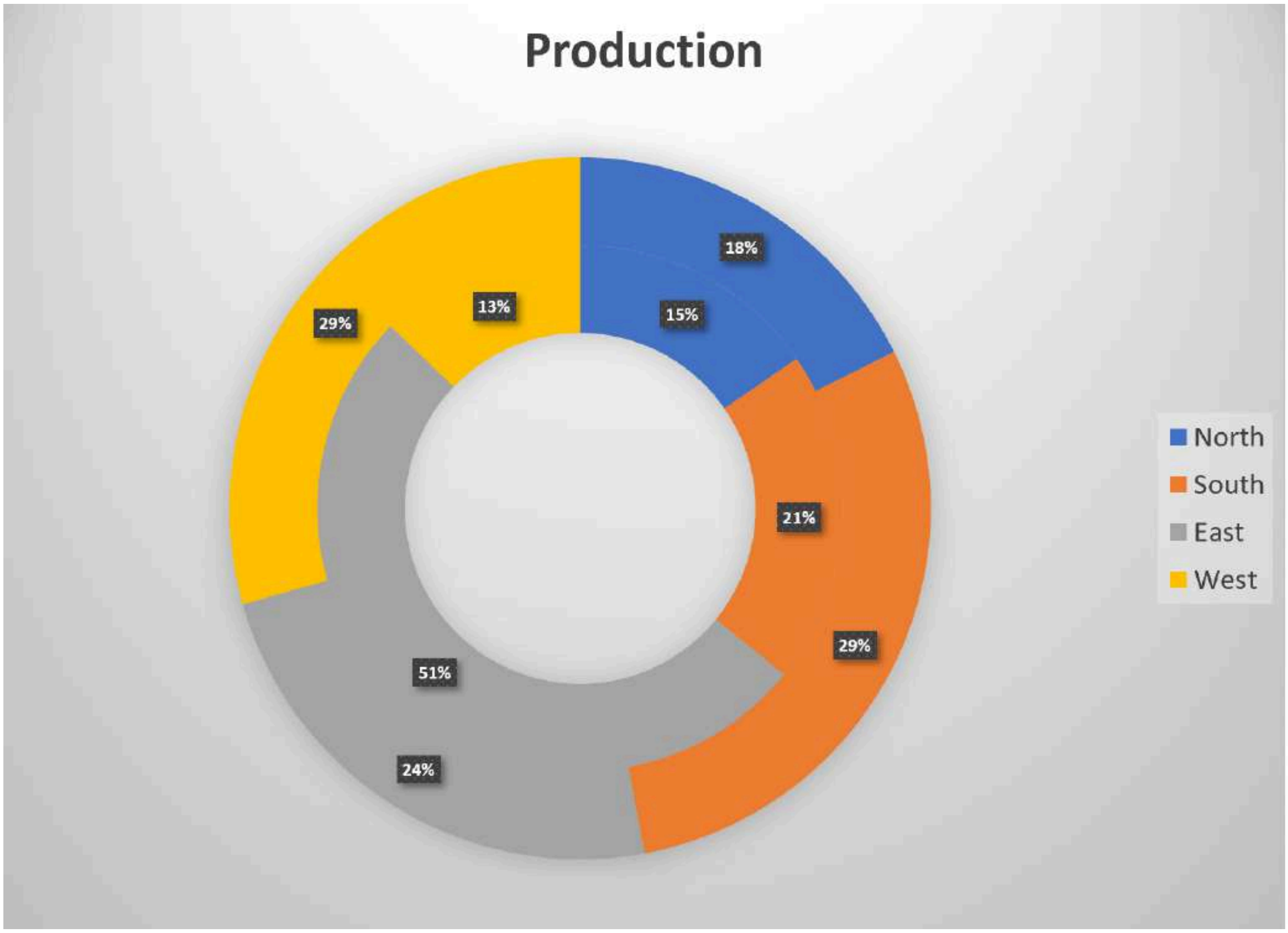


Figure 4.37 Doughnut Chart

Task 3 Sales Pipeline

Create the below Funnel Chart to provide our sales team a visual snapshot of the company's **sales** process, outlining deals that are expected to close within the month.

1. From the **Sales** sheet, create the below Funnel Chart.
2. Note to leave the chart embedded in the sheet. Resize, and move the chart accordingly.
3. Check the spelling on all of the worksheets and make any necessary changes. Save your work and submit **SC4 Sales** as directed by your instructor.

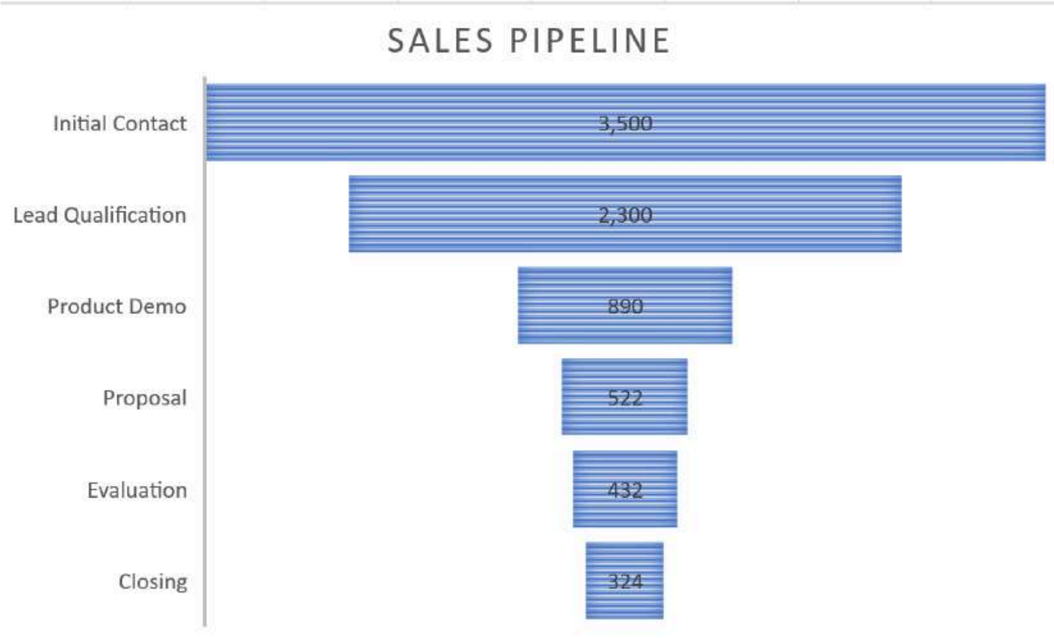


Figure 4.38 Funnel Chart

Attribution

"4.6 Scored Assessment" by Hallie Puncochar, Portland Community College is licensed under CC BY 4.0

SECTION VI

4. FREQUENCY DISTRIBUTIONS

4.1 FREQUENCY DISTRIBUTIONS FOR QUANTITATIVE DATA

4.1: Frequency Distributions for Quantitative Data

4.1.1: Guidelines for Plotting Frequency Distributions

The frequency distribution of events is the number of times each event occurred in an experiment or study.

Learning Objective

Define statistical frequency and illustrate how it can be depicted graphically.

Key Takeaways

Key Points

- Frequency distributions can be displayed in a table, histogram, line graph, dot plot, or a pie chart, just to name a few.
- A histogram is a graphical representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval.

- There is no "best" number of bins, and different bin sizes can reveal different features of the data.
- Frequency distributions can be displayed in a table, histogram, line graph, dot plot, or a pie chart, to just name a few.

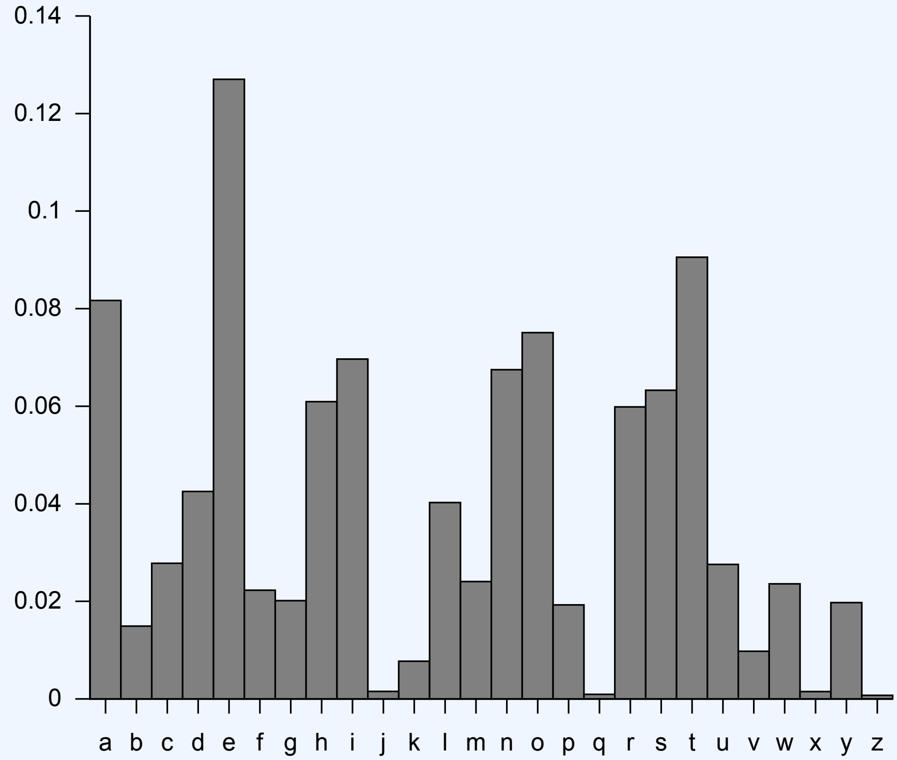
Key Terms

frequency
: number of times an event occurred in an experiment (absolute frequency)

histogram
: a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval

In statistics, the frequency (or absolute frequency) of an event is the number of times the event occurred in an experiment or study. These frequencies are often graphically represented in histograms. The relative frequency (or empirical probability) of an event refers to the absolute frequency normalized by the total number of events. The values of all events can be plotted to produce a frequency distribution.

A histogram is a graphical representation of tabulated frequencies , shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval. The height of a rectangle is also equal to the frequency density of the interval, i.e., the frequency divided by the width of the interval. The total area of the histogram is equal to the number of data. An example of the frequency distribution of letters of the alphabet in the English language is shown in the histogram in .



Letter frequency in the English language

A typical distribution of letters in English language text.

A histogram may also be normalized displaying relative frequencies. It then shows the proportion of cases that fall into each of several categories, with the total area equaling 1. The categories are usually specified as consecutive, non-overlapping intervals of a variable. The categories (intervals) must be adjacent, and often are chosen to be of the same size. The rectangles of a histogram are drawn so that they touch each other to indicate that the original variable is continuous.

There is no "best" number of bins, and different bin sizes can reveal different features of the data. Some theoreticians have attempted to determine an optimal number of bins, but these methods generally make strong assumptions about the shape of the distribution. Depending on the actual data distribution and the goals of the analysis, different bin widths may be appropriate, so experimentation is usually needed to determine an appropriate width.

4.1.2: Outliers

In statistics, an outlier is an observation that is numerically distant from the rest of the data.

Learning Objectives

Discuss outliers in terms of their causes and consequences, identification, and exclusion.

Key Takeaways

Key Points

- Outliers can occur by chance, by human error, or by equipment malfunction.
- Outliers may be indicative of a non-normal distribution, or they may just be natural deviations that occur in a large sample.
- Unless it can be ascertained that the deviation is not significant, it is not wise to ignore the presence of outliers.
- There is no rigid mathematical definition of what constitutes an outlier; thus, determining whether or not an observation is an outlier is ultimately a subjective experience.

Key Terms

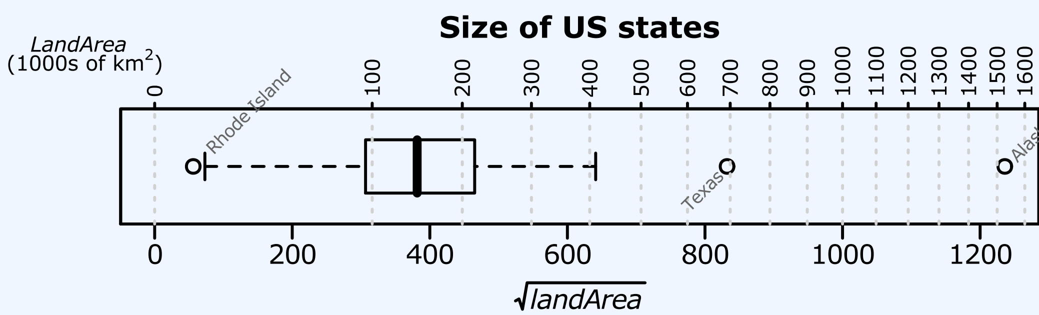
skewed
: Biased or distorted (pertaining to statistics or information).

standard deviation
: a measure of how spread out data values are around the mean, defined as the square root of the variance

interquartile range
The difference between the first and third quartiles; a robust measure of sample dispersion.

What is an Outlier?

In statistics, an outlier is an observation that is numerically distant from the rest of the data. Outliers can occur by chance in any distribution, but they are often indicative either of measurement error or of the population having a heavy-tailed distribution. In the former case, one wishes to discard the outliers or use statistics that are robust against them. In the latter case, outliers indicate that the distribution is skewed and that one should be very cautious in using tools or intuitions that assume a normal distribution.



Outliers

This box plot shows where the US states fall in terms of their size. Rhode Island, Texas, and Alaska are outside the normal data range, and therefore are considered outliers in this case.

In most larger samplings of data, some data points will be further away from the sample mean than what is deemed reasonable. This can be due to incidental systematic error or flaws in the theory that generated an assumed family of probability distributions, or it may be that some observations are far from the center of the data. Outlier points can therefore indicate faulty data, erroneous procedures, or areas where a certain theory might not be valid. However, in large samples, a small number of outliers is to be expected, and they typically are not due to any anomalous condition.

Outliers, being the most extreme observations, may include the sample maximum or sample minimum, or both, depending on whether they are extremely high or low. However, the sample maximum and minimum are not always outliers because they may not be unusually far from other observations.

Interpretations of statistics derived from data sets that include outliers may be misleading. For example, imagine that we calculate the average temperature of 10 objects in a room. Nine of them are between 20° and 25° Celsius, but an oven is at 175°C. In this case, the median of the data will be between 20° and 25°C, but the mean temperature will be between 35.5° and 40 °C. The median better reflects the temperature of a randomly sampled object than the mean; however, interpreting the mean as "a typical sample", equivalent to the median, is incorrect. This case illustrates that outliers may be indicative of data points that belong to a different population than the rest of the sample set. Estimators capable of coping with outliers are said to be robust. The median is a robust statistic, while the mean is not.

Causes for Outliers

Outliers can have many anomalous causes. For example, a physical apparatus for taking measurements may have suffered a transient malfunction, or there may have been an error in data transmission or transcription. Outliers can also arise due to changes in system behavior, fraudulent behavior, human error, instrument error or simply through natural deviations in populations. A sample may have been contaminated with elements from outside the population being examined. Alternatively, an outlier could be the result of a flaw in the assumed theory, calling for further investigation by the researcher.

Unless it can be ascertained that the deviation is not significant, it is ill-advised to ignore the presence of outliers. Outliers that cannot be readily explained demand special attention.

Identifying Outliers

There is no rigid mathematical definition of what constitutes an outlier. Thus, determining whether or not an observation is an outlier is ultimately a subjective exercise. Model-based methods, which are commonly used for identification, assume that the data is from a normal distribution and identify observations which are deemed "unlikely" based on mean and standard deviation. Other methods flag observations based on measures such as the interquartile range (IQR). For example, some people use the 1.5⋅IQR">1.5·IQR rule. This defines an outlier to be any observation that falls 1.5⋅IQR">1.5·IQR below the first quartile or any observation that falls 1.5⋅IQR">1.5·IQR above the third quartile.

Working With Outliers

Deletion of outlier data is a controversial practice frowned on by many scientists and science instructors. While mathematical criteria provide an objective and quantitative method for data rejection, they do not make the practice more scientifically or methodologically sound — especially in small sets or where a normal distribution cannot be assumed. Rejection of outliers is more acceptable in areas of practice where the underlying model of the process being measured and the usual distribution of measurement error are confidently known. An outlier resulting from an instrument reading error may be excluded, but it is desirable that the reading is at least verified.

Even when a normal distribution model is appropriate to the data being analyzed, outliers are expected for large sample sizes and should not automatically be discarded if that is the case. The application should use a classification algorithm that is robust to outliers to model data with naturally occurring outlier points. Additionally, the possibility should be considered that the underlying distribution of the data is not approximately normal, but rather skewed.

4.1.3: Relative Frequency Distributions

A relative frequency is the fraction or proportion of times a value occurs in a data set.

Learning Objective

Define relative frequency and construct a relative frequency distribution.

Key Takeaways

Key Points

- To find the relative frequencies, divide each frequency by the total number of data points in the sample.
- Relative frequencies can be written as fractions, percents, or decimals. The column should add up to 1 (or 100%).

- The only difference between a relative frequency distribution graph and a frequency distribution graph is that the vertical axis uses proportional or relative frequency rather than simple frequency.
- Cumulative relative frequency (also called an ogive) is the accumulation of the previous relative frequencies.

Key Terms

cumulative relative frequency
: the accumulation of the previous relative frequencies

relative frequency
: the fraction or proportion of times a value occurs

histogram
: a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval

What is a Relative Frequency Distribution?

A relative frequency is the fraction or proportion of times a value occurs. To find the relative frequencies, divide each frequency by the total number of data points in the sample. Relative frequencies can be written as fractions, percents, or decimals.

How to Construct a Relative Frequency Distribution

Constructing a relative frequency distribution is not that much different than from constructing a regular frequency distribution. The beginning process is the same, and the same guidelines must be used when creating classes for the data. Recall the following:

- Each data value should fit into one class only (classes are mutually exclusive).
- The classes should be of equal size.
- Classes should not be open-ended.
- Try to use between 5 and 20 classes.

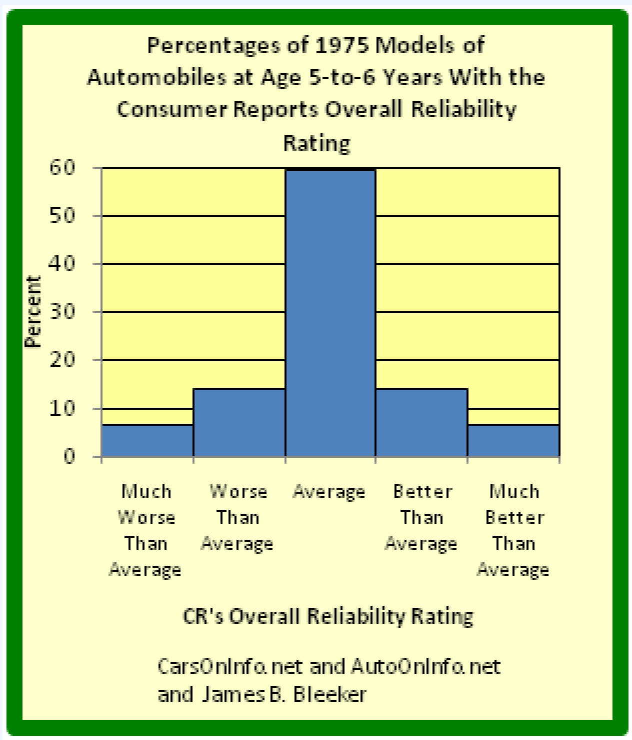
Create the frequency distribution table, as you would normally. However, this time, you will need to add a third column. The first column should be labeled *Class* or *Category*. The second column should be labeled *Frequency*. The third column should be labeled *Relative Frequency*. Fill in your class limits in column one. Then, count the number of data points that fall in each class and write that number in column two.

Next, start to fill in the third column. The entries will be calculated by dividing the frequency of that class by the total number of data points. For example, suppose we have a frequency of 5 in one class, and there are a total of 50 data points. The relative frequency for that class would be calculated by the following:

5/50=0.10

You can choose to write the relative frequency as a decimal (0.10), as a fraction (110">1/10), or as a percent (10%). Since we are dealing with proportions, the relative frequency column should add up to 1 (or 100%). It may be slightly off due to rounding.

Relative frequency distributions is often displayed in histograms and in frequency polygons. The only difference between a relative frequency distribution graph and a frequency distribution graph is that the vertical axis uses proportional or relative frequency rather than simple frequency.



Relative Frequency Histogram

This graph shows a relative frequency histogram. Notice the vertical axis is labeled with percentages rather than simple frequencies.

Cumulative Relative Frequency Distributions

Just like we use cumulative frequency distributions when discussing simple frequency distributions, we often use cumulative frequency distributions when dealing with relative frequency as well. Cumulative relative frequency (also called an *ogive*) is the accumulation of the previous relative frequencies. To find the cumulative relative frequencies, add all the previous relative frequencies to the relative frequency for the current row.

4.1.4: Cumulative Frequency Distributions

A cumulative frequency distribution displays a running total of all the preceding frequencies in a frequency distribution.

Learning Objectives

Define cumulative frequency and construct a cumulative frequency distribution.

Key Takeaways

Key Points

- To create a cumulative frequency distribution, start by creating a regular frequency distribution with one extra column added.
- To complete the cumulative frequency column, add all the frequencies at that class and all preceding classes.
- Cumulative frequency distributions are often displayed in histograms and in frequency polygons.

Key Terms

histogram
: a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval

frequency distribution
: a representation, either in a graphical or tabular format, which displays the number of observations within a given interval

What is a Cumulative Frequency Distribution?

A cumulative frequency distribution is the sum of the class and all classes below it in a frequency distribution. Rather than displaying the frequencies from each class, a cumulative frequency distribution displays a running total of all the preceding frequencies.

How to Construct a Cumulative Frequency Distribution

Constructing a cumulative frequency distribution is not that much different than constructing a regular frequency distribution. The beginning process is the same, and the same guidelines must be used when creating classes for the data. Recall the following:

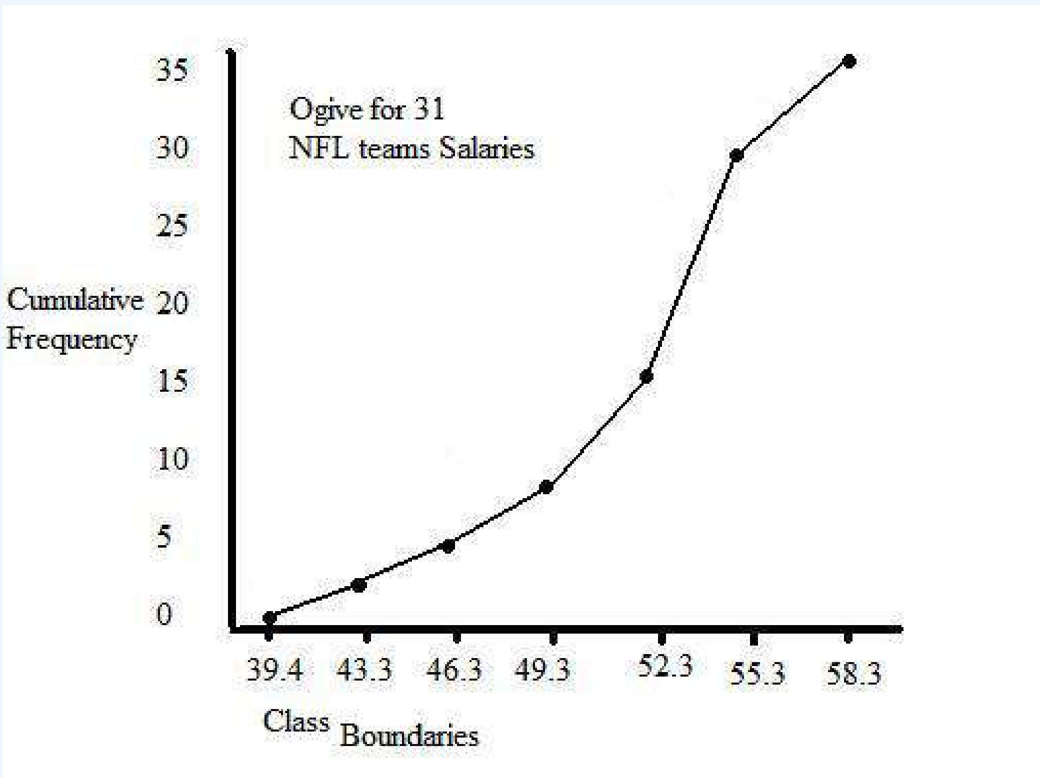
- Each data value should fit into one class only (classes are mutually exclusive).
- The classes should be of equal size.
- Classes should not be open-ended.
- Try to use between 5 and 20 classes.

Create the frequency distribution table, as you would normally. However, this time, you will need to add a third column. The first column should be labeled *Class* or *Category*. The second column should be labeled *Frequency*. The third column should be labeled *Cumulative Frequency*. Fill in your class limits in column one. Then, count the number of data points that falls in each class and write that number in column two.

Next, start to fill in the third column. The first entry will be the same as the first entry in the *Frequency* column. The second entry will be the sum of the first two entries in the *Frequency* column, the third entry will be the sum of the first three entries in the *Frequency* column, etc. The last entry in the *Cumulative Frequency* column should equal the number of total data points, if the math has been done correctly.

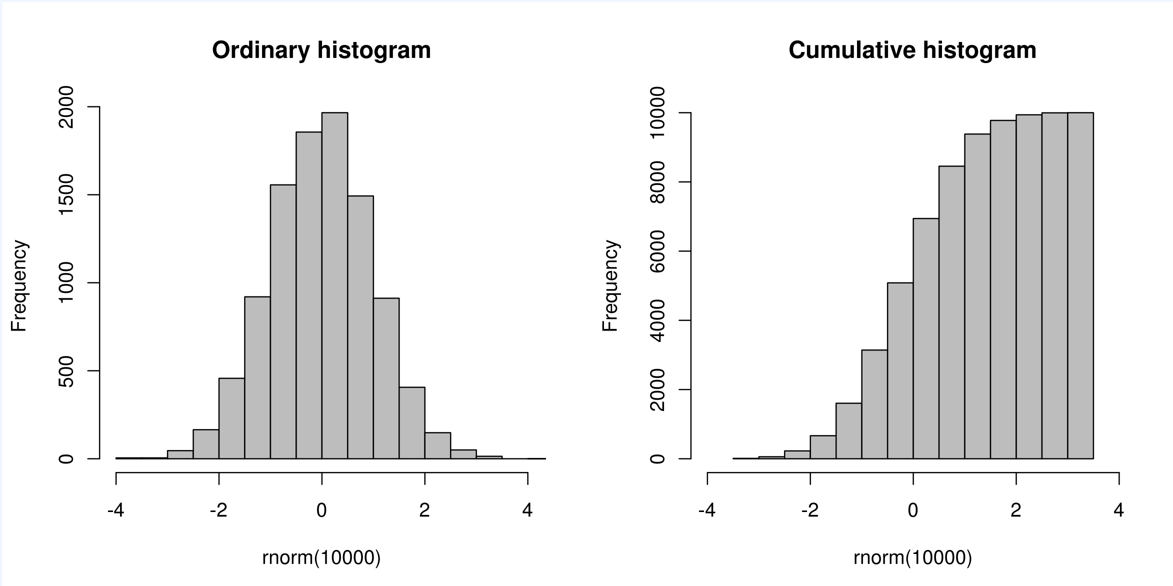
Graphical Displays of Cumulative Frequency Distributions

There are a number of ways in which cumulative frequency distributions can be displayed graphically. Histograms are common , as are frequency polygons . Frequency polygons are a graphical device for understanding the shapes of distributions. They serve the same purpose as histograms, but are especially helpful in comparing sets of data.



Frequency Polygon

This graph shows an example of a cumulative frequency polygon.



Frequency Histograms

This image shows the difference between an ordinary histogram and a cumulative frequency histogram.

4.1.5: Graphs for Quantitative Data

A plot is a graphical technique for representing a data set, usually as a graph showing the relationship between two or more variables.

Learning Objective

Identify common plots used in statistical analysis.

Key Takeaways

Key Points

- Graphical procedures such as plots are used to gain insight into a data set in terms of testing assumptions, model selection, model validation, estimator selection, relationship identification, factor effect determination, or outlier detection.
- Statistical graphics give insight into aspects of the underlying structure of the data.
- Graphs can also be used to solve some mathematical equations, typically by finding where two plots intersect.

Key Terms

histogram
: a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval

plot
: a graph or diagram drawn by hand or produced by a mechanical or electronic device

scatter plot
: A type of display using Cartesian coordinates to display values for two variables for a set of data.

A plot is a graphical technique for representing a data set, usually as a graph showing the relationship between two or more variables. Graphs of functions are used in mathematics, sciences, engineering, technology, finance, and other areas where a visual representation of the relationship between variables would be useful. Graphs can also be used to read off the value of an unknown variable plotted as a function of a known one. Graphical procedures are also used to gain insight into a data set in terms of:

- testing assumptions,
- model selection,
- model validation,
- estimator selection,

- relationship identification,
- factor effect determination, or
- outlier detection.

Plots play an important role in statistics and data analysis. The procedures here can broadly be split into two parts: quantitative and graphical. Quantitative techniques are the set of statistical procedures that yield numeric or tabular output. Some examples of quantitative techniques include:

- hypothesis testing,
- analysis of variance,
- point estimates and confidence intervals, and
- least squares regression.

There are also many statistical tools generally referred to as graphical techniques which include:

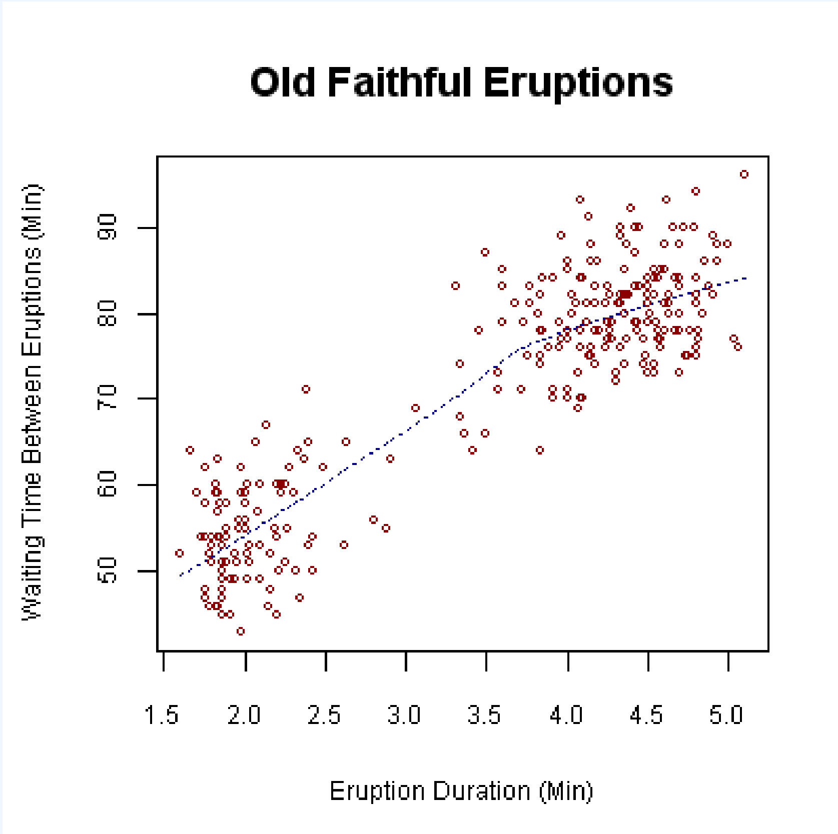
- scatter plots ,
- histograms,
- probability plots,
- residual plots,
- box plots, and
- block plots.

Below are brief descriptions of some of the most common plots:

Scatter plot: This is a type of mathematical diagram using Cartesian coordinates to display values for two variables for a set of data. The data is displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis. This kind of plot is also called a scatter chart, scattergram, scatter diagram, or scatter graph.

Histogram: In statistics, a histogram is a graphical representation of the distribution of data. It is an estimate of the probability distribution of a continuous variable or can be used to plot the frequency of an event (number of times an event occurs) in an experiment or study.

Box plot: In descriptive statistics, a boxplot, also known as a box-and-whisker diagram, is a convenient way of graphically depicting groups of numerical data through their five-number summaries (the smallest observation, lower quartile (Q1), median (Q2), upper quartile (Q3), and largest observation). A boxplot may also indicate which observations, if any, might be considered outliers.



Scatter Plot

This is an example of a scatter plot, depicting the waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA.

4.1.6: Typical Shapes

Distributions can be symmetrical or asymmetrical depending on how the data falls.

Learning Objective

Evaluate the shapes of symmetrical and asymmetrical frequency distributions.

Key Takeaways

Key Points

- A normal distribution is a symmetric distribution in which the mean and median are equal. Most data are clustered in the center.
- An asymmetrical distribution is said to be positively skewed (or skewed to the right) when the tail on the right side of the histogram is longer than the left side.
- An asymmetrical distribution is said to be negatively skewed (or skewed to the left) when the tail on the left side of the histogram is longer than the right side.
- Distributions can also be uni-modal, bi-modal, or multi-modal.

Key Terms

skewness
: A measure of the asymmetry of the probability distribution of a real-valued random variable; is the third standardized moment, defined as where is the third moment about the mean and is the standard deviation.

empirical rule
: That a normal distribution has 68% of its observations within one standard deviation of the mean, 95% within two, and 99.7% within three.

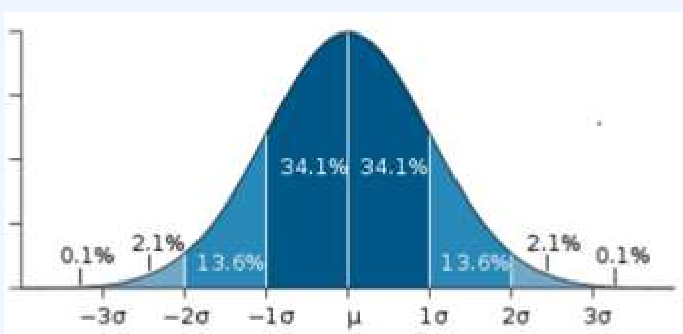
standard deviation
: a measure of how spread out data values are around the mean, defined as the square root of the variance

Distribution Shapes

In statistics, distributions can take on a variety of shapes. Considerations of the shape of a distribution arise in statistical data analysis, where simple quantitative descriptive statistics and plotting techniques, such as histograms, can lead to the selection of a particular family of distributions for modelling purposes.

Symmetrical Distributions

In a symmetrical distribution, the two sides of the distribution are mirror images of each other. A normal distribution is an example of a truly symmetric distribution of data item values. When a histogram is constructed on values that are normally distributed, the shape of the columns form a symmetrical bell shape. This is why this distribution is also known as a "normal curve" or "bell curve. " In a true normal distribution, the mean and median are equal, and they appear in the center of the curve. Also, there is only one mode, and most of the data are clustered around the center. The more extreme values on either side of the center become more rare as distance from the center increases. About 68% of values lie within one standard deviation (σ) away from the mean, about 95% of the values lie within two standard deviations, and about 99.7% lie within three standard deviations . This is known as the empirical rule or the 3-sigma rule.



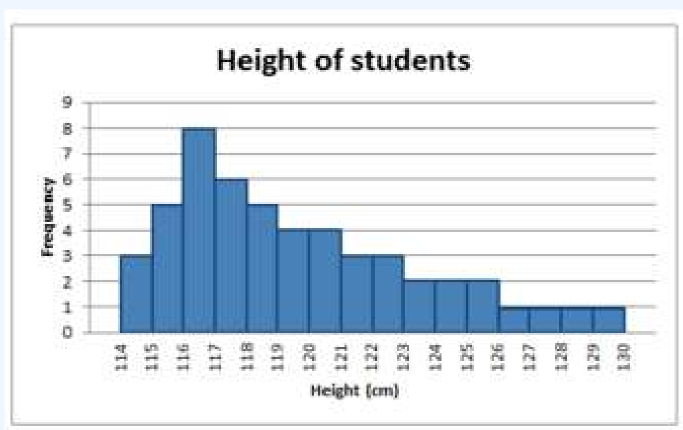
Normal Distribution

This image shows a normal distribution. About 68% of data fall within one standard deviation, about 95% fall within two standard deviations, and 99.7% fall within three standard deviations.

Asymmetrical Distributions

In an asymmetrical distribution, the two sides will not be mirror images of each other. Skewness is the tendency for the values to be more frequent around the high or low ends of the x-axis. When a histogram is constructed for skewed data, it is possible to identify skewness by looking at the shape of the distribution.

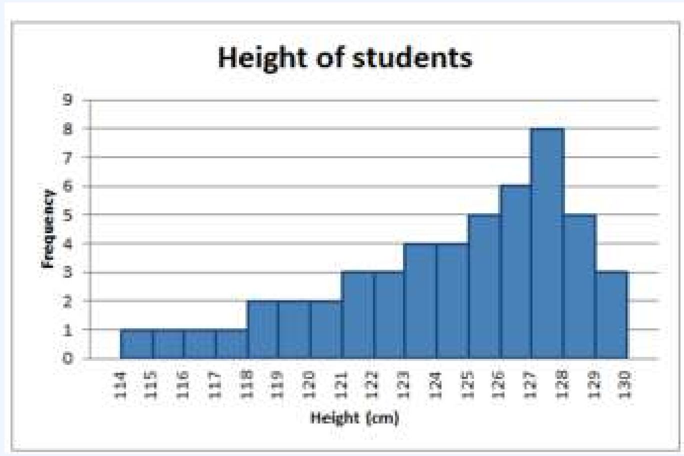
A distribution is said to be positively skewed (or skewed to the right) when the tail on the right side of the histogram is longer than the left side. Most of the values tend to cluster toward the left side of the x-axis (i.e., the smaller values) with increasingly fewer values at the right side of the x-axis (i.e., the larger values). In this case, the median is less than the mean .



Positively Skewed Distribution

This distribution is said to be positively skewed (or skewed to the right) because the tail on the right side of the histogram is longer than the left side.

A distribution is said to be negatively skewed (or skewed to the left) when the tail on the left side of the histogram is longer than the right side. Most of the values tend to cluster toward the right side of the x-axis (i.e., the larger values), with increasingly less values on the left side of the x-axis (i.e., the smaller values). In this case, the median is greater than the mean .



Negatively Skewed Distribution

This distribution is said to be negatively skewed (or skewed to the left) because the tail on the left side of the histogram is longer than the right side.

When data are skewed, the median is usually a more appropriate measure of central tendency than the mean.

Other Distribution Shapes

A uni-modal distribution occurs if there is only one "peak" (or highest point) in the distribution, as seen previously in the normal distribution. This means there is one mode (a value that occurs more frequently than any other) for the data. A bi-modal distribution occurs when there are two modes. Multi-modal distributions with more than two modes are also possible.

4.1.7: Z-Scores and Location in a Distribution

A z-score is the signed number of standard deviations an observation is above the mean of a distribution.

Learning Objectives

Define z-scores and demonstrate how they are converted from raw scores.

Key Takeaways

Key Points

- A positive z-score represents an observation above the mean, while a negative z-score represents an observation below the mean.
- We obtain a z-score through a conversion process known as standardizing or normalizing.
- Z-scores are most frequently used to compare a sample to a standard normal deviate (standard normal distribution, with with μ=0">μ=0 and σ=1">σ=1).
- While z-scores can be defined without assumptions of normality, they can only be defined if one knows the population parameters.
- z-scores provide an assessment of how off-target a process is operating.

Key Terms

Student's t-statistic
: a ratio of the departure of an estimated parameter from its notional value and its standard error

z-score
: The standardized value of observation x from a distribution that has mean mu and standard deviation $sigma$.

raw score
: an original observation that has not been transformed to a z-score

A z-score is the signed number of standard deviations an observation is above the mean of a distribution. Thus, a positive z-score represents an observation above the mean, while a negative z-score represents an observation below the mean. We obtain a z-score through a conversion process known as standardizing or normalizing.

z-scores are also called standard scores, z-values, normal scores or standardized variables. The use of "z" is because the normal distribution is also known as the "z distribution." z-scores are most frequently used to compare a sample to a standard normal deviate (standard normal distribution, with μ=0″>$\mu=0$ and σ=1″>$\sigma=1$).

While z-scores can be defined without assumptions of normality, they can only be defined if one knows the population parameters. If one only has a sample set, then the analogous computation with sample mean and sample standard deviation yields the Student's t″>$t-$ statistic.

Calculation From a Raw Score

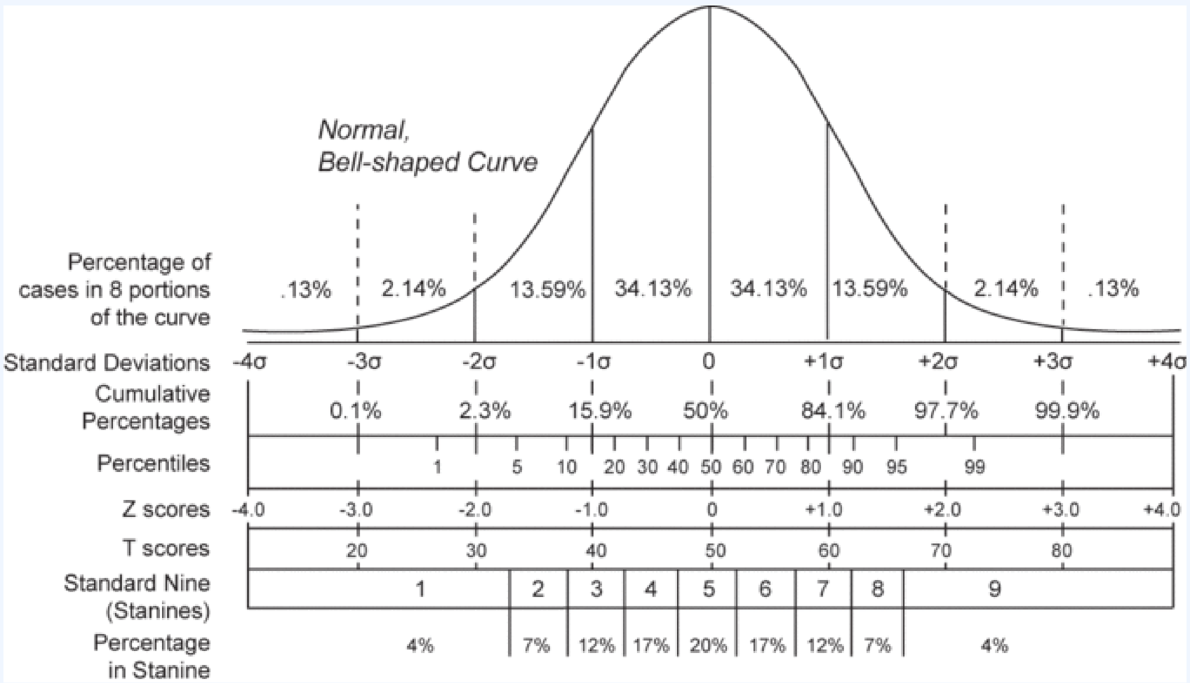
A raw score is an original datum, or observation, that has not been transformed. This may include, for example, the original result obtained by a student on a test (i.e., the number of correctly answered items) as opposed to that score after transformation to a standard score or percentile rank. The z-score, in turn, provides an assessment of how off-target a process is operating.

The conversion of a raw score, x″>x, to a z″>z-score can be performed using the following equation:

z=x−μσ″>$z=(x-\mu)/\sigma$

where μ″>μ is the mean of the population and σ″>σ is the standard deviation of the population. The absolute value of z″>z represents the distance between the raw score and the population mean in units of the standard deviation. z″>z is negative when the raw score is below the mean and positive when the raw score is above the mean.

A key point is that calculating z″>z requires the population mean and the population standard deviation, not the sample mean nor sample deviation. It requires knowing the population parameters, not the statistics of a sample drawn from the population of interest. However, in cases where it is impossible to measure every member of a population, the standard deviation may be estimated using a random sample.



Normal Distribution and Scales

Shown here is a chart comparing the various grading methods in a normal distribution. z-scores for this standard normal distribution can be seen in between percentiles and z-scores.

Attributions

- Guidelines for Plotting Frequency Distributions
    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "Histogram."
      https://en.wikipedia.org/wiki/Histogram.
      Wikipedia
      CC BY-SA 3.0.
    - "Statistical frequency."
      https://en.wikipedia.org/wiki/Statistical_frequency.
      Wikipedia

CC BY-SA 3.0.

- "Frequency analysis."
  https://en.wikipedia.org/wiki/Frequency_analysis.
  Wikipedia
  CC BY-SA 3.0.
- "histogram."
  http://en.wikipedia.org/wiki/histogram.
  Wikipedia
  CC BY-SA 3.0.
- "frequency."
  http://en.wiktionary.org/wiki/frequency.
  Wiktionary
  CC BY-SA 3.0.
- "Chapter 2.
  Summarizing Data: listing and grouping – Statistics."
  http://statistics.wikidot.com/ch2.
  Wikidot
  CC BY.
- "English letter frequency (alphabetic)."
  https://en.wikipedia.org/wiki/File:English_letter_frequency_(alphabetic).svg.
  Wikipedia
  Public domain.

- Outliers
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "interquartile range."
    http://en.wiktionary.org/wiki/interquartile_range.
    Wiktionary
    CC BY-SA 3.0.
  - "Outliers in statistics."
    http://en.wikipedia.org/wiki/Outliers_in_statistics.
    Wikipedia
    CC BY-SA 3.0.
  - "standard deviation."
    http://en.wiktionary.org/wiki/standard_deviation.

Wiktionary
CC BY-SA 3.0.
    - "skewed."
      http://en.wiktionary.org/wiki/skewed.
      Wiktionary
      CC BY-SA 3.0.
    - "Outlier."
      http://en.wikipedia.org/wiki/Outlier.
      Wikipedia
      GNU FDL.
- Relative Frequency Distributions
    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "histogram."
      http://en.wikipedia.org/wiki/histogram.
      Wikipedia
      CC BY-SA 3.0.
    - "Susan Dean and Barbara Illowsky, Sampling and Data: Frequency, Relative Frequency, and Cumulative Frequency. September 17, 2013."
      http://cnx.org/content/m16012/latest/.
      OpenStax CNX
      CC BY 3.0.
    - "Histogram of Consumer Reports."
      http://commons.wikimedia.org/wiki/File:Histogram_of_Consumer_Reports.
      Wikimedia
      CC BY-SA.
- Cumulative Frequency Distributions
    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "histogram."
      http://en.wikipedia.org/wiki/histogram.
      Wikipedia
      CC BY-SA 3.0.

    - "David Lane, Frequency Polygons. September 17, 2013."
      http://cnx.org/content/m11214/latest/.
      OpenStax CNX
      CC BY 3.0.
    - "David Lane, Frequency Polygons. September 17, 2013."
      http://cnx.org/content/m11214/latest/.
      OpenStax CNX
      CC BY 3.0.
    - "Terms of Use – Wikimedia Foundation."
      http://wikimediafoundation.org/wiki/Terms_of_Use.
      Wikimedia Foundation
      CC BY-SA.
    - "Terms of Use – Wikimedia Foundation."
      http://wikimediafoundation.org/wiki/Terms_of_Use.
      Wikimedia Foundation
      CC BY-SA.
- Graphs for Quantitative Data
    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "Relative frequency."
      https://en.wikipedia.org/wiki/Relative_frequency.
      Wikipedia
      CC BY-SA 3.0.
    - "Histogram."
      https://en.wikipedia.org/wiki/Histogram.
      Wikipedia
      CC BY-SA 3.0.
    - "Scatter plot."
      https://en.wikipedia.org/wiki/Scatter_plot.
      Wikipedia
      CC BY-SA 3.0.
    - "Data plot."
      http://en.wikipedia.org/wiki/Data_plot.
      Wikipedia
      CC BY-SA 3.0.

  - "histogram."
    http://en.wikipedia.org/wiki/histogram.
    Wikipedia
    CC BY-SA 3.0.
  - "plot."
    http://en.wiktionary.org/wiki/plot.
    Wiktionary
    CC BY-SA 3.0.
  - "scatter plot."
    http://en.wiktionary.org/wiki/scatter_plot.
    Wiktionary
    CC BY-SA 3.0.
  - Wikimedia.
    http://upload.wikimedia.org/wikipedia/commons/0/0f/Oldfaithful3.png.
    Public domain.
- Typical Shapes
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Statistical Language – Measures of Shape."
    http://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+measures+of+shape.
    Austrailian Bureau of Statistics
    CC BY.
  - "Shape of the distribution."
    http://en.wikipedia.org/wiki/Shape_of_the_distribution.
    Wikipedia
    CC BY-SA 3.0.
  - "empirical rule."
    http://en.wiktionary.org/wiki/empirical_rule.
    Wiktionary
    CC BY-SA 3.0.
  - "skewness."
    http://en.wiktionary.org/wiki/skewness.

Wiktionary
CC BY-SA 3.0.
  - "standard deviation."
    http://en.wiktionary.org/wiki/standard_deviation.
    Wiktionary
    CC BY-SA 3.0.
  - "Error 404."
    http://www.abs.gov.au/websitedbs/a3121120.nsf/89a5f3d8684682b6ca256de4002c809b/81a53a0a10c05d3bca257949001281b5!OpenDocument.
    Austrailian Bureau of Statistics
    CC BY.
  - "Error 404."
    http://www.abs.gov.au/websitedbs/a3121120.nsf/89a5f3d8684682b6ca256de4002c809b/81a53a0a10c05d3bca257949001281b5!OpenDocument.
    Austrailian Bureau of Statistics
    CC BY.
  - "Error 404."
    http://www.abs.gov.au/websitedbs/a3121120.nsf/89a5f3d8684682b6ca256de4002c809b/81a53a0a10c05d3bca257949001281b5!OpenDocument.
    Austrailian Bureau of Statistics
    CC BY.
- Z-Scores and Location in a Distribution
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Student's t-statistic."
    http://en.wikipedia.org/wiki/Student's%20t-statistic.
    Wikipedia
    CC BY-SA 3.0.
  - "Standard score."
    http://en.wikipedia.org/wiki/Standard_score.
    Wikipedia
    CC BY-SA 3.0.
  - "Raw score."
    http://en.wikipedia.org/wiki/Raw_score.
    Wikipedia

CC BY-SA 3.0.

- "raw score."
  http://en.wikipedia.org/wiki/raw%20score.
  Wikipedia
  CC BY-SA 3.0.
- "Normal distribution and scales."
  http://en.wikipedia.org/wiki/File:Normal_distribution_and_scales.gif.
  Wikipedia
  Public domain.

4.2 FREQUENCY DISTRIBUTIONS FOR QUALITATIVE DATA

4.2: Frequency Distributions for Qualitative Data

4.2.1: Describing Qualitative Data

Qualitative data is a categorical measurement expressed not in terms of numbers, but rather by means of a natural language description.

Learning Objectives

Summarize the processes available to researchers that allow qualitative data to be analyzed similarly to quantitative data.

Key Takeaways

Key Points

- Observer impression is when expert or bystander observers examine the data, interpret it via forming an impression and report their impression in a structured and sometimes quantitative form.

- To discover patterns in qualitative data, one must try to find frequencies, magnitudes, structures, processes, causes, and consequences.
- The Ground Theory Method (GTM) is an inductive approach to research in which theories are generated solely from an examination of data rather than being derived deductively.
- Coding is an interpretive technique that both organizes the data and provides a means to introduce the interpretations of it into certain quantitative methods.
- Most coding requires the analyst to read the data and demarcate segments within it.

Key Terms

nominal
: Having values whose order is insignificant.

ordinal
: Of a number, indicating position in a sequence.

qualitative analysis
: The numerical examination and interpretation of observations for the purpose of discovering underlying meanings and patterns of relationships.

Qualitative data is a categorical measurement expressed not in terms of numbers, but rather by means of a natural language description. In statistics, it is often used interchangeably with "categorical" data. When there is not a natural ordering of the categories, we call these nominal categories. Examples might be gender, race, religion, or sport.

When the categories may be ordered, these are called ordinal variables. Categorical variables that judge size (small, medium, large, etc.) are ordinal variables. Attitudes (strongly disagree, disagree, neutral, agree, strongly agree) are also ordinal variables; however, we may not know which value is the best or worst of these issues. Note that the distance between these categories is not something we can measure.

Qualitative Analysis

Qualitative Analysis is the numerical examination and interpretation of observations for the purpose of discovering underlying meanings and patterns of relationships. The most common form of qualitative qualitative analysis is observer impression—when an expert or bystander observers examine the data, interpret it via forming an impression and report their impression in a structured and sometimes quantitative form.

An important first step in qualitative analysis and observer impression is to discover patterns. One must try to find frequencies, magnitudes, structures, processes, causes, and consequences. One method of this is through cross-case analysis, which is analysis that involves an examination of more than one case. Cross-case analysis can be further broken down into *variable-oriented analysis* and *case-oriented analysis*. Variable-oriented analysis is that which describes and/or explains a particular variable, while case-oriented analysis aims to understand a particular case or several cases by looking closely at the details of each.

The Ground Theory Method (GTM) is an inductive approach to research, introduced by Barney Glaser and Anselm Strauss, in which theories are generated solely from an examination of data rather than being derived deductively. A component of the Grounded Theory Method is the *constant comparative method*, in which observations are compared with one another and with the evolving inductive theory.

Four Stages of the Constant Comparative Method

1. comparing incident application to each category
2. integrating categories and their properties
3. delimiting the theory
4. writing theory

Other methods of discovering patterns include semiotics and conversation analysis. Semiotics is the study of signs and the meanings associated with them. It is commonly associated with content analysis. Conversation analysis is a meticulous analysis of the details of conversation, based on a complete transcript that includes pauses and other non-verbal communication.

Conceptualization and Coding

In quantitative analysis, it is usually obvious what the variables to be analyzed are, for example, race, gender, income, education, etc. Deciding what is a variable, and how to code each subject on each variable, is more difficult in qualitative data analysis.

Concept formation is the creation of variables (usually called *themes*) out of raw qualitative data. It is more sophisticated in qualitative data analysis. Casing is an important part of concept formation. It is the process of determining what represents a case. Coding is the actual transformation of qualitative data into themes.

More specifically, coding is an interpretive technique that both organizes the data and provides a means to introduce the interpretations of it into certain quantitative methods. Most coding requires the analyst to read the data and demarcate segments within it, which may be done at different times throughout the process. Each segment is labeled with a “code” – usually a word or short phrase that suggests how the associated data segments inform the research objectives. When coding is complete, the analyst prepares reports via a mix of: summarizing the prevalence of codes, discussing similarities and differences in related codes across distinct original sources/contexts, or comparing the relationship between one or more codes.

Some qualitative data that is highly structured (e.g., close-end responses from surveys or tightly defined interview questions) is typically coded without additional segmenting of the content. In these cases, codes are often applied as a layer on top of the data. Quantitative analysis of these codes is typically the capstone analytical step for this type of qualitative data.

A frequent criticism of coding method is that it seeks to transform qualitative data into empirically valid data that contain actual value range, structural proportion, contrast ratios, and scientific objective properties. This can tend to drain the data of its variety, richness, and individual character. Analysts respond to this criticism by thoroughly expositing their definitions of codes and linking those codes soundly to the underlying data, therein bringing back some of the richness that might be absent from a mere list of codes.

Alternatives to Coding

Alternatives to coding include recursive abstraction and mechanical techniques. Recursive abstraction involves the summarizing of datasets. Those summaries are then further summarized and so on. The end result is a more compact summary that would have been difficult to accurately discern without the preceding steps of distillation.

Mechanical techniques rely on leveraging computers to scan and reduce large sets of qualitative data. At their most basic level, mechanical techniques rely on counting words, phrases, or coincidences of tokens within the data. Often referred to as content analysis, the output from these techniques is amenable to many advanced statistical analyses.

4.2.2: Interpreting Distributions Constructed by Others

Graphs of distributions created by others can be misleading, either intentionally or unintentionally.

Learning Objective

Demonstrate how distributions constructed by others may be misleading, either intentionally or unintentionally

Key Takeaways

Key Points

- Misleading graphs will misrepresent data, constituting a misuse of statistics that may result in an incorrect conclusion being derived from them.
- Graphs can be misleading if they're used excessively, if they use the third dimensions where it is unnecessary, if they are improperly scaled, or if they're truncated.
- The use of biased or loaded words in the graph's title, axis labels, or caption may inappropriately prime the reader.

Key Terms

bias
: (Uncountable) Inclination towards something; predisposition, partiality, prejudice, preference, predilection.

distribution
: the set of relative likelihoods that a variable will have a value in a given interval

truncate
: To shorten something as if by cutting off part of it.

Distributions Constructed by Others

Unless you are constructing a graph of a distribution on your own, you need to be very careful about how you read and interpret graphs. Graphs are made in order to display data; however, some people may intentionally try to mislead the reader in order to convey certain information.

In statistics, these types of graphs are called misleading graphs (or distorted graphs). They misrepresent data, constituting a misuse of statistics that may result in an incorrect conclusion being derived from them. Graphs may be misleading through being excessively complex or poorly constructed. Even when well-constructed to accurately display the characteristics of their data, graphs can be subject to different interpretation.

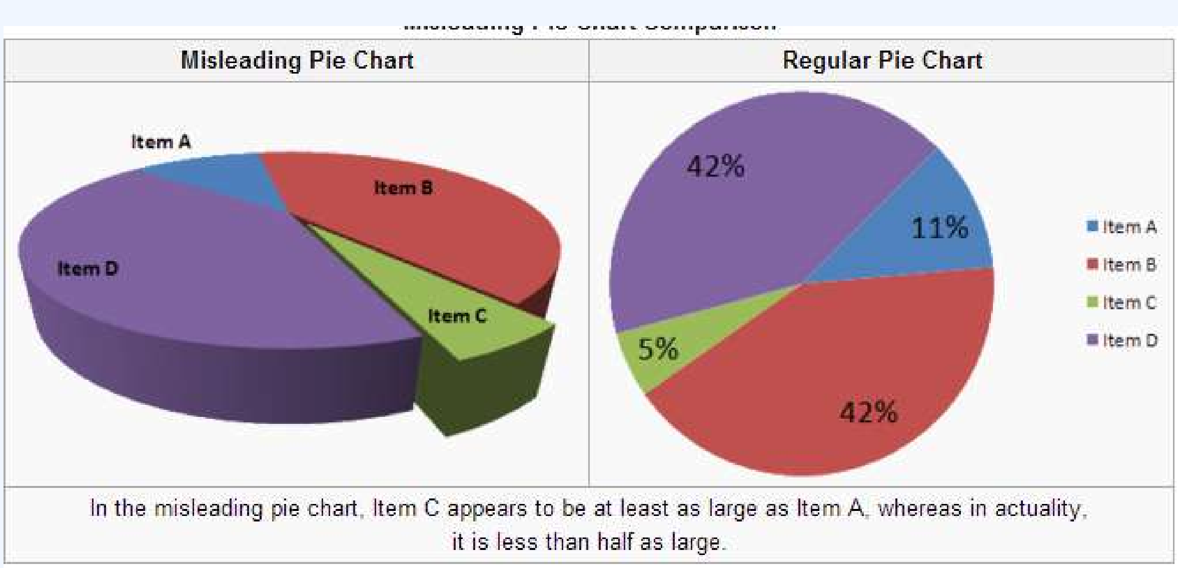
Misleading graphs may be created intentionally to hinder the proper interpretation of data, but can also be created accidentally by users for a variety of reasons including unfamiliarity with the graphing software, the misinterpretation of the data, or because the data cannot be accurately conveyed. Misleading graphs are often used in false advertising.

Types of Misleading Graphs

The use of graphs where they are not needed can lead to unnecessary confusion/interpretation. Generally, the more explanation a graph needs, the less the graph itself is needed. Graphs do not always convey information better than tables. This is often called excessive usage.

The use of biased or loaded words in the graph's title, axis labels, or caption may inappropriately prime the reader.

Pie charts can be especially misleading. Comparing pie charts of different sizes could be misleading as people cannot accurately read the comparative area of circles. The usage of thin slices which are hard to discern may be difficult to interpret. The usage of percentages as labels on a pie chart can be misleading when the sample size is small. A perspective (3D) pie chart is used to give the chart a 3D look. Often used for aesthetic reasons, the third dimension does not improve the reading of the data; on the contrary, these plots are difficult to interpret because of the distorted effect of perspective associated with the third dimension. In a 3D pie chart, the slices that are closer to the reader appear to be larger than those in the back due to the angle at which they're presented .

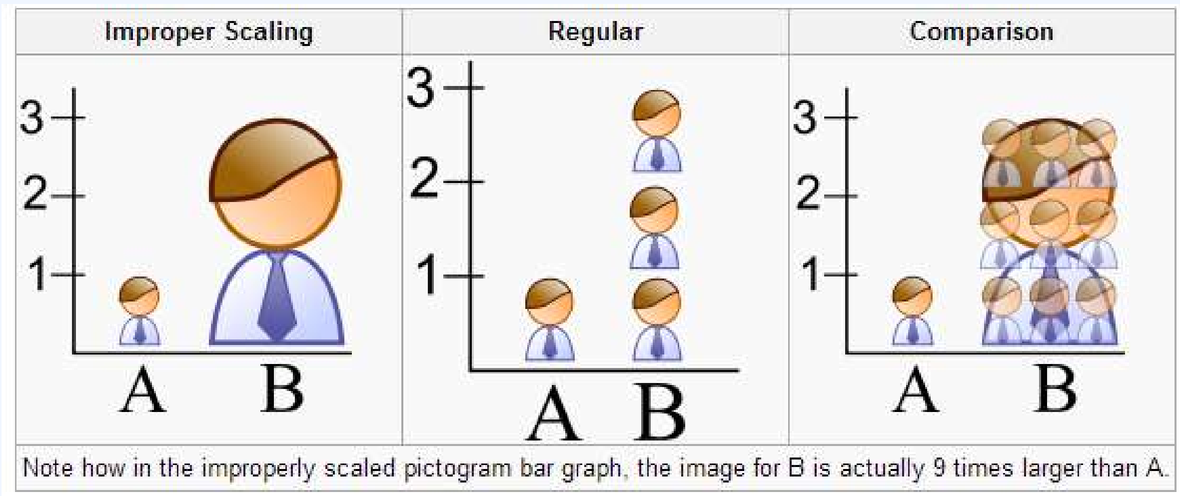


In the misleading pie chart, Item C appears to be at least as large as Item A, whereas in actuality, it is less than half as large.

3-D Pie Chart

In the misleading pie chart, Item C appears to be at least as large as Item A, whereas in actuality, it is less than half as large.

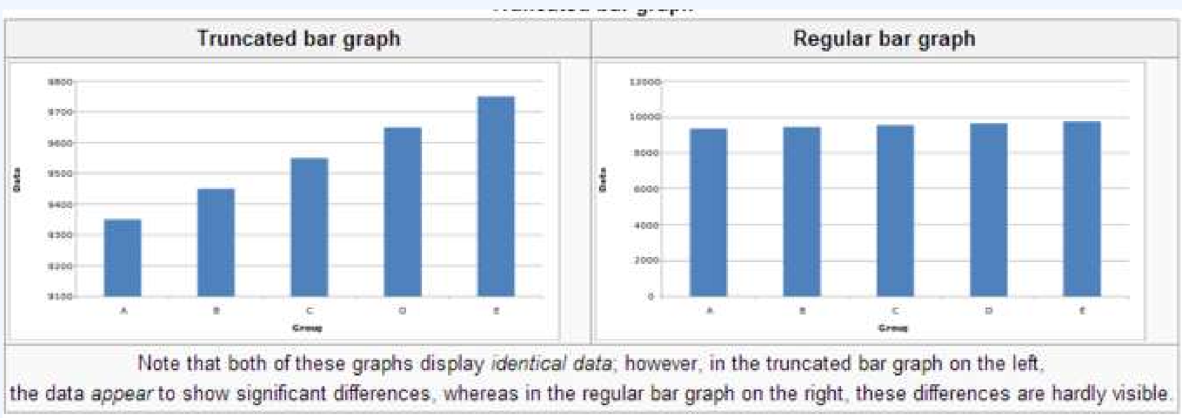
When using pictogram in bar graphs, they should not be scaled uniformly as this creates a perceptually misleading comparison. The area of the pictogram is interpreted instead of only its height or width. This causes the scaling to make the difference appear to be squared .



Improper Scaling

Note how in the improperly scaled pictogram bar graph, the image for B is actually 9 times larger than A.

A truncated graph has a y-axis that does not start at 0. These graphs can create the impression of important change where there is relatively little change .



Truncated Bar Graph

Note that both of these graphs display identical data; however, in the truncated bar graph on the left, the data appear to show significant differences, whereas in the regular bar graph on the right, these differences are hardly visible.

Usage in the Real World

Graphs are useful in the summary and interpretation of financial data. Graphs allow for trends in large data sets to be seen while also allowing the data to be interpreted by non-specialists. Graphs are often used in corporate annual reports as a form of impression management. In the United States, graphs do not have to be audited as they fall under AU Section 550 Other Information in Documents Containing Audited Financial Statements. Several published studies have looked at the usage of graphs in corporate reports for different corporations in different countries and have found frequent usage of improper design, selectivity, and measurement distortion within these reports. The presence of misleading graphs in annual reports have led to requests for standards to be set. Research has found that while readers with poor levels of financial understanding have a greater chance of being misinformed by misleading graphs, even those with financial understanding, such as loan officers, may be misled.

4.2.3: Graphs of Qualitative Data

Qualitative data can be graphed in various ways, including using pie charts and bar charts.

Learning Objective

Create a pie chart and bar chart representing qualitative data.

Key Takeaways

Key Points

- Since qualitative data represent individual categories, calculating descriptive statistics is limited. Mean, median, and measures of spread cannot be calculated; however, the mode can be calculated.
- One way in which we can graphically represent qualitative data is in a pie chart. Categories are represented by slices of the pie, whose areas are proportional to the percentage of items in that category.
- The key point about the qualitative data is that they do not come with a pre-established ordering (the way numbers are ordered).
- Bar charts can also be used to graph qualitative data. The Y axis displays the frequencies and the X axis displays the categories.

Key Term

descriptive statistics
: A branch of mathematics dealing with summarization and description of collections of data sets, including the concepts of arithmetic mean, median, and mode.

Qualitative Data

Recall the difference between quantitative and qualitative data. Quantitative data are data about numeric values. Qualitative data are measures of types and may be represented as a name or symbol. Statistics that describe or summarize can be produced for quantitative data and to a lesser extent for qualitative data. As quantitative data are always numeric they can be ordered, added together, and the frequency of an observation can be counted. Therefore, all descriptive statistics can be calculated using quantitative data. As qualitative data represent individual (mutually exclusive) categories, the descriptive statistics that can be calculated are limited, as many of these techniques require numeric values which can be logically ordered from lowest to highest and which express a count. Mode can be calculated, as it it the most frequency observed value. Median, measures of shape, measures of spread such as the range and interquartile range, require an ordered data set with a logical low-end value and high-end value. Variance and standard deviation require the mean to be calculated, which is not appropriate for categorical variables as they have no numerical value.

Graphing Qualitative Data

There are a number of ways in which qualitative data can be displayed. A good way to demonstrate the different types of graphs is by looking at the following example:

When Apple Computer introduced the iMac computer in August 1998, the company wanted to learn whether the iMac was expanding Apple's market share. Was the iMac just attracting previous Macintosh owners? Or was it purchased by newcomers to the computer market, and by previous Windows users who were switching over? To find out, 500 iMac customers were interviewed. Each customer was categorized as a previous Macintosh owners, a previous Windows owner, or a new computer purchaser. The qualitative data results were displayed in a frequency table.

Previous Ownership	Frequency	Relative Frequency
None	85	0.17
Windows	60	0.12
Mac	355	0.71
Total	500	1.00

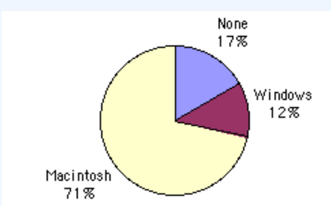
Frequency Table for Mac Data

The frequency table shows how many people in the study were previous Mac owners, previous Windows owners, or neither.

The key point about the qualitative data is that they do not come with a pre-established ordering (the way numbers are ordered). For example, there is no natural sense in which the category of previous Windows users comes before or after the category of previous iMac users. This situation may be contrasted with quantitative data, such as a person's weight. People of one weight are naturally ordered with respect to people of a different weight.

Pie Charts

One way in which we can graphically represent this qualitative data is in a pie chart. In a pie chart, each category is represented by a slice of the pie. The area of the slice is proportional to the percentage of responses in the category. This is simply the relative frequency multiplied by 100. Although most iMac purchasers were Macintosh owners, Apple was encouraged by the 12% of purchasers who were former Windows users, and by the 17% of purchasers who were buying a computer for the first time .



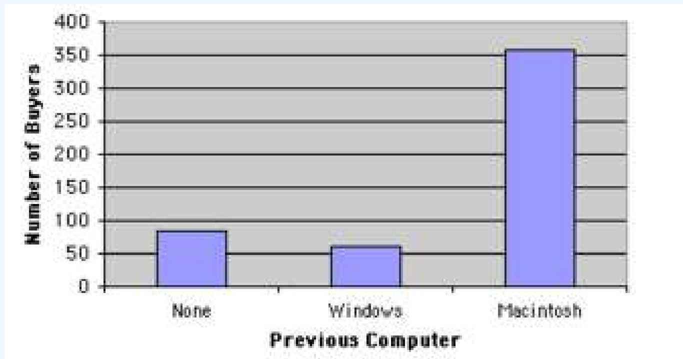
Pie Chart for Mac Data

The pie chart shows how many people in the study were previous Mac owners, previous Windows owners, or neither.

Pie charts are effective for displaying the relative frequencies of a small number of categories. They are not recommended, however, when you have a large number of categories. Pie charts can also be confusing when they are used to compare the outcomes of two different surveys or experiments.

Here is another important point about pie charts. If they are based on a small number of observations, it can be misleading to label the pie slices with percentages. For example, if just 5 people had been interviewed by Apple Computers, and 3 were former Windows users, it would be misleading to display a pie chart with the Windows slice showing 60%. With so few people interviewed, such a large percentage of Windows users might easily have accord since chance can cause large errors with small samples. In this case, it is better to alert the user of the pie chart to the actual numbers involved. The slices should therefore be labeled with the actual frequencies observed (e.g., 3) instead of with percentages.

Bar Charts



Bar Chart for Mac Data

The bar chart shows how many people in the study were previous Mac owners, previous Windows owners, or neither.

Bar charts can also be used to represent frequencies of different categories . Frequencies are shown on the Y axis and the type of computer previously owned is shown on the X axis. Typically the Y-axis shows the number of observations rather than the percentage of observations in each category as is typical in pie charts.

4.2.4: Misleading Graphs

A misleading graph misrepresents data and may result in incorrectly derived conclusions.

Learning Objectives

Criticize the practices of excessive usage, biased labeling, improper scaling, truncating, and the addition of a third dimension that often result in misleading graphs.

Key Takeaways

Key Points

- Misleading graphs may be created intentionally to hinder the proper interpretation of data, but can be also created accidentally by users for a variety of reasons.
- The use of graphs where they are not needed can lead to unnecessary confusion/interpretation. This is referred to as excessive usage.
- The use of biased or loaded words in the graph's title, axis labels, or caption may inappropriately sway the reader. This is called biased labeling.
- Graphs can also be misleading if they are improperly labeled, if they are truncated, if there is an axis change, if they lack a scale, or if they are unnecessarily displayed in the third dimension.

Key Terms

pictogram
: a picture that represents a word or an idea by illustration; used often in graphs

volatility
: the state of sharp and regular fluctuation

What is a Misleading Graph?

In statistics, a misleading graph, also known as a distorted graph, is a graph which misrepresents data, constituting a misuse of statistics and with the result that an incorrect conclusion may be derived from it. Graphs may be misleading through being excessively complex or poorly constructed. Even when well-constructed to accurately display the characteristics of their data, graphs can be subject to different interpretation.

Misleading graphs may be created intentionally to hinder the proper interpretation of data, but can be also created accidentally by users for a variety of reasons including unfamiliarity with the graphing software, the misinterpretation of the data, or because the data cannot be accurately conveyed. Misleading graphs are often used in false advertising. One of the first authors to write about misleading graphs was Darrell Huff, who published the best-selling book *How to Lie With Statistics* in 1954. It is still in print.

Excessive Usage

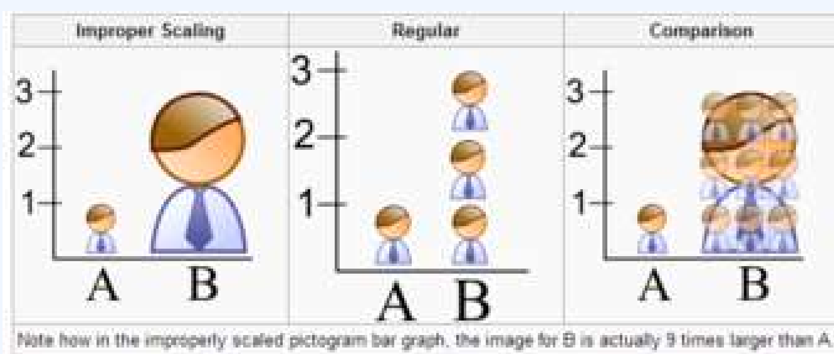
There are numerous ways in which a misleading graph may be constructed. The use of graphs where they are not needed can lead to unnecessary confusion/interpretation. Generally, the more explanation a graph needs, the less the graph itself is needed. Graphs do not always convey information better than tables.

Biased Labeling

The use of biased or loaded words in the graph's title, axis labels, or caption may inappropriately sway the reader.

Improper Scaling

When using pictogram in bar graphs, they should not be scaled uniformly as this creates a perceptually misleading comparison. The area of the pictogram is interpreted instead of only its height or width. This causes the scaling to make the difference appear to be squared.

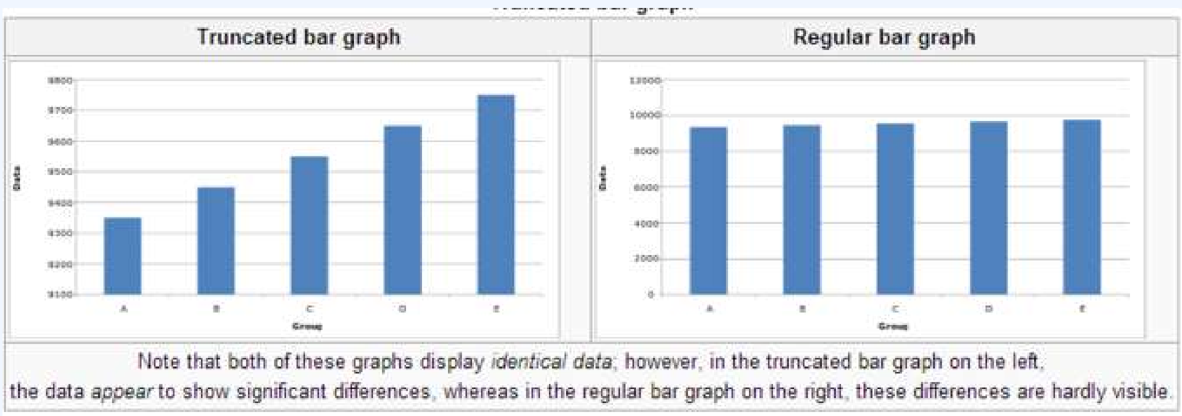


Improper Scaling

In the improperly scaled pictogram bar graph, the image for B is actually 9 times larger than A.

Truncated Graphs

A truncated graph has a y-axis that does not start at zero. These graphs can create the impression of important change where there is relatively little change.Truncated graphs are useful in illustrating small differences. Graphs may also be truncated to save space. Commercial software such as MS Excel will tend to truncate graphs by default if the values are all within a narrow range.

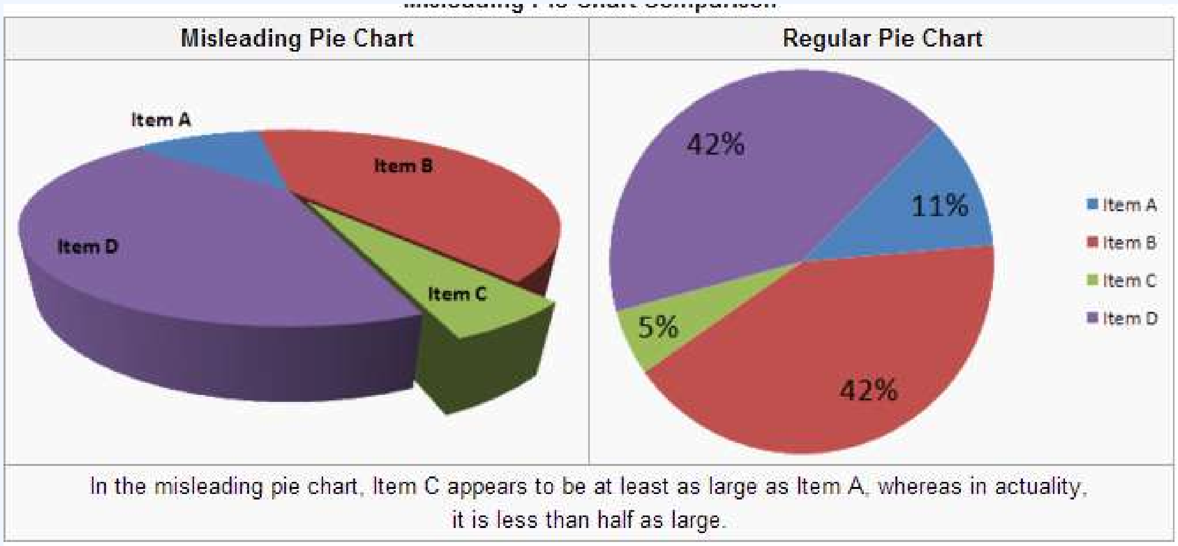


Truncated Bar Graph

Both of these graphs display identical data; however, in the truncated bar graph on the left, the data appear to show significant differences, whereas in the regular bar graph on the right, these differences are hardly visible.

Misleading 3D Pie Charts

A perspective (3D) pie chart is used to give the chart a 3D look. Often used for aesthetic reasons, the third dimension does not improve the reading of the data; on the contrary, these plots are difficult to interpret because of the distorted effect of perspective associated with the third dimension. The use of superfluous dimensions not used to display the data of interest is discouraged for charts in general, not only for pie charts. In a 3D pie chart, the slices that are closer to the reader appear to be larger than those in the back due to the angle at which they're presented .



Misleading 3D Pie Chart

In the misleading pie chart, Item C appears to be at least as large as Item A, whereas in actuality, it is less than half as large.

Other Misleading Graphs

Graphs can also be misleading for a variety of other reasons. An axis change affects how the graph appears in terms of its growth and volatility. A graph with no scale can be easily manipulated to make the difference between bars look larger or smaller than they actually are. Improper intervals can affect the appearance of a graph, as well as omitting data. Finally, graphs can also be misleading if they are overly complex or poorly constructed.

Graphs in Finance and Corporate Reports

Graphs are useful in the summary and interpretation of financial data. Graphs allow for trends in large data sets to be seen while also allowing the data to be interpreted by non-specialists. Graphs are often used in corporate annual reports as a form of impression management. In the United States, graphs do not have to be audited. Several published studies have looked at the usage of graphs in corporate reports for different corporations in different countries and have found frequent usage of improper design, selectivity, and measurement distortion within these reports. The presence of misleading graphs in annual reports have led to requests for standards to be set. Research has found that while readers with poor levels of financial understanding have a greater chance of being misinformed by misleading graphs, even those with financial understanding, such as loan officers, may be misled.

4.2.5: Do It Yourself: Plotting Qualitative Frequency Distributions

Qualitative frequency distributions can be displayed in bar charts, Pareto charts, and pie charts.

Learning Objectives

Outline the steps necessary to plot a frequency distribution for qualitative data.

Key Takeaways

Key Points

- The first step to plotting a qualitative frequency distributions is to create a frequency table.
- If drawing a bar graph or Pareto chart, first draw two axes. The y-axis is labeled with the frequency (or relative frequency) and the x-axis is labeled with the category.
- In bar graphs and Pareto graphs, draw rectangles of equal width and heights that correspond to their frequencies/relative frequencies.
- A pie chart shows the distribution in a different way, where each percentage is a slice of the pie.

Key Terms

relative frequency distribution
: a representation, either in graphical or tabular format, which displays the fraction of

observations in a certain category

frequency distribution
: a representation, either in a graphical or tabular format, which displays the number of observations within a given interval

Pareto chart
: a type of bar graph where where the bars are drawn in decreasing order of frequency or relative frequency

Ways to Organize Data

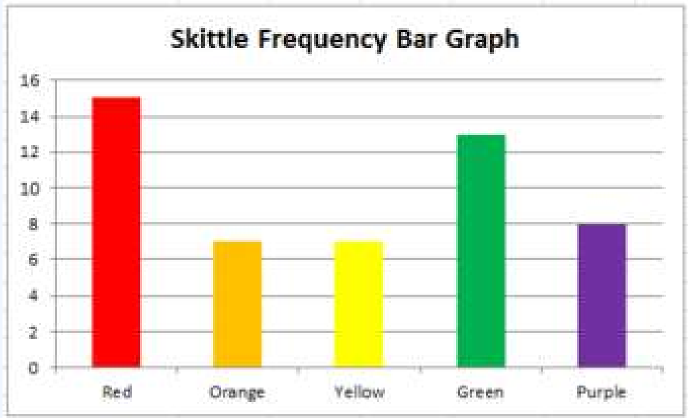
When data is collected from a survey or an experiment, they must be organized into a manageable form. Data that is not organized is referred to as raw data. A few different ways to organize data include tables, graphs, and numerical summaries.

One common way to organize qualitative, or categorical, data is in a frequency distribution. A frequency distribution lists the number of occurrences for each category of data.

Step-by-Step Guide to Plotting Qualitative Frequency Distributions

The first step towards plotting a qualitative frequency distribution is to create a table of the given or collected data. For example, let's say you want to determine the distribution of colors in a bag of Skittles. You open up a bag, and you find that there are 15 red, 7 orange, 7 yellow, 13 green, and 8 purple. Create a two column chart, with the titles of Color and Frequency, and fill in the corresponding data.

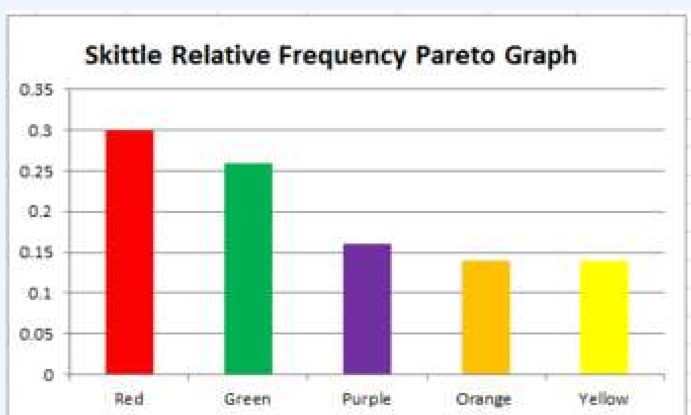
To construct a frequency distribution in the form of a bar graph, you must first draw two axes. The y-axis (vertical axis) should be labeled with the frequencies and the x-axis (horizontal axis) should be labeled with each category (in this case, Skittle color). The graph is completed by drawing rectangles of equal width for each color, each as tall as their frequency .



Bar Graph

This graph shows the frequency distribution of a bag of Skittles.

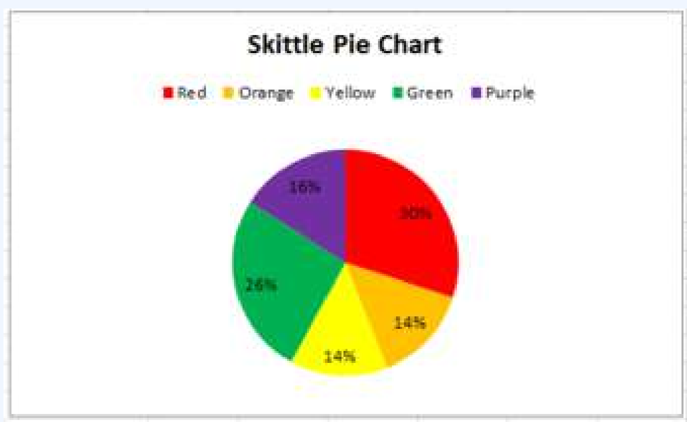
Sometimes a relative frequency distribution is desired. If this is the case, simply add a third column in the table called Relative Frequency. This is found by dividing the frequency of each color by the total number of Skittles (50, in this case). This number can be written as a decimal, a percentage, or as a fraction. If we decided to use decimals, the relative frequencies for the red, orange, yellow, green, and purple Skittles are respectively 0.3, 0.14, 0.14, 0.26, and 0.16. The decimals should add up to 1 (or very close to it due to rounding). Bar graphs for relative frequency distributions are very similar to bar graphs for regular frequency distributions, except this time, the y-axis will be labeled with the relative frequency rather than just simply the frequency. A special type of bar graph where the bars are drawn in decreasing order of relative frequency is called a Pareto chart .



Pareto Chart

This graph shows the relative frequency distribution of a bag of Skittles.

The distribution can also be displayed in a pie chart, where the percentages of the colors are broken down into slices of the pie. This may be done by hand, or by using a computer program such as Microsoft Excel . If done by hand, you must find out how many degrees each piece of the pie corresponds to. Since a circle has 360 degrees, this is found out by multiplying the relative frequencies by 360. The respective degrees for red, orange, yellow, green, and purple in this case are 108, 50.4, 50.4, 93.6, and 57.6. Then, use a protractor to properly draw in each slice of the pie.



Pie Chart

This pie chart shows the frequency distribution of a bag of Skittles.

4.2.6: Summation Notation

In statistical formulas that involve summing numbers, the Greek letter sigma is used as the summation notation.

Learning Objectives

Discuss the summation notation and identify statistical situations in which it may be useful or even essential.

Key Takeaways

Key Points

- There is no special notation for the summation of explicit sequences (such as 1+2+4+2), as the corresponding repeated addition expression will do.
- If the terms of the sequence are given by a regular pattern, possibly of variable length, then the summation notation may be useful or even essential.
- In general, mathematicians use the following sigma notation: ∑i=mnai">$\sum_{i=m}^{n} a_i$, where m is the lower bound, n is the upper bound, i is the index of summation, and a(i) represents each successive term to be added.

Key Terms

summation notation
: a notation, given by the Greek letter sigma, that denotes the operation of adding a sequence of numbers

ellipsis
: a mark consisting of three periods, historically with spaces in between, before, and after them ” . . . “, nowadays a single character ” (used in printing to indicate an omission)

Summation

Many statistical formulas involve summing numbers. Fortunately there is a convenient notation for expressing summation. This section covers the basics of this summation notation.

Summation is the operation of adding a sequence of numbers, the result being their sum or total. If numbers are added sequentially from left to right, any intermediate result is a partial sum, prefix sum, or running total of the summation. The numbers to be summed (called addends, or sometimes summands) may be integers, rational numbers, real numbers, or complex numbers. Besides numbers, other types of values can be added as well: vectors, matrices, polynomials and, in general, elements of any additive group. For finite sequences of such elements, summation always produces a well-defined sum.

The summation of the sequence [1, 2, 4, 2] is an expression whose value is the sum of each of the members of the sequence. In the example, 1+2+4+2=9. Since addition is associative, the value does not depend on how the additions are grouped. For instance (1+2) + (4+2) and 1 + ((2+4) + 2) both have the value 9; therefore, parentheses are usually omitted in repeated additions. Addition is also commutative, so changing the order of the terms of a finite sequence does not change its sum.

Notation

There is no special notation for the summation of such explicit sequences as the example above, as the corresponding repeated addition expression will do. If, however, the terms of the sequence are given by a regular pattern, possibly of variable length, then a summation operator may be useful or even essential.

For the summation of the sequence of consecutive integers from 1 to 100 one could use an addition expression involving an ellipsis to indicate the missing terms: 1+2+3+4+⋯+99+100">1+2+3+4+⋯+99+100. In this case the reader easily guesses the pattern; however, for more complicated patterns, one needs to be precise about the rule used to find successive terms. This can be achieved by using the summation notation "Σ">Σ " Using this sigma notation, the above summation is written as:

∑i=1100i">$\sum_{i=1}^{100} i$

In general, mathematicians use the following sigma notation: ∑i=mnai">$\sum_{i=m}^{n} i$

In this notation, i">i represents the index of summation, ai">ai is an indexed variable representing each successive term in the series, m">m is the lower bound of summation, and n">n is the upper bound of summation. The "i=m">i=m" under the summation symbol means that the index i">i starts out equal to m">m. The index, i">i, is incremented by 1 for each successive term, stopping when i=n">i=n.

Here is an example showing the summation of exponential terms (terms to the power of 2):

∑i=3612=32+42+52+62=86">$\sum_{i=3}^{6} 1^2 = 3^2 + 4^2 + 5^2 + 6^2 = 86$

Informal writing sometimes omits the definition of the index and bounds of summation when these are clear from context, as in:

∑ai2=∑i=1nai2">$\sum a_i^2 = \sum_{i=1}^{n} a_i^2$

One often sees generalizations of this notation in which an arbitrary logical condition is supplied, and the sum is intended to be taken over all values satisfying the condition. For example, the sum of f(k)">f(k) over all integers k">k in the specified range can be written as: ∑0≤k">$\sum_{0\leq k}$

The sum of f(x)">f(x) over all elements x">x in the set S">S can be written as: ∑xϵSf(x)"> $\sum_{xeS} f(x)$

4.2.7: Graphing Bivariate Relationships

We can learn much more by displaying bivariate data in a graphical form that maintains the pairing of variables.

Learning Objectives

Compare the strengths and weaknesses of the various methods used to graph bivariate data.

Key Takeaways

Key Points

- When one variable increases with the second variable, we say that x and y have a positive association.
- Conversely, when y decreases as x increases, we say that they have a negative association.
- The presence of qualitative data leads to challenges in graphing bivariate relationships.
- If both variables are qualitative, we would be able to graph them in a contingency table.

Key Terms

bivariate
: Having or involving exactly two variables.

contingency table
: a table presenting the joint distribution of two categorical variables

skewed
: Biased or distorted (pertaining to statistics or information).

Introduction to Bivariate Data

Measures of central tendency, variability, and spread summarize a single variable by providing important information about its distribution. Often, more than one variable is collected on each individual. For example, in large health studies of populations it is common to obtain variables such as age, sex, height, weight, blood pressure, and total cholesterol on each individual. Economic studies may be interested in, among other things, personal income and years of education. As a third example, most university admissions committees ask for an applicant's high school grade point average and standardized admission test scores (e.g., SAT). In the following text, we consider bivariate data, which for now consists of two quantitative variables for each individual. Our first interest is in summarizing such data in a way that is analogous to summarizing univariate (single variable) data.

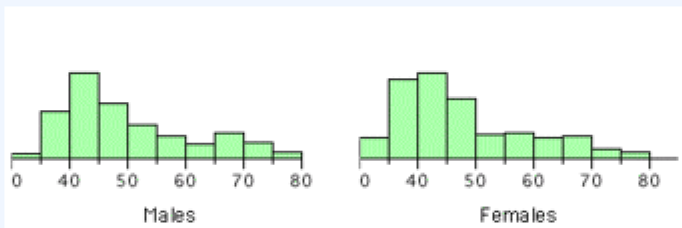
By way of illustration, let's consider something with which we are all familiar: age. More specifically, let's consider if people tend to marry other people of about the same age. One way to address the question is to look at pairs of ages for a sample of married couples. Bivariate Sample 1 shows the ages of 10 married couples. Going across the columns we see that husbands and wives tend to be of about the same age, with men having a tendency to be slightly older than their wives.

Couple	A	B	C	D	E	F	G	H	I	J
Husband	36	72	37	36	51	50	47	50	37	41
Wife	35	67	33	35	50	46	47	42	36	41

Bivariate Sample 1

Sample of spousal ages of 10 white American couples.

These pairs are from a dataset consisting of 282 pairs of spousal ages (too many to make sense of from a table). What we need is a way to graphically summarize the 282 pairs of ages, such as a histogram. as in .



Bivariate Histogram

Histogram of spousal ages.

Each distribution is fairly skewed with a long right tail. From the first figure we see that not all husbands are older than their wives. It is important to see that this fact is lost when we separate the variables. That is, even though we provide summary statistics on each variable, the pairing within couples is lost by separating the variables. Only by maintaining the pairing can meaningful answers be found about couples, per se.

Therefore, we can learn much more by displaying the bivariate data in a graphical form that maintains the pairing. shows a scatter plot of the paired ages. The x-axis represents the age of the husband and the y-axis the age of the wife.

Bivariate Scatterplot

Scatterplot showing wife age as a function of husband age.

There are two important characteristics of the data revealed by this figure. First, it is clear that there is a strong relationship between the husband's age and the wife's age: the older the husband, the older the wife. When one variable increases with the second variable, we say that x and y have a positive association. Conversely, when y decreases as x increases, we say that they have a negative association. Second, the points cluster along a straight line. When this occurs, the relationship is called a linear relationship.

Bivariate Relationships in Qualitative Data

The presence of qualitative data leads to challenges in graphing bivariate relationships. We could have one qualitative variable and one quantitative variable, such as SAT subject and score. However, making a scatter plot would not be possible as only one variable is numerical. A bar graph would be possible.

If both variables are qualitative, we would be able to graph them in a contingency table. We can then use this to find whatever information we may want. In , this could include what percentage of the group are female and right-handed or what percentage of the males are left-handed.

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Totals	87	13	100

Contingency Table

Contingency tables are useful for graphically representing qualitative bivariate relationships.

Attributions

- Describing Qualitative Data
    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "qualitative analysis."
      http://en.wikipedia.org/wiki/qualitative%20analysis.
      Wikipedia
      CC BY-SA 3.0.
    - "Qualitative research."
      http://en.wikipedia.org/wiki/Qualitative_research.
      Wikipedia
      CC BY-SA 3.0.
    - "ordinal."
      http://en.wiktionary.org/wiki/ordinal.
      Wiktionary
      CC BY-SA 3.0.
    - "nominal."
      http://en.wiktionary.org/wiki/nominal.
      Wiktionary
      CC BY-SA 3.0.
    - "Statistics/Different Types of Data/Quantitative and Qualitative Data."

http://en.wikibooks.org/wiki/Statistics/Different_Types_of_Data/Quantitative_and_Qualitative_Data%23Qualitative_data.
Wikibooks
CC BY-SA 3.0.
    - "Social Research Methods/Qualitative Research."
      http://en.wikibooks.org/wiki/Social_Research_Methods/Qualitative_Research.
      Wikibooks
      CC BY-SA 3.0.
- Interpreting Distributions Constructed by Others
    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "Misleading graph."
      http://en.wikipedia.org/wiki/Misleading_graph.
      Wikipedia
      CC BY-SA 3.0.
    - "distribution."
      http://en.wiktionary.org/wiki/distribution.
      Wiktionary
      CC BY-SA 3.0.
    - "truncate."
      http://en.wiktionary.org/wiki/truncate.
      Wiktionary
      CC BY-SA 3.0.
    - "bias."
      http://en.wiktionary.org/wiki/bias.
      Wiktionary
      CC BY-SA 3.0.
    - "Misleading graph."
      http://en.wikipedia.org/wiki/Misleading_graph.
      Wikipedia
      GNU FDL.
    - "Misleading graph."
      http://en.wikipedia.org/wiki/Misleading_graph.
      Wikipedia
      GNU FDL.

    - "Misleading graph."
      http://en.wikipedia.org/wiki/Misleading_graph.
      Wikipedia
      GNU FDL.
- Graphs of Qualitative Data
    - "Boundless."
      http://www.boundless.com/.
      Boundless Learning
      CC BY-SA 3.0.
    - "descriptive statistics."
      http://en.wiktionary.org/wiki/descriptive_statistics.
      Wiktionary
      CC BY-SA 3.0.
    - "David Lane, Graphing Qualitative Variables. September 17, 2013."
      http://cnx.org/content/m10927/latest/.
      OpenStax CNX
      CC BY 3.0.
    - "Error 404."
      http://www.abs.gov.au/websitedbs/a3121120.nsf/89a5f3d8684682b6ca256de4002c809b/e200e8e572a2ae52ca25794900127f4f!OpenDocument.
      Austrailian Bureau of Statistics
      CC BY.
    - "David Lane, Graphing Qualitative Variables. April 22, 2013."
      http://cnx.org/content/m10927/latest/.
      OpenStax CNX
      CC BY 3.0.
    - "David Lane, Graphing Qualitative Variables. April 22, 2013."
      http://cnx.org/content/m10927/latest/.
      OpenStax CNX
      CC BY 3.0.
    - "David Lane, Graphing Qualitative Variables. April 22, 2013."
      http://cnx.org/content/m10927/latest/.
      OpenStax CNX
      CC BY 3.0.
- Misleading Graphs
    - "Boundless."
      http://www.boundless.com/.

Boundless Learning
CC BY-SA 3.0.
    - "Misleading graph."
http://en.wikipedia.org/wiki/Misleading_graph.
Wikipedia
CC BY-SA 3.0.
    - "volatility."
http://en.wiktionary.org/wiki/volatility.
Wiktionary
CC BY-SA 3.0.
    - "pictogram."
http://en.wiktionary.org/wiki/pictogram.
Wiktionary
CC BY-SA 3.0.
    - "Misleading graph."
http://en.wikipedia.org/wiki/Misleading_graph.
Wikipedia
GNU FDL.
    - "Misleading graph."
http://en.wikipedia.org/wiki/Misleading_graph.
Wikipedia
GNU FDL.
    - "Misleading Graph."
http://en.wikipedia.org/wiki/Misleading_graph.
Wikipedia
GNU FDL.
- Do It Yourself: Plotting Qualitative Frequency Distributions
    - "Boundless."
http://www.boundless.com/.
Boundless Learning
CC BY-SA 3.0.
    - "Frequency distribution."
http://en.wikipedia.org/wiki/Frequency_distribution.
Wikipedia
CC BY-SA 3.0.
    - "Microsoft Excel – spreadsheet software – Office.com."
http://office.microsoft.com/en-us/excel/.

Microsoft
License: Other.
  - "Microsoft Excel – spreadsheet software – Office.com."
    http://office.microsoft.com/en-us/excel/.
    Microsoft
    License: Other.
  - "Microsoft Excel – spreadsheet software – Office.com."
    http://office.microsoft.com/en-us/excel/.
    Microsoft
    License: Other.
- Summation Notation
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "summation notation."
    http://en.wikipedia.org/wiki/summation%20notation.
    Wikipedia
    CC BY-SA 3.0.
  - "Summation."
    http://en.wikipedia.org/wiki/Summation%23Capital-sigma_notation.
    Wikipedia
    CC BY-SA 3.0.
  - "ellipsis."
    http://en.wiktionary.org/wiki/ellipsis.
    Wiktionary
    CC BY-SA 3.0.
- Graphing Bivariate Relationships
  - "Boundless."
    http://www.boundless.com/.
    Boundless Learning
    CC BY-SA 3.0.
  - "Bivariate Data Tutorial | Sophia Learning."
    http://www.sophia.org/bivariate-data-tutorial.
    Sophia Learning Online
    CC BY.
  - "bivariate."

http://en.wiktionary.org/wiki/bivariate.
Wiktionary
CC BY-SA 3.0.

- "contingency table."
  http://en.wiktionary.org/wiki/contingency_table.
  Wiktionary
  CC BY-SA 3.0.
- "skewed."
  http://en.wiktionary.org/wiki/skewed.
  Wiktionary
  CC BY-SA 3.0.
- "David Lane, Introduction to Bivariate Data. September 17, 2013."
  http://cnx.org/content/m10949/latest/.
  OpenStax CNX
  CC BY 3.0.
- "David Lane, Introduction to Bivariate Data. May 6, 2013."
  http://cnx.org/content/m10949/latest/.
  OpenStax CNX
  CC BY 3.0.
- "David Lane, Introduction to Bivariate Data. May 6, 2013."
  http://cnx.org/content/m10949/latest/.
  OpenStax CNX
  CC BY 3.0.
- "Contingency table."
  http://en.wikipedia.org/wiki/Contingency_table.
  Wikipedia
  GNU FDL.
- "David Lane, Introduction to Bivariate Data. May 6, 2013."
  http://cnx.org/content/m10949/latest/.
  OpenStax CNX
  CC BY 3.0.

SECTION VII

4.XLSX – EXCEL CHALLENGE - MATHEMATICAL COMPUTATIONS

Perhaps the most valuable feature of Excel is its ability to produce mathematical outputs using the data in a workbook. This chapter reviews several mathematical outputs that you can produce in Excel through the construction of formulas and functions. The chapter begins with the construction of formulas for basic and complex mathematical computations. The second section reviews statistical functions, such as SUM, AVERAGE, MIN, and MAX, which can be applied to a range of cells. The last section of the chapter addresses functions used to calculate mortgage and lease payments as well as the valuation of investments. This chapter also shows how you can use data from multiple worksheets to construct formulas and functions. These skills will be demonstrated in the context of a personal cash budget, which is a vital tool for managing your money for long-term financial security. The personal budget objective will also provide you with several opportunities to demonstrate Excel's what-if scenario capabilities, which highlight how formulas and functions automatically produce new outputs when one or more inputs are changed.

Attribution

Adapted by Mary Schatz from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

4.XLSX.1 FORMULAS

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Learning Objectives

- Learn how to create basic formulas.
- Understand relative referencing when copying and pasting formulas.
- Work with complex formulas by controlling the order of mathematical operations.
- Understand formula auditing tools.

This section reviews the fundamental skills for entering formulas into an Excel worksheet. The example used for this chapter is the construction of a personal budget. Most financial advisors recommend that all households construct and maintain a personal budget to achieve and maintain strong financial health. Organizing and maintaining a personal budget is a skill you can practice at any point in your life. Whether you are managing your expenses during college or maintaining the finances of a family of four, a personal budget can be a vital tool when making financial decisions. Excel can make managing your money a fun and rewarding exercise.

Open the Data File

Download Data File: CH2 Data

1. Open the Data file named **CH2 Data** and use the File/Save As command to save it with the new name **CH2 Personal Budget**.

Figure 2.1 shows the completed workbook that will be demonstrated in this chapter. Notice that this workbook contains four worksheets. The first worksheet, **Budget Summary**, serves as an overview of the data that was entered and calculated in the second and third worksheets, **Budget Detail** and **Loan Payments**. The second worksheet, **Budget Detail**, provides a detailed list of all the expenses and the third worksheet,

Loan Payments, provides information regarding car payment and mortgage payment amounts. The last worksheet, **Prepare to Print**, has data that is unrelated to the budget worksheets but will be used in Section 2.4 – Preparing to Print.

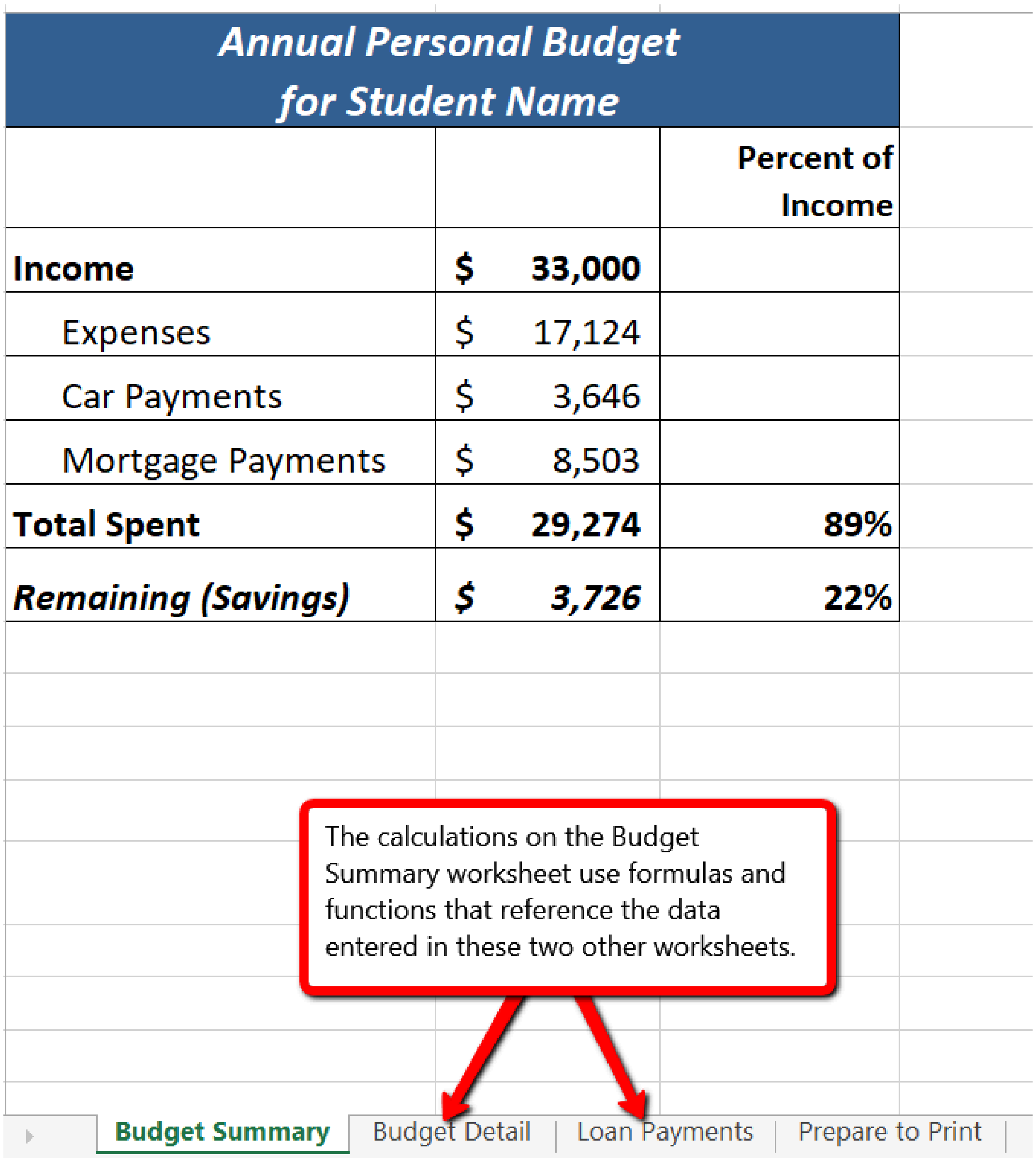


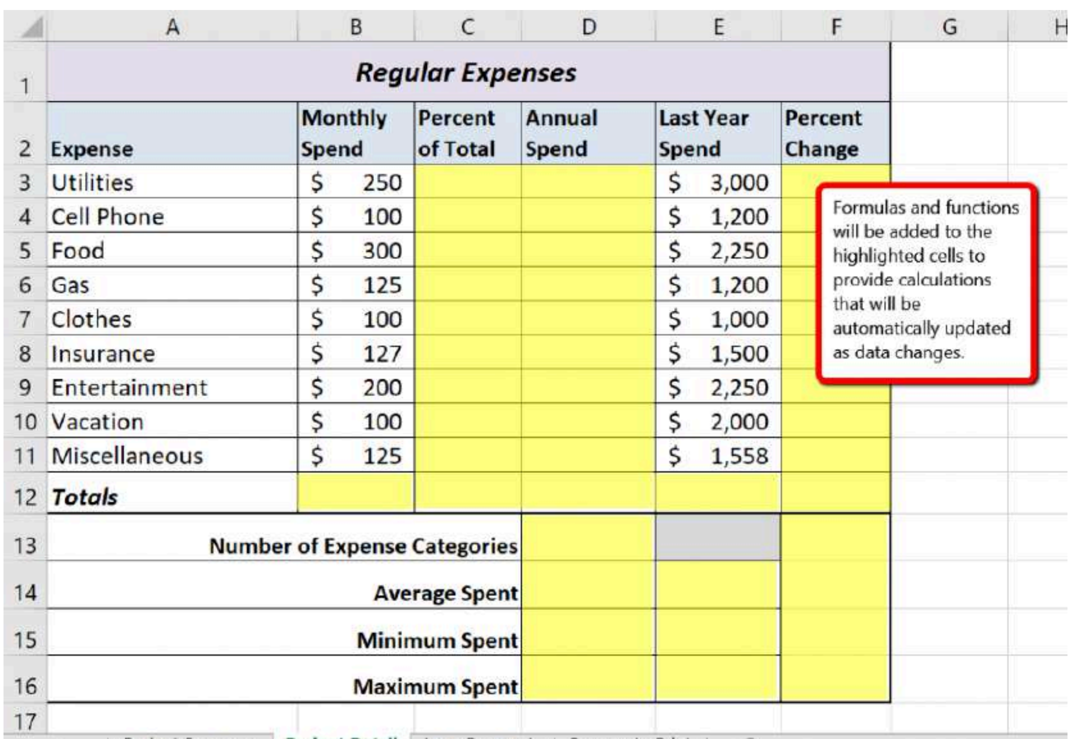
Figure 2.1 – The completed Budget Summary worksheet

Creating a Basic Formula

When formulas and cell references are used Excel will automatically recalculate when data is changed

Formulas are used to calculate a variety of mathematical outputs in Excel and can be used to create virtually any custom calculation required for your objective. Furthermore, when constructing a formula in Excel, you use cell addresses that, when added to a formula, become cell references. This means that Excel uses, or references, the number entered into the cell location when performing the calculation. As a result, when the numbers in the cells that are referenced are changed, Excel automatically recalculates the formula and produces a new result. This is what gives Excel the ability to create a variety of what-if scenarios, which will be explained later in the chapter.

To demonstrate the construction of a basic formula, we will begin working on the **Budget Detail** worksheet, which is shown in **Figure 2.2**. To complete this worksheet, we will enter some data, and then create several formulas and functions. **Table 2.1** provides definitions for each of the spend categories listed in the range A3:A11. When you develop a personal budget, these categories are defined on the basis of how you spend your money. It is likely that every person could have different categories or define the same categories differently. Therefore, it is important to review the definitions in **Table 2.1** to understand how we are defining these categories before proceeding.



	A	B	C	D	E	F
1	***Regular Expenses***					
2	**Expense**	**Monthly Spend**	**Percent of Total**	**Annual Spend**	**Last Year Spend**	**Percent Change**
3	Utilities	$ 250			$ 3,000	
4	Cell Phone	$ 100			$ 1,200	
5	Food	$ 300			$ 2,250	
6	Gas	$ 125			$ 1,200	
7	Clothes	$ 100			$ 1,000	
8	Insurance	$ 127			$ 1,500	
9	Entertainment	$ 200			$ 2,250	
10	Vacation	$ 100			$ 2,000	
11	Miscellaneous	$ 125			$ 1,558	
12	***Totals***					
13	**Number of Expense Categories**					
14	**Average Spent**					
15	**Minimum Spent**					
16	**Maximum Spent**					
17						

Figure 2.2 Budget Detail Worksheet

Table 2.1 Spend Category Definitions

Category	Definition
Utilities	Electricity, heat, water, home phone, cable, Internet access
Cell Phone	Cell phone plan and equipment charges
Food	Groceries
Gas	Cost of gas for vehicle
Clothes	Clothes, shoes, and accessories
Insurance	Renter, homeowner, and/or car insurance
Entertainment	Activities like dining out, movie and theater tickets, parties, and so on
Vacation	Vacation expenses
Miscellaneous	Any other spending categories

The amount of money spent each month for each category, as well as the amount of money spent last year, is already entered into the worksheet. We will write formulas that will calculate the annual (yearly) amount spent, the percent of the total spent each category represents, as well as the percent change from last year's spending to the current year.

The first formula will calculate the Annual Spend values. The formula will be constructed so that it takes the values in the Monthly Spend column and multiplies them by 12 (the number of months in a year). This will show how much money will be spent per year for each of the categories listed in Column A. Since the first category is **Utilities**, we will start by creating the formula to multiply the Monthly Spend amount in B3 by 12. This formula will be created in D3 – the Annual Spend cell for the Utilities category. This formula will be written as: =**B3*12**

1. Switch to the **Budget Detail** worksheet if needed. Click cell **D3**.
2. Type an equal sign =
   When the first character entered into a cell is an equal sign, it signals Excel to perform a calculation.
3. Type **B3**. *This adds B3 to the formula, which is now a cell reference. Excel will use whatever value is entered into cell B3 in the calculation.*
4. Type the * . *This is the symbol for multiplication in Excel. As shown in* ***Table 2.2*** *the mathematical operators in Excel are slightly different from those found on a typical calculator.*
5. Type the number **12**. *This multiplies the value in cell B3 by 12. In this formula, a number, or constant, is used instead of a cell reference because it will not change. In other words, there will always be 12 months in*

Formulas always start with the equal sign. This signifies to Excel that the contents of the cell should be calculated, not just displayed as basic text or numbers.

a year.

6. Press the **ENTER** key. *This enters the formula into the cell.*

Table 2.2 Excel Mathematical Operators (move up)

Symbol	Operation
+	Addition
–	Subtraction
/	Division
*	Multiplication
^	Power/Exponent

Why?

Use Cell References

Cell references enable Excel to automatically recalculate when one or more inputs in the referenced cells are changed. Cell references also allow you to trace how results are being calculated in a formula. You should **never** use a calculator to determine a mathematical output and type it into the cell location of a worksheet. Doing so eliminates Excel's cell-referencing benefits as well as your ability to trace a formula to determine how results are being calculated.

Use Universal Constants

There will be times when you are writing formulas that you will need to use universal constants, or numbers that do not change, such as the number of days in a week, weeks or months in a year, and so on. For example, if you are calculating the monthly cost of an item when you know the yearly cost, you will always divide by 12 since there are 12 months in a year. In this case, you use the constant of 12 instead of a cell reference because the number of months in a year never changes.

Figure 2.3 shows how the formula appears in cell D3 before you press the ENTER key. **Figure 2.4** shows the result of the formula after you press the ENTER key, as well as the **formula bar** which displays the formula as it was entered in the cell.

The Annual Spend for Utilities is $3,000 because the formula is taking the Monthly Spend in cell B3 and multiplying it by 12. If the value in cell B3 is changed, the formula automatically produces a new result.

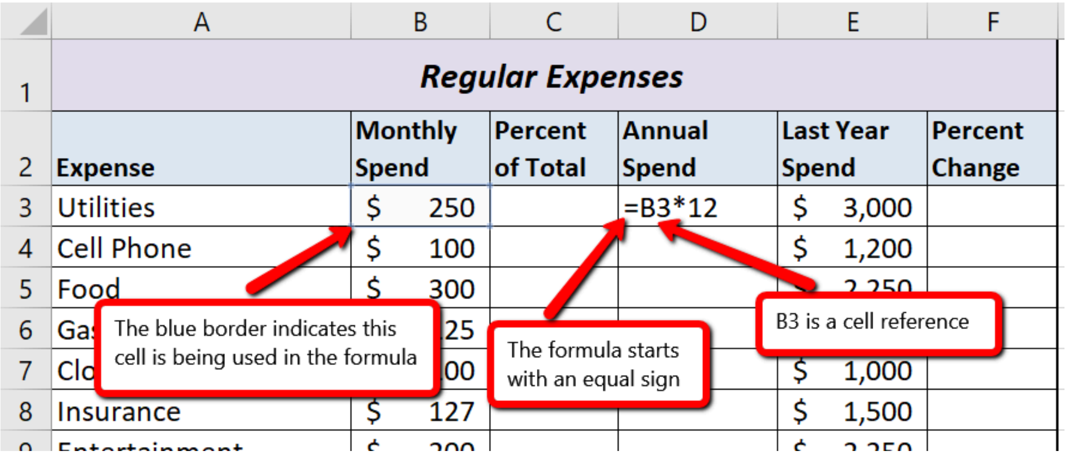


Figure 2.3 Adding a Formula to a Worksheet

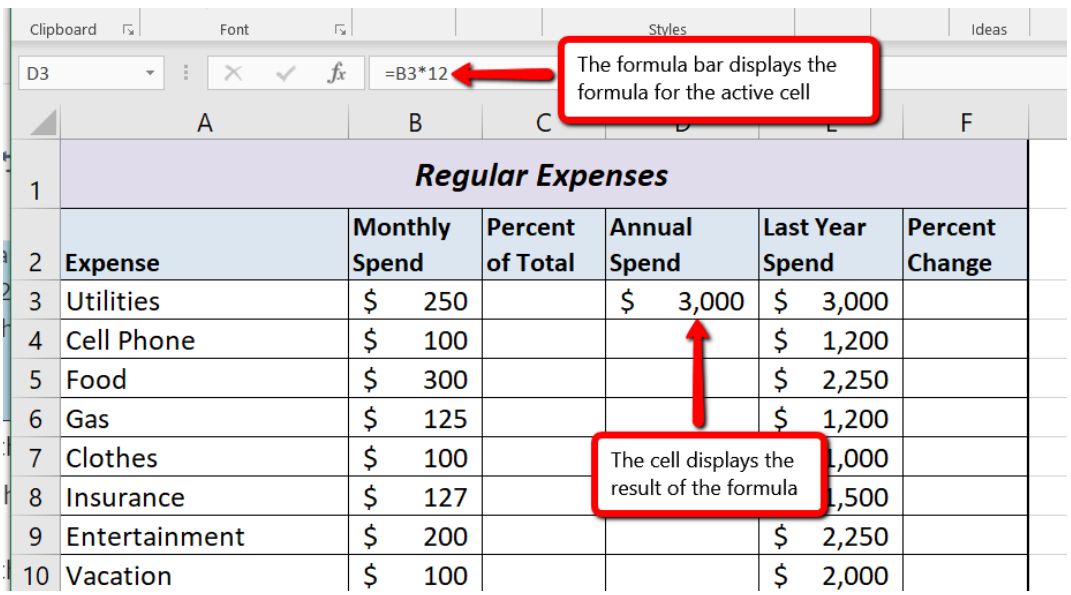


Figure 2.4 Formula Output for Annual Spend

Relative References (Copying and Pasting Formulas)

Once a formula is typed into a worksheet, it can be copied and pasted to other cell locations. For example, in cell D3 we have calculated the annual spend for the Utilities category, but this calculation needs to be performed for the rest of the cell locations in Column D. Since we used the B3 cell reference in the formula, Excel automatically adjusts that cell reference when the formula is copied and pasted into the rest of the cell locations in the column. This is called relative referencing and is demonstrated as follows:

1. Click cell D3.
2. Place the mouse pointer over the Auto Fill Handle in the bottom right corner of the cell.

3. When the mouse pointer turns from a white block plus sign to a black plus sign, click and drag down to cell D11. This pastes the formula into the range D4:D11.
4. Double click cell D6. Notice that the cell reference in the formula is automatically changed to B6.
5. Press the ENTER key.

Figure 2.5 shows the results added to the rest of the cell locations in the Annual Spend column. For each row, the formula takes the value in the Monthly Spend column and multiplies it by 12. You will also see that cell D6 has been double clicked to show the formula. Notice that Excel automatically changed the original cell reference of B3 to B6. This is the result of relative referencing, which means Excel automatically adjusts a cell reference relative to its original location when it is pasted into new cell locations. In this example, the formula was pasted into eight cell locations below the original cell location. As a result, Excel increased the row number of the original cell reference by a value of one for each row it was pasted into.

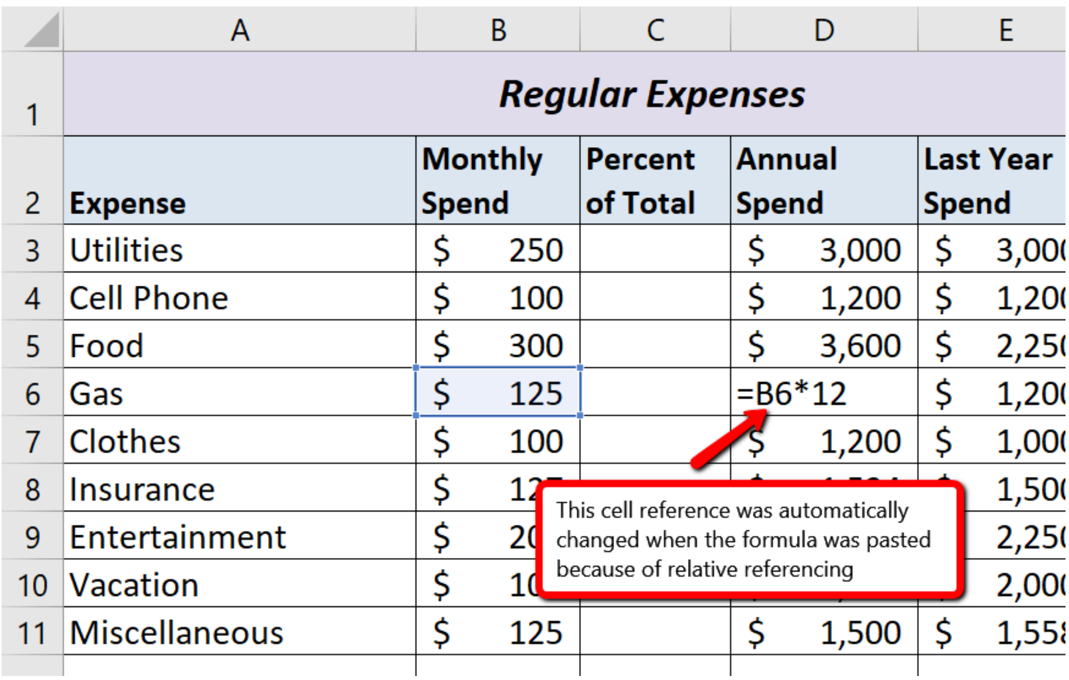


Figure 2.5 Relative Reference Example

Why?

Use Relative Referencing

Relative referencing is a convenient feature in Excel. When you use cell references in a formula, Excel automatically adjusts the cell references when the formula is pasted into new cell locations. If this feature were not available, you would have to manually retype the formula when you want the same calculation applied to other cell locations in a column or row.

Creating Complex Formulas (Controlling the Order of Operations)

The next formula to be added to the Personal Budget workbook is the percent change over last year (Column F). This formula determines the difference between this year's Annual Spend values (Column D) and the values in the Last Year Spend column (Column E) and shows the difference in terms of a percentage. This requires that the order of mathematical operations be controlled to get an accurate result.

Excel uses the standard mathematical order of operations, as defined in Table 2.3. When writing complex formulas it is important to remember this **order of operations.** You want to be sure that your formulas will calculate in the order you intend. To help you remember which operations will be performed first, you can use the acronym PEMDAS.

P – parentheses

E – exponents

MD – multiplication and division

AS – addition and subtraction

Table 2.3 shows the standard order of operations (PEMDAS) for a typical formula. To change the order of operations shown in the table, you can use parentheses to process certain mathematical calculations first.

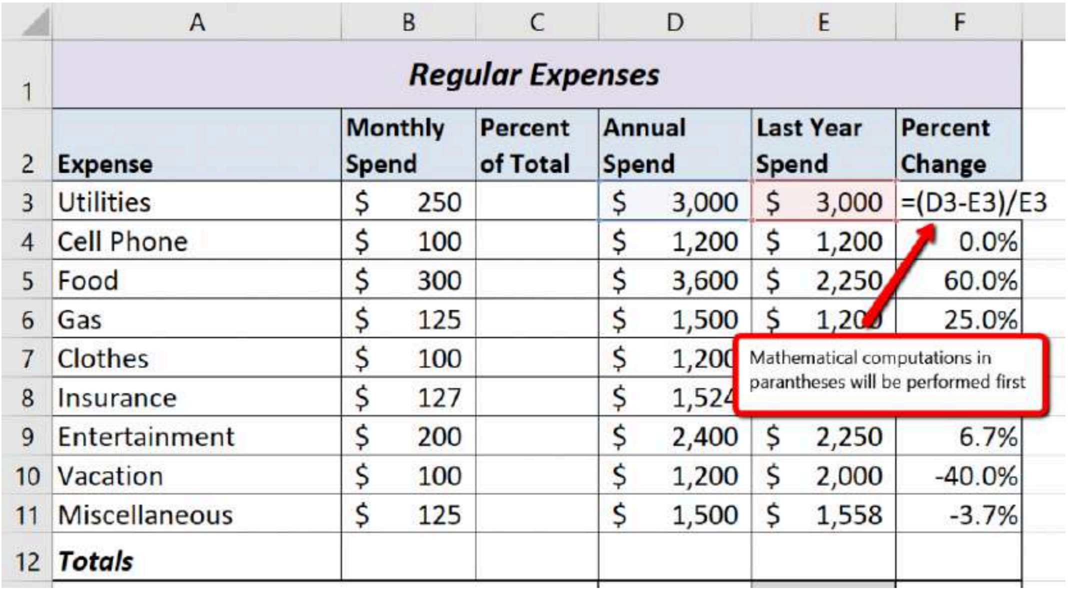
Table 2.3 Standard Order of Mathematical Operations (PEMDAS)

Symbol	Order
( )	Any calculation inside parentheses will be done first. If there are layers of parentheses used in a formula, Excel computes the innermost parentheses first and the outermost parentheses last.
^	Excel executes any exponential computations next.
* or /	Excel performs any multiplication or division computations next. When there are multiple instances of these computations in a formula, they are executed in order from left to right.
+ or –	Excel performs any addition or subtraction computations last. When there are multiple instances of these computations in a formula, they are executed in order from left to right.

To create the Percent Change formula, we will need to use parentheses to control the order of the calculations. We need the difference of the two values to be found before the division is done, so we will use parentheses around the subtraction portion of the formula to indicate that calculation needs to be done first. This formula is added to the worksheet as follows:

1. Click cell F3 in the **Budget Detail** worksheet.
2. Type an equal sign =.
3. Type an open parenthesis **(**.
4. Click cell D3. This will add a cell reference to cell D3 to the formula. *When building formulas, you can click cell locations instead of typing them.*
5. Type a minus sign –.
6. Click cell E3 to add this cell reference to the formula.
7. Type a closing parenthesis **)**.
8. Type the slash / symbol for division.
9. Click cell E3. This completes the formula that will calculate the percent change of last year's actual spent dollars vs. this year's budgeted spend dollars (**see Figure 2.6**).
10. Press the ENTER key.
11. Click cell F3 to activate it.
12. Place the mouse pointer over the Auto Fill Handle.
13. When the mouse pointer turns from a white block plus sign to a black plus sign, click and drag down to cell F11. This pastes the formula into the range F4:F11.

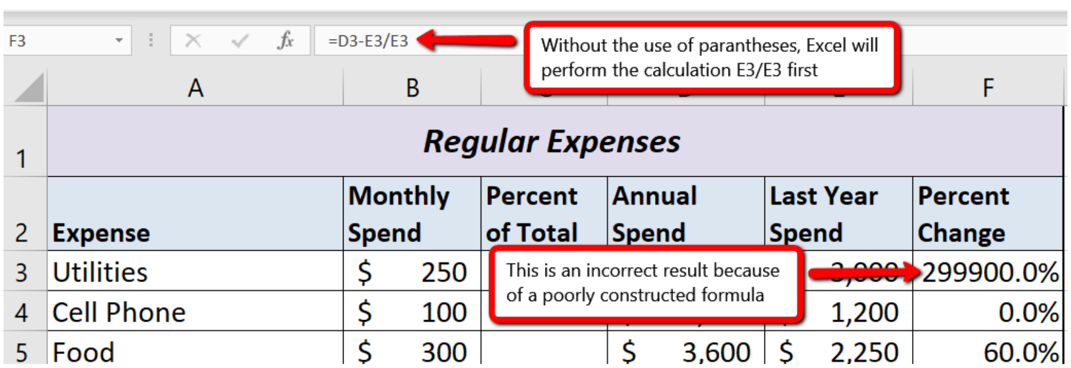
Figure 2.6 shows the formula that was added to the **Budget Detail** worksheet to calculate the percent change in spending. The parentheses were added to this formula to control the order of operations. Any mathematical computations placed in parentheses are executed first before the standard order of mathematical operations (see **Table 2.3)**. In this case, if parentheses were not used, Excel would produce an erroneous result for this worksheet.



	A	B	C	D	E	F
1	***Regular Expenses***					
2	**Expense**	**Monthly Spend**	**Percent of Total**	**Annual Spend**	**Last Year Spend**	**Percent Change**
3	Utilities	$ 250		$ 3,000	$ 3,000	=(D3-E3)/E3
4	Cell Phone	$ 100		$ 1,200	$ 1,200	0.0%
5	Food	$ 300		$ 3,600	$ 2,250	60.0%
6	Gas	$ 125		$ 1,500	$ 1,200	25.0%
7	Clothes	$ 100		$ 1,200		
8	Insurance	$ 127		$ 1,524		
9	Entertainment	$ 200		$ 2,400	$ 2,250	6.7%
10	Vacation	$ 100		$ 1,200	$ 2,000	-40.0%
11	Miscellaneous	$ 125		$ 1,500	$ 1,558	-3.7%
12	***Totals***					

Figure 2.6 Adding the Percent Change Formula

Figure 2.7 shows the result of the percent change formula if the parentheses are removed. The formula produces a result of a 299900% increase. Since there is no change between the LY spend and the budget Annual Spend, the result should be 0%. However, without the parentheses, Excel is following the standard order of operations. This means the value in cell E3 will be divided by E3 first (3,000/3,000), which is 1. Then, the value of 1 will be subtracted from the value in cell D3 (3,000–1), which is 2,999. Since cell F3 is formatted as a percentage, Excel expresses the output as an increase of 299900%.



	A	B	C	D	E	F
1	***Regular Expenses***					
2	**Expense**	**Monthly Spend**	**Percent of Total**	**Annual Spend**	**Last Year Spend**	**Percent Change**
3	Utilities	$ 250				299900.0%
4	Cell Phone	$ 100			1,200	0.0%
5	Food	$ 300		$ 3,600	$ 2,250	60.0%

Figure 2.7 Removing the Parentheses from the Percent Change Formula

Integrity Check<

Does the Output of Your Formula Make Sense?

It is important to note that the accuracy of the output produced by a formula depends on how it is constructed. Therefore, always check the result of your formula to see whether it makes sense with data in your worksheet. As shown in **Figure 2.7**, a poorly constructed formula can give you an inaccurate result. In other words, you can see that there is no change between the Annual Spend and LY Spend for Household Utilities. Therefore, the result of the formula should be 0%. However, since the parentheses were removed in this case, the formula is clearly producing an erroneous result.

Skill Refresher

Formulas

1. Type an equal sign **=**.
2. Click or type a cell location. If using constants, type a number.
3. Type a mathematical operator.
4. Click or type a cell location. If using constants, type a number.
5. Use parentheses where necessary to control the order of operations.
6. Press the ENTER key.

Auditing Formulas

Excel provides a few tools that you can use to review the formulas entered into a worksheet. For example, instead of showing the outputs for the formulas used in a worksheet, you can have Excel show the formula as it was entered in the cell locations. This is demonstrated as follows:

1. With the **Budget Detail** worksheet open, click the Formulas tab of the Ribbon.
2. Click the Show Formulas button in the Formula Auditing group of commands. This displays the formulas in the worksheet instead of showing the mathematical outputs.
3. Click the Show Formulas button again. The worksheet returns to showing the output of the formulas.

You can also toggle Show Formulas on and off using the keyboard. Hold down the CTRL key while pressing the ` key.

Figure 2.8 shows the **Budget Detail** worksheet after activating the Show Formulas command in the Formulas tab of the Ribbon. As shown in the figure, this command allows you to view and check all the formulas in a worksheet without having to click each cell individually. After activating this command, the column widths in your worksheet increase significantly. The column widths were adjusted for the worksheet shown in **Figure 2.8** so all columns can be seen. The column widths return to their previous width when the Show Formulas command is deactivated.

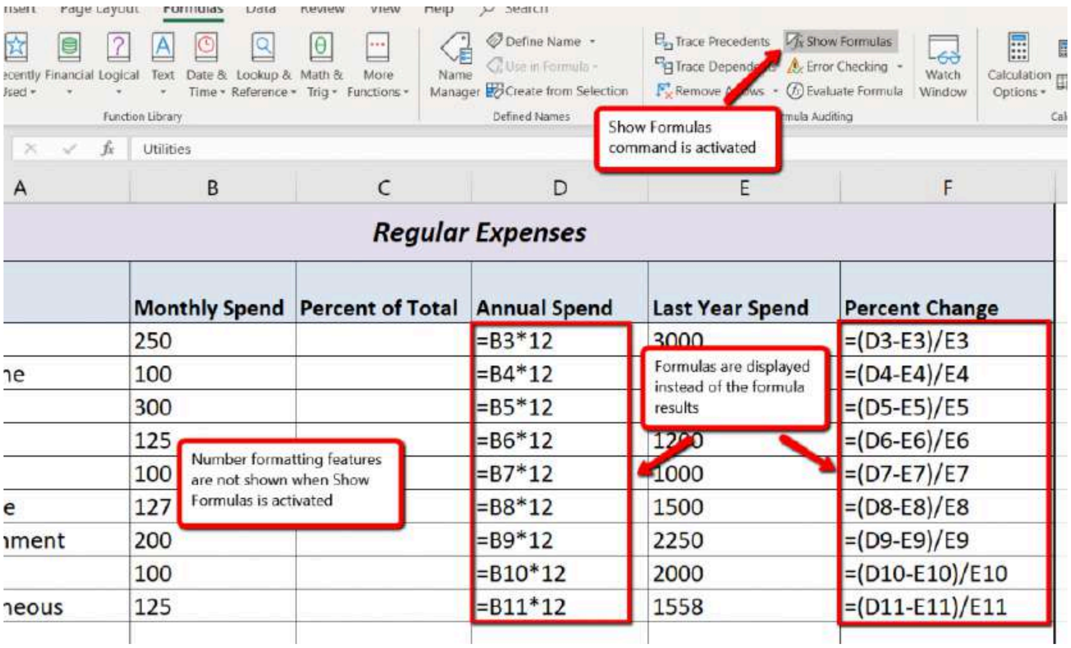


Figure 2.8 Show Formulas Command

Skill Refresher

Show Formulas

1. Click the Formulas tab on the Ribbon.
2. Click the Show Formulas button in the Formula Auditing group of commands.
3. Click the Show Formulas button again to show formula outputs.

Keyboard Shortcuts

Show Formulas

- Hold down the CTRL key while pressing the accent symbol **`**. Same for Excel for Mac.

Two other tools in the Formula Auditing group of commands are the **Trace Precedents** and **Trace Dependents** commands. These commands are used to trace the cell references used in a formula. A precedent cell is a cell whose value is used in other cells. The Trace Precedents command shows an arrow to indicate the cells or ranges (precedents) which affect the active cell's value. A dependent cell is a cell whose value depends on the values of other cells in the workbook. The Trace Dependents command shows where any given cell is referenced in a formula. The following is a demonstration of these commands:

1. Click cell D3.
2. Click the Trace Dependents button in the Formula Auditing group of commands in the Formulas tab of the Ribbon. A blue arrow appears, pointing to cell F3 (see **Figure 2.9**). This indicates that cell D3 is referenced in a formula entered in cell F3.
3. Click the Remove Arrows command in the Formula Auditing group of commands in the Formulas tab of the Ribbon. This removes the Trace Dependents arrow.
4. Click cell F3.
5. Click the Trace Precedents button in the Formula Auditing group of commands in the Formulas tab of the Ribbon. A blue arrow with dots in cells D3 and E3, and pointing to cell F3 appears (see **Figure 2.10**). This indicates that cells D3 and E3 are references in a formula entered in cell F3.
6. Click the Remove Arrows command in the Formula Auditing group of commands in the Formulas tab of the Ribbon. This removes the Trace Precedents arrow.
7. Save the **CH2 Personal Budget** file.

Figure 2.9 shows the Trace Dependents arrow on the **Budget Detail** worksheet. The blue dot represents the activated cell. The arrows indicate where the cell is referenced in formulas.

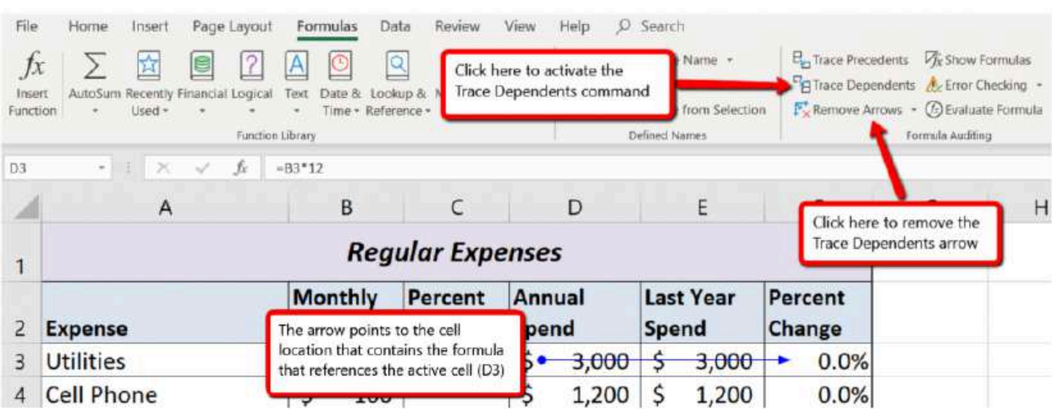


Figure 2.9 Trace Dependents Example

Figure 2.10 shows the Trace Precedents arrow on the **Budget Detail** worksheet. The blue dots on this arrow indicate the cells that are referenced in the formula contained in the activated cell. The arrow is pointing to the activated cell location that contains the formula.

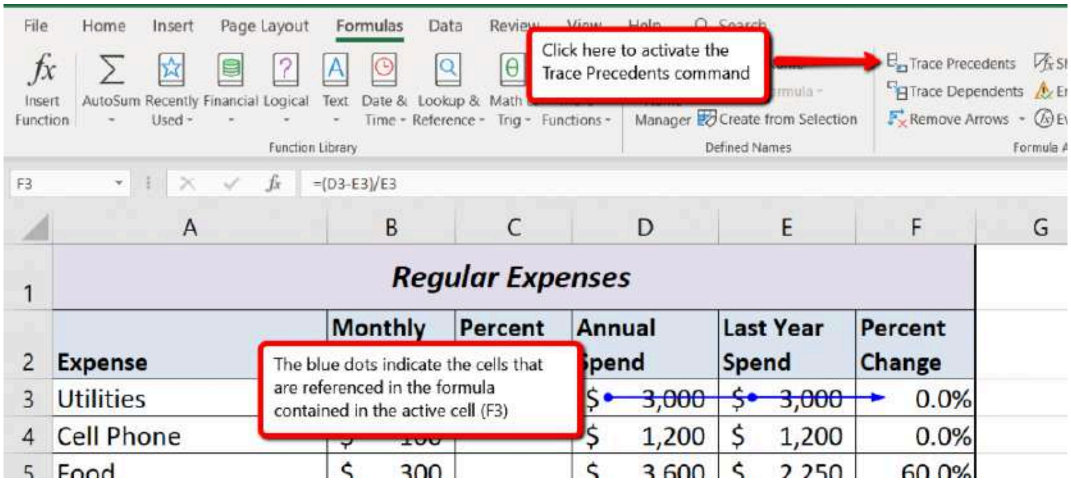


Figure 2.10 Trace Precedents Example

Skill Refresher

Trace Dependents

1. Click a cell location that contains a number or formula.
2. Click the Formulas tab on the Ribbon.
3. Click the Trace Dependents button in the Formula Auditing group of commands.
4. Use the arrow(s) to determine where the cell is referenced in formulas and functions.
5. Click the Remove Arrows button to remove the arrows from the worksheet.

Trace Precedents

1. Click a cell location that contains a formula or function.
2. Click the Formulas tab on the Ribbon.
3. Click the Trace Precedents button in the Formula Auditing group of commands.
4. Use the dot(s) along the line to determine what cells are referenced in the formula or function.
5. Click the Remove Arrows button to remove the line with the dots.

Key Takeaways

- Mathematical computations are conducted through formulas and functions.
- An equal sign **=** precedes all formulas and functions.
- Formulas and functions must be created with cell references to conduct what-if scenarios where mathematical outputs are recalculated when one or more inputs are changed.
- Mathematical operators on a typical calculator are different from those used in Excel. **Table 2.2** "Excel Mathematical Operators" lists Excel mathematical operators.
- When using numerical values in formulas and functions, only use universal constants that do not change, such as days in a week, months in a year, and so on.
- Relative referencing automatically adjusts the cell references in formulas and functions when they are pasted into new locations on a worksheet. This eliminates the need to retype formulas and functions when they are needed in multiple rows or columns on a worksheet.
- Parentheses must be used to control the order of operations when necessary for complex formulas.
- Formula auditing tools such as Trace Dependents, Trace Precedents, and Show Formulas should be used to check the integrity of formulas that have been entered into a worksheet.

Attribution

Adapted by Mary Schatz from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

4.XLSX.2 INTRODUCTORY STATISTICAL FUNCTIONS

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Learning Objectives

1. Use the SUM function to calculate totals.
2. Use the COUNT function to count cell locations with numerical values.
3. Use the AVERAGE function to calculate the arithmetic mean.
4. Use the MAX and MIN functions to find the highest and lowest values in a range of cells.
5. Learn how to copy and paste formulas without formats applied to a cell location.
6. Use absolute references to calculate percent of totals.
7. Learn how to set a multiple level sort sequence for data sets that have duplicate values or outputs.

In addition to formulas, another way to conduct mathematical computations in Excel is through functions. Excel functions apply a mathematical process to a group of cells in a worksheet. For example, the SUM function is used to add the values contained in a range of cells. Functions are more efficient than formulas when you are applying a mathematical process to a group of cells. If you use a formula to add the values in a range of cells, you would have to add each cell location to the formula one at a time. This can be very time-consuming if you have to add the values in a few hundred cell locations. However, when you use a function, you can highlight all the cells that contain values you wish to sum in just one step.

The components of a function are as follows:

=FunctionName(Arguments)

Functions are a type of formula, therefore they start with an equal sign. The next component is the name of the function. A list of commonly used functions is shown in **Table 2.4**. After the function name comes the arguments for the function, which are always enclosed in parentheses. The arguments are the cell locations and/or values that will be used in the function. The number and type of arguments varies based on the the function being used, although in this section we will only work with a range of cells for the function arguments. Some examples of different functions with their arguments are:

=SUM(B2:B15) – adds the values in B2 through B15
=SQRT(A5) – finds the square root of the value in A5
=COUNTA(A1:A20) – finds the number of cells from A1 through A20 that contain text or a number

Throughout Section 2.2 we will add a variety of mathematical functions to the Personal Budget workbook. In addition to creating functions, this section also reviews percent of total calculations and the use of absolute references.

Table 2.4 Commonly Used Functions

Function	Output
ABS	The absolute value of a number
AVERAGE	The average or arithmetic mean for a group of numbers
COUNT	The number of cell locations in a range that contain a numeric value
COUNTA	The number of cell locations in a range that contain text or a numeric value
MAX	The highest numeric value in a group of numbers
MEDIAN	The middle number in a group of numbers (half the numbers in the group are higher than the median and half the numbers in the group are lower than the median)
MIN	The lowest numeric value in a group of numbers
MODE	The number that appears most frequently in a group of numbers
PRODUCT	The result of multiplying all the values in a range of cell locations
SQRT	The positive square root of a number
SUM	The total of all numeric values in a group

It is important to note that there are several methods for adding a function to a worksheet, and we will explore each of them throughout this section.

- Typing the function directly into a cell
- Selecting from the function list
- Using the Function Library on the ribbon
- Using the Insert Function button

The SUM Function

The SUM function is used when you need to calculate totals for a range of cells or a group of selected cells on a worksheet. With regard to the **Budget Detail** worksheet, we will use the SUM function to calculate the totals in row 12, starting with the Monthly Spend total in B12. The following illustrates how a function can be added to a worksheet by typing it into a cell location:

1. Switch to the **Budget Detail** worksheet if needed.
2. Click cell B12.
3. Type an equal sign =.
4. Type the function name **SUM**.
5. Type an open parenthesis **(**.
6. Click cell B3 and drag down to cell B11. This places the range B3:B11 into the function.
7. Type a closing parenthesis **)**.
8. Press the ENTER key. The function calculates the total for the Monthly Spend column, which is $1,427.

Figure 2.11 shows the appearance of the SUM function added to the **Budget Detail** worksheet before pressing the ENTER key.

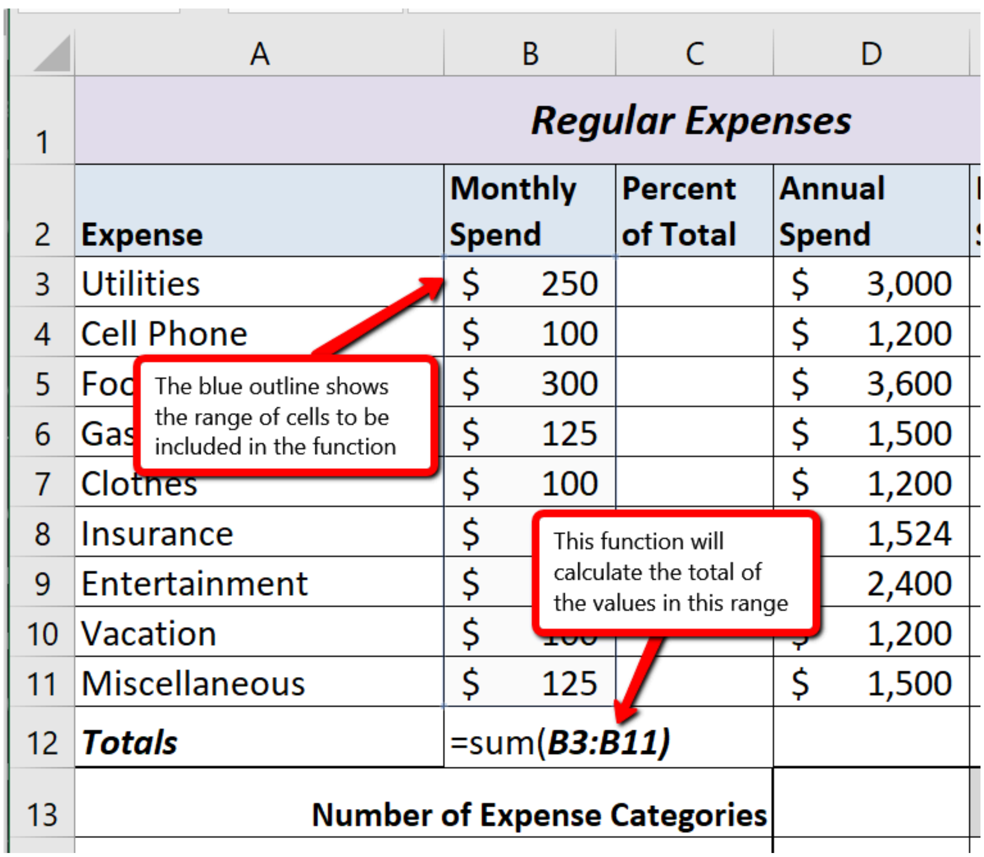
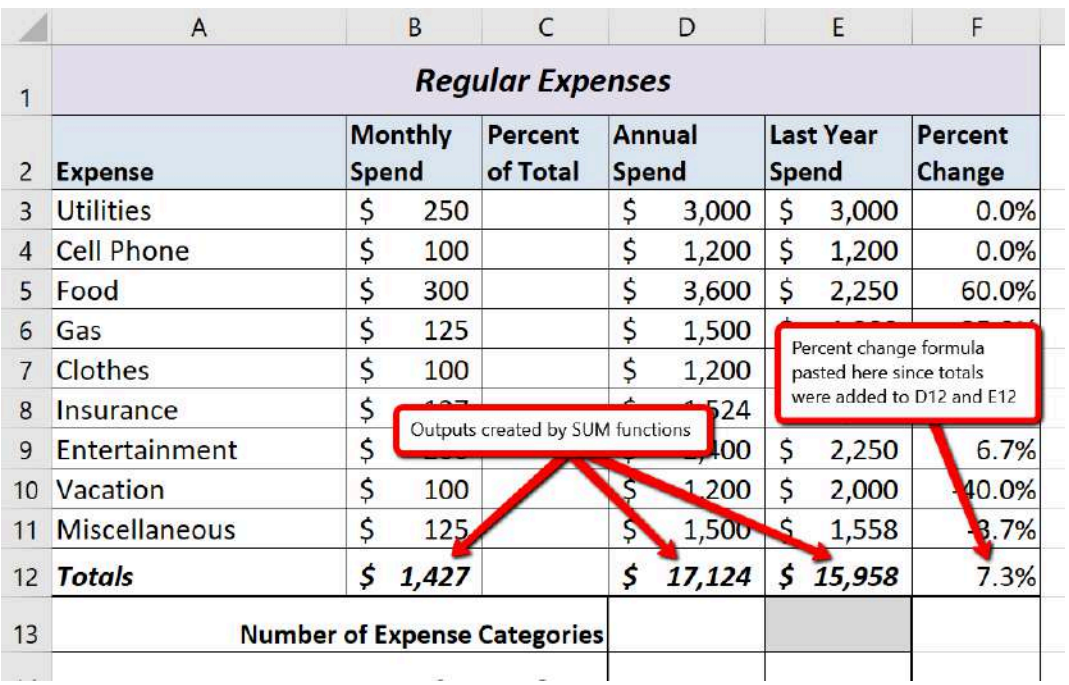


Figure 2.11 Adding the SUM Function to the Budget Detail Worksheet

As shown in **Figure 2.11**, the SUM function was added to cell B12. However, this function is also needed to calculate the totals in the Annual Spend and Last Year Spend columns. The function can be copied and pasted into these cell locations because of relative referencing. Relative referencing serves the same purpose for functions as it does for formulas. To complete the Totals in row 12, we need to copy and paste the SUM function into D12 and E12. Since we will then have totals in D12 and E12, we can paste the percent change formula into F12.

1. Click cell B12 in the **Budget Detail** worksheet.
2. Click the Copy button in the Home tab of the Ribbon.
3. Highlight cells D12 and E12.
4. Click the Paste button in the Home tab of the Ribbon. This pastes the SUM function into cells D12 and E12 and calculates the totals for these columns.
5. Click cell F11.
6. Click the Copy button in the Home tab of the Ribbon.
7. Click cell F12, then click the Paste button in the Home tab of the Ribbon.

Figure 2.12 shows the output of the SUM function that was added to cells B12, D12, and E12. In addition, the percent change formula was copied and pasted into cell F12. Notice that this version of the budget is planning an increase in spending compared to last year.



	A	B	C	D	E	F
1	***Regular Expenses***					
2	**Expense**	**Monthly Spend**	**Percent of Total**	**Annual Spend**	**Last Year Spend**	**Percent Change**
3	Utilities	$ 250		$ 3,000	$ 3,000	0.0%
4	Cell Phone	$ 100		$ 1,200	$ 1,200	0.0%
5	Food	$ 300		$ 3,600	$ 2,250	60.0%
6	Gas	$ 125		$ 1,500	[illegible]	[illegible]
7	Clothes	$ 100		$ 1,200	[illegible]	[illegible]
8	Insurance	$ [illegible]		[illegible]24	[illegible]	[illegible]
9	Entertainment	$ [illegible]		[illegible]400	$ 2,250	6.7%
10	Vacation	$ 100		$ 1,200	$ 2,000	40.0%
11	Miscellaneous	$ 125		$ 1,500	$ 1,558	3.7%
12	***Totals***	***$ 1,427***		***$ 17,124***	***$ 15,958***	7.3%
13	**Number of Expense Categories**					

Figure 2.12 Results of the SUM Function in the Budget Detail worksheet

Integrity Check

Cell Ranges in Functions

When you intend to use a function on a range of cells in a worksheet, make sure there are two cell locations separated by a **colon** and not a comma. If you enter two cell locations separated by a comma, the function will calculate only the two cell locations listed instead of an entire range of cells. For example, the SUM function shown in **Figure 2.13** will add only the values in cells C3 and C11, not the range C3:C11.

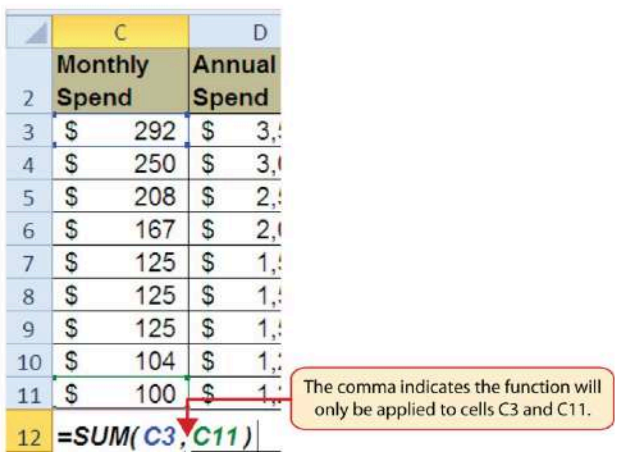


Figure 2.13 SUM Function Adding Two Cell Locations

The COUNT Function

Data file: Continue with CH2 Personal Budget.

The next function that we will add to the **Budget Detail** worksheet is the COUNT function. The COUNT function is used to determine how many cells in a range contain a numeric entry. The COUNT function will not work for counting text or other non-numeric entries. If you want to count text instead of, or in addition to, numeric entries you use the COUNTA function. For the **Budget Detail** worksheet, we will use the COUNT function to count the number of items that are planned in the Annual Spend column (Column D). The following explains how the COUNT function is added to the worksheet by selecting from the function list:

1. Click cell D13.
2. Type an equal sign =.

3. Type the letter **C** (to start spelling the name of the function).
4. Click the down arrow on the scroll bar of the function list (see **Figure 2.14**) and find the word COUNT.

   Mac Users can scroll down with touchpad or mouse to find COUNT
5. Double click the word COUNT from the function list.

   Mac Users should single click the word "COUNT" do not double-click
6. Highlight the range D3:D11.
7. You can type a closing parenthesis **)** and then press the ENTER key, or simply press the ENTER key and Excel will close the function for you. The function produces an output of 9 since there are 9 items planned on the worksheet.

Figure 2.14 shows the function list box that appears after completing steps 2 and 3 for the COUNT function. The function list provides an alternative method for adding a function to a worksheet.

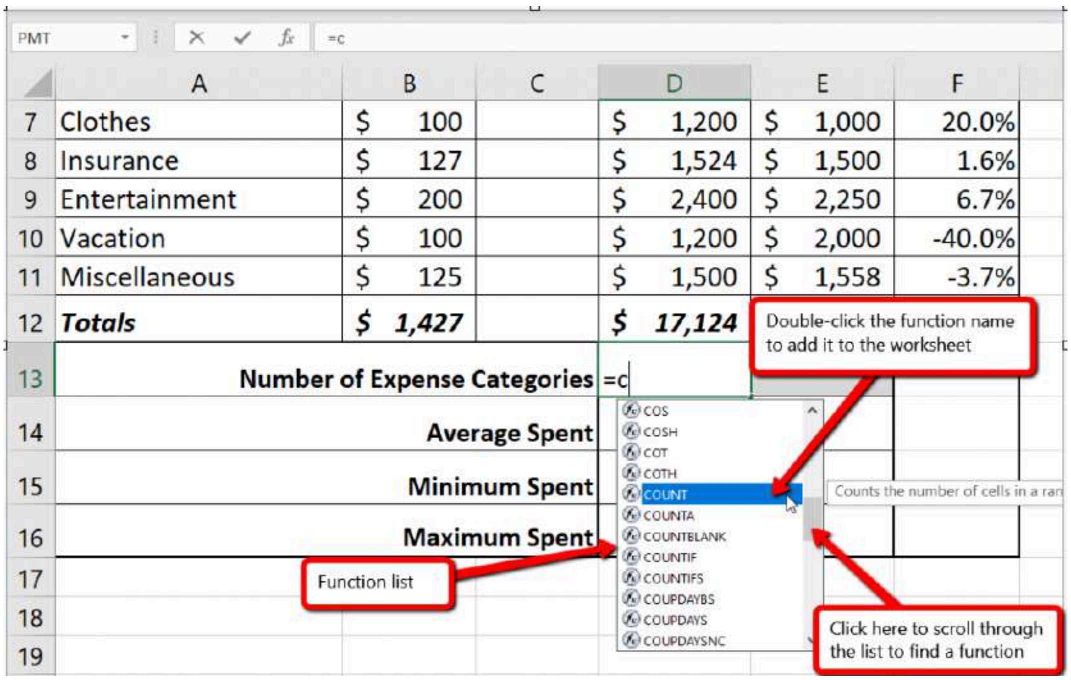


Figure 2.14 Using the Function List to Add the COUNT Function

Figure 2.15 shows the output of the COUNT function after pressing the ENTER key. The function counts the number of cells in the range D3:D11 that contain a numeric value. The result of 9 indicates that there are 9 categories planned for this budget.

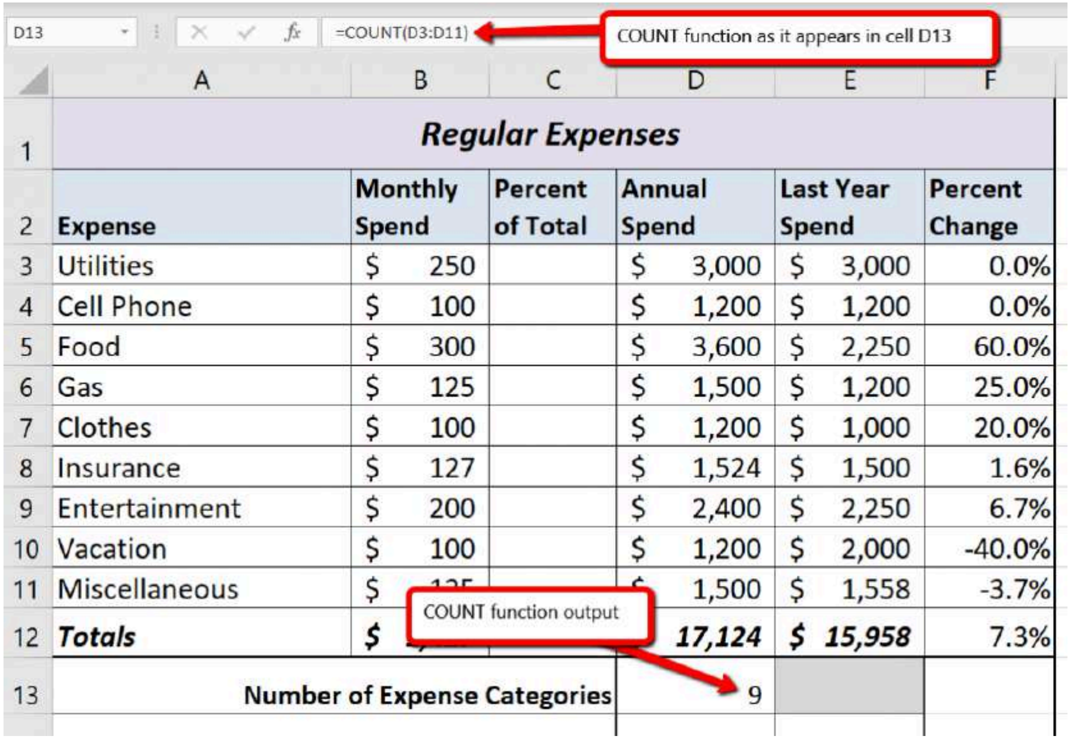


Figure 2.15 Completed COUNT Function in the Budget Detail Worksheet

The AVERAGE Function

The next function we will add to the **Budget Detail** worksheet is the AVERAGE function. This function is used to calculate the arithmetic mean for a group of numbers. For the **Budget Detail** worksheet, we will use the function to calculate the average of the values in the Annual Spend column. We will add this to the worksheet by using the Function Library on the Formulas ribbon. The following steps explain how this is accomplished:

1. Click cell D14 in the **Budget Detail** worksheet.
2. Click the Formulas tab on the Ribbon.
3. Click the More Functions button in the Function Library group of commands.
4. Place the mouse pointer over the Statistical option from the drop-down list of options.
5. Click the AVERAGE function name from the list of functions that appear in the menu (see **Figure 2.16**). This opens the Function Arguments dialog box.
6. Click the Collapse Dialog button in the Function Arguments dialog box (see **Figure 2.17**).
   For Mac Users, the Collapse Dialog button may not collapse. Just continue with Step 7 and press Enter after selecting the range.
7. Highlight the range D3:D11.
8. Click the Expand Dialog button in the Function Arguments dialog box (see **Figure 2.18**). You can also press the ENTER key to get the same result.

9. Click the OK button on the Function Arguments dialog box. This adds the AVERAGE function to the worksheet.
   Mac Users should click the **DONE** button

Figure 2.16 illustrates how a function is selected from the Function Library in the Formulas tab of the Ribbon.

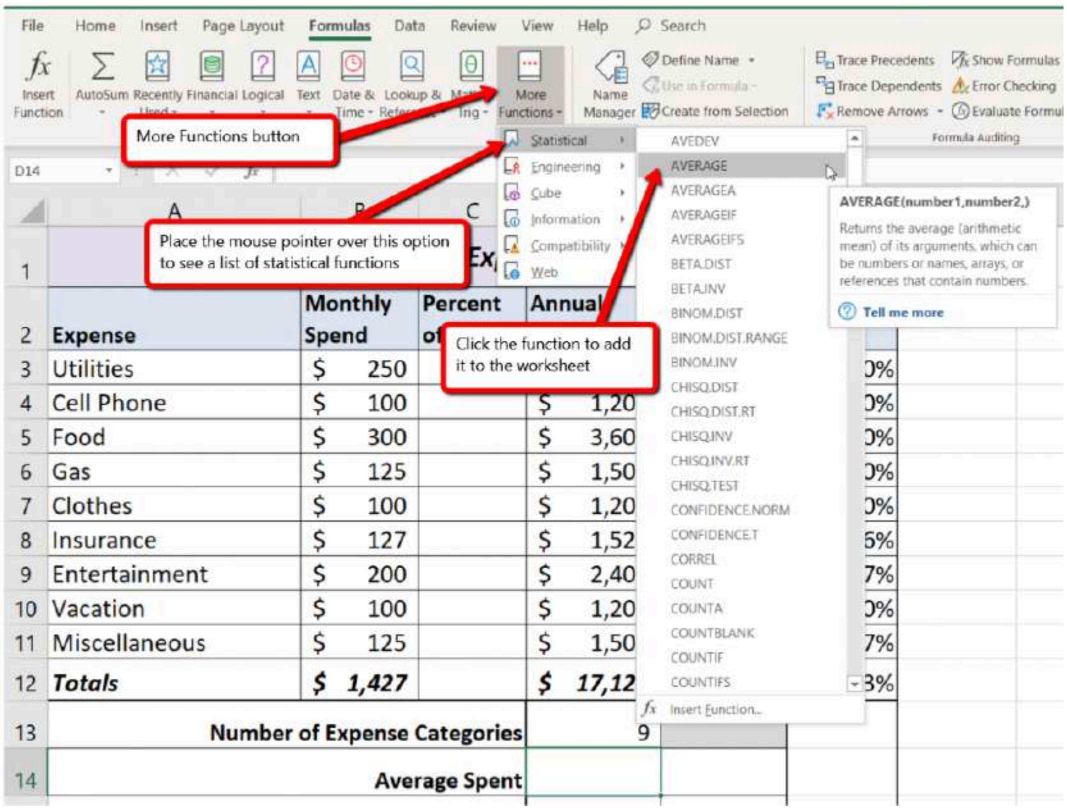


Figure 2.16 Selecting the AVERAGE function from the Function Library

Figure 2.17 shows the Function Arguments dialog box. This appears after a function is selected from the Function Library. The Collapse Dialog button is used to hide the dialog box so a range of cells can be highlighted on the worksheet and then added to the function.

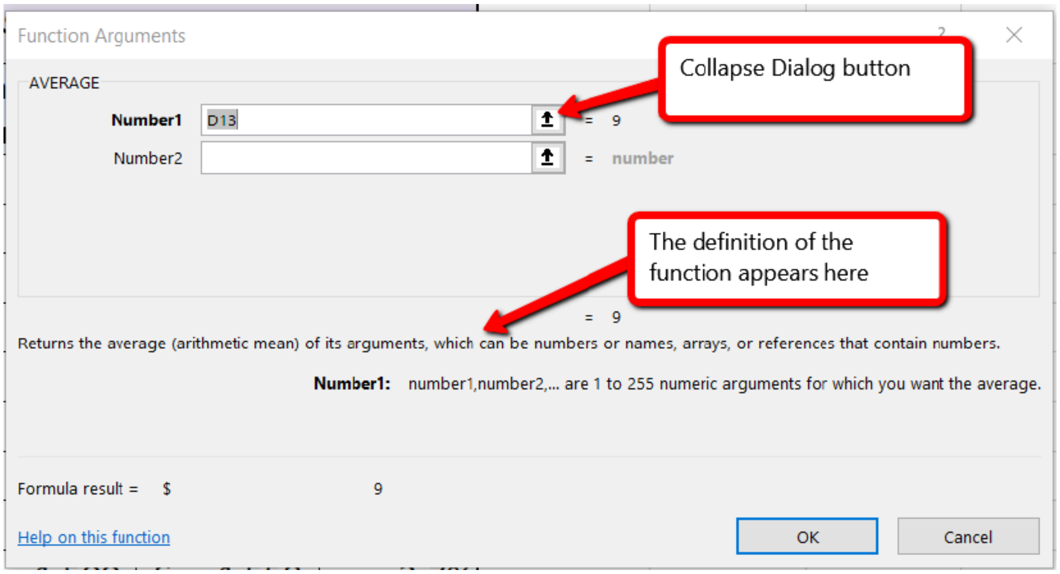


Figure 2.17 Function Arguments Dialog Box

Figure 2.18 shows how a range of cells can be selected from the Function Arguments dialog box once it has been collapsed.

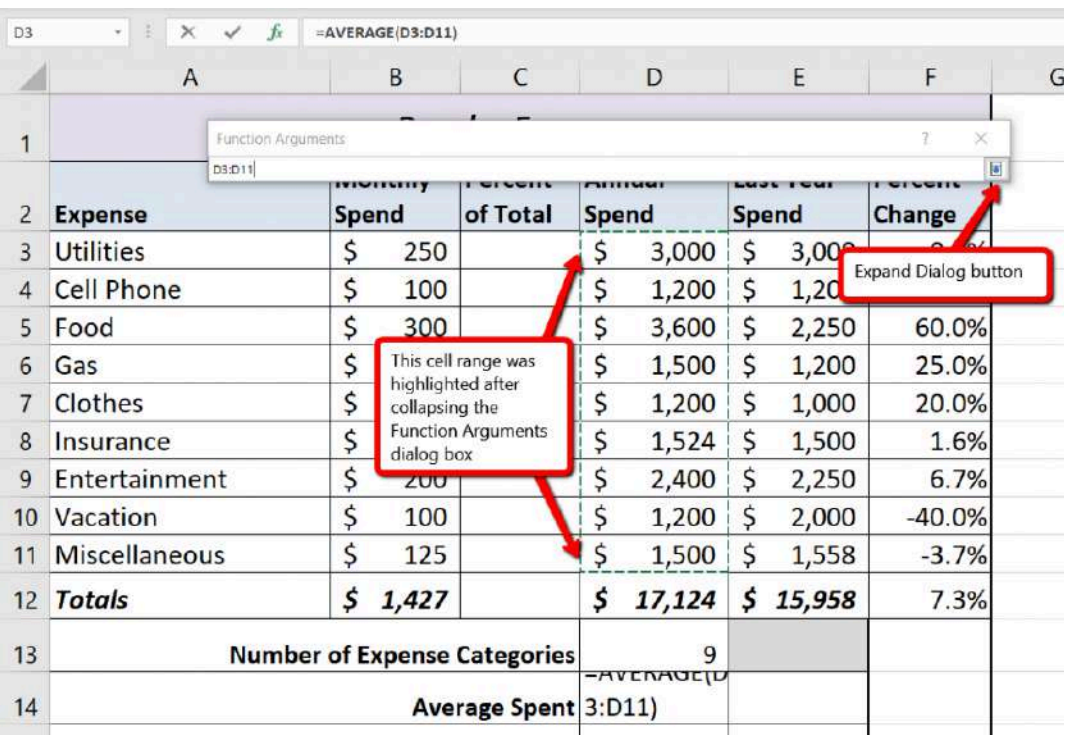


Figure 2.18 Selecting a range from the Function Arguments Dialog Box

Figure 2.19 shows the Function Arguments dialog box after the cell range is defined for the AVERAGE function. The dialog box shows the result of the function before it is added to the cell location. This allows you to assess the function output to determine whether it makes sense before adding it to the worksheet.

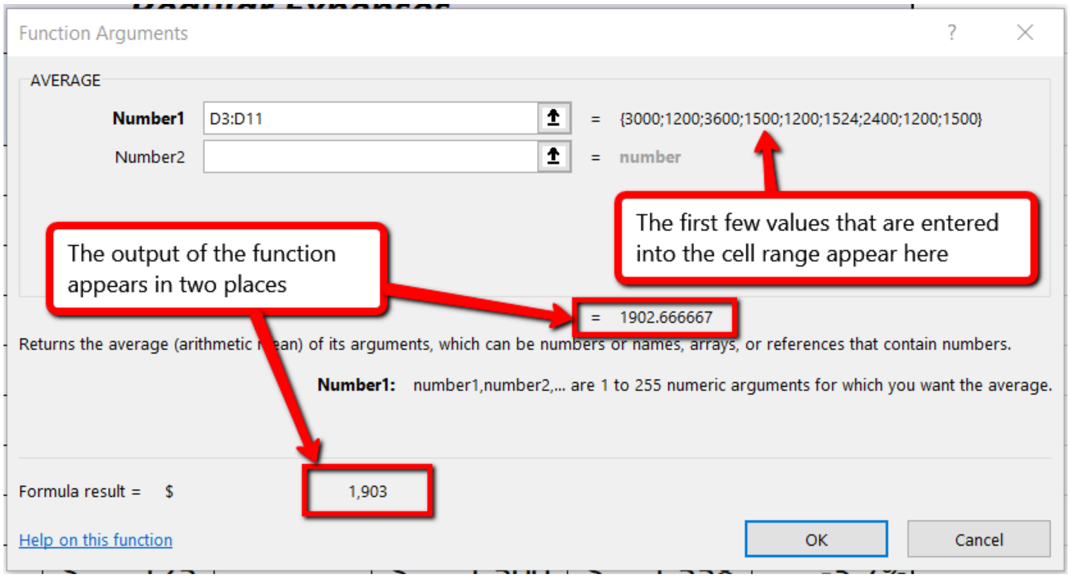
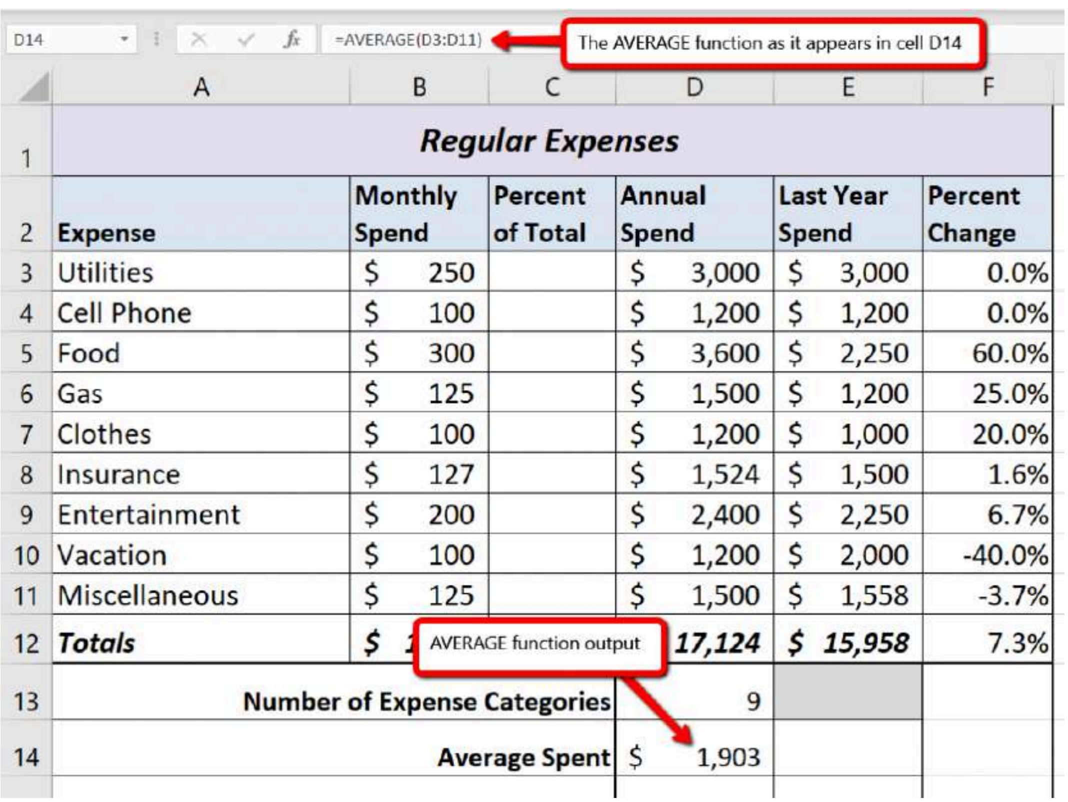


Figure 2.19 Function Arguments Dialog Box after a Cell Range Is Defined for a Function

Figure 2.20 shows the completed AVERAGE function in the **Budget Detail** worksheet. The output of the function shows that on average we expect to spend $1,903 for each of the categories listed in Column A of the budget. This average spend calculation per category can be used as an indicator to determine which categories are costing more or less than the average budgeted spend dollars.



	A	B	C	D	E	F
1	*Regular Expenses*					
2	**Expense**	**Monthly Spend**	**Percent of Total**	**Annual Spend**	**Last Year Spend**	**Percent Change**
3	Utilities	$ 250		$ 3,000	$ 3,000	0.0%
4	Cell Phone	$ 100		$ 1,200	$ 1,200	0.0%
5	Food	$ 300		$ 3,600	$ 2,250	60.0%
6	Gas	$ 125		$ 1,500	$ 1,200	25.0%
7	Clothes	$ 100		$ 1,200	$ 1,000	20.0%
8	Insurance	$ 127		$ 1,524	$ 1,500	1.6%
9	Entertainment	$ 200		$ 2,400	$ 2,250	6.7%
10	Vacation	$ 100		$ 1,200	$ 2,000	-40.0%
11	Miscellaneous	$ 125		$ 1,500	$ 1,558	-3.7%
12	***Totals***	***$ 1***		***17,124***	***$ 15,958***	7.3%
13	**Number of Expense Categories**			9		
14	**Average Spent**			$ 1,903		

Figure 2.20 Completed AVERAGE function

The MAX and MIN Functions

Data file: Continue with CH2 Personal Budget.

The final two statistical functions that we will add to the **Budget Detail** worksheet are the MAX and MIN functions. These functions identify the highest and lowest values in a range of cells. The following steps explain how to add these functions to the **Budget Detail** worksheet using the Insert Function button:

1. Click cell D15 in the **Budget Det**ail worksheet.
2. Click the Insert Function button on the Formulas ribbon. (see **Figure 2.21**)
3. This brings up the Insert Function dialog box. Type the word **MIN** in the search box and then click the Go button. (see **Figure 2.22**)
4. Double-click MIN in the list. This opens the Function Arguments dialog box.
5. Click the Collapse Dialog button in the Function Arguments dialog box.
6. Highlight the range D3:D11.
7. Click the Expand Dialog button in the Function Arguments dialog box.
8. Click the OK button on the Function Arguments dialog box. This adds the MIN function to the worksheet. (see **Figure 2.23)**
9. Click cell D16.
10. Repeat steps 2-8 (using MAX instead of MIN) to add the MAX function to the worksheet. (see **Figure 2.24**)

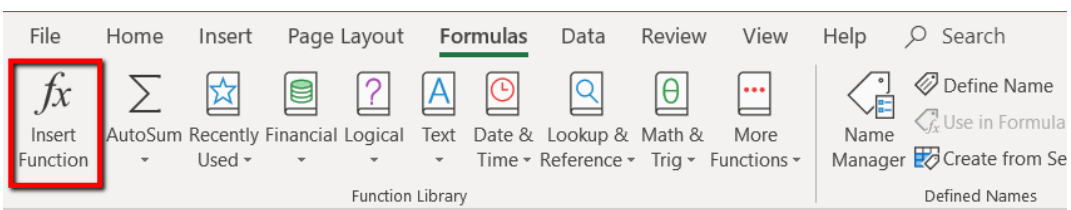


Figure 2.21 Insert Function button on the Formulas Ribbon

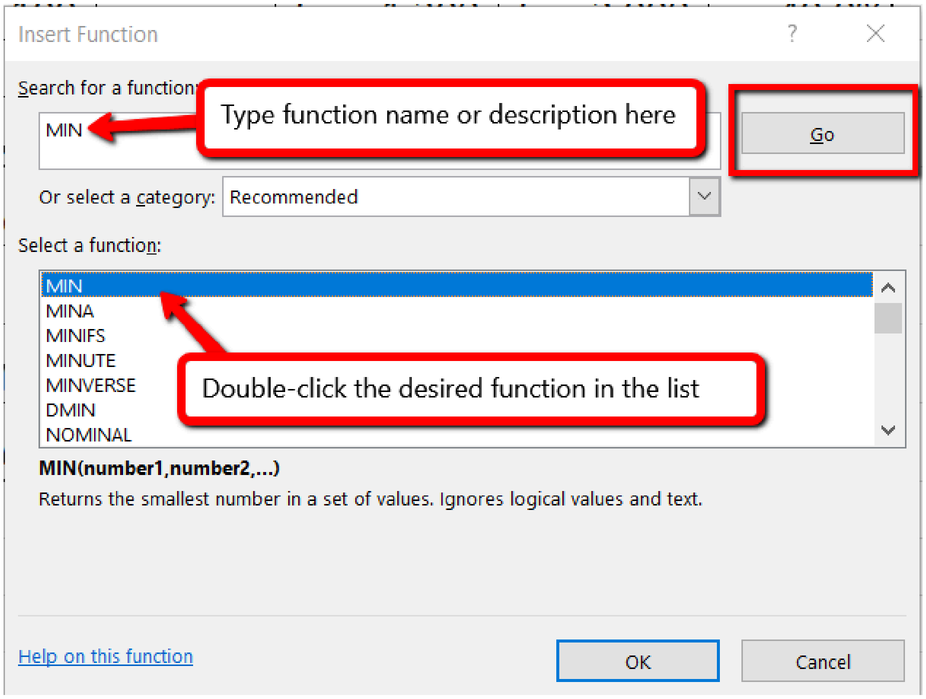
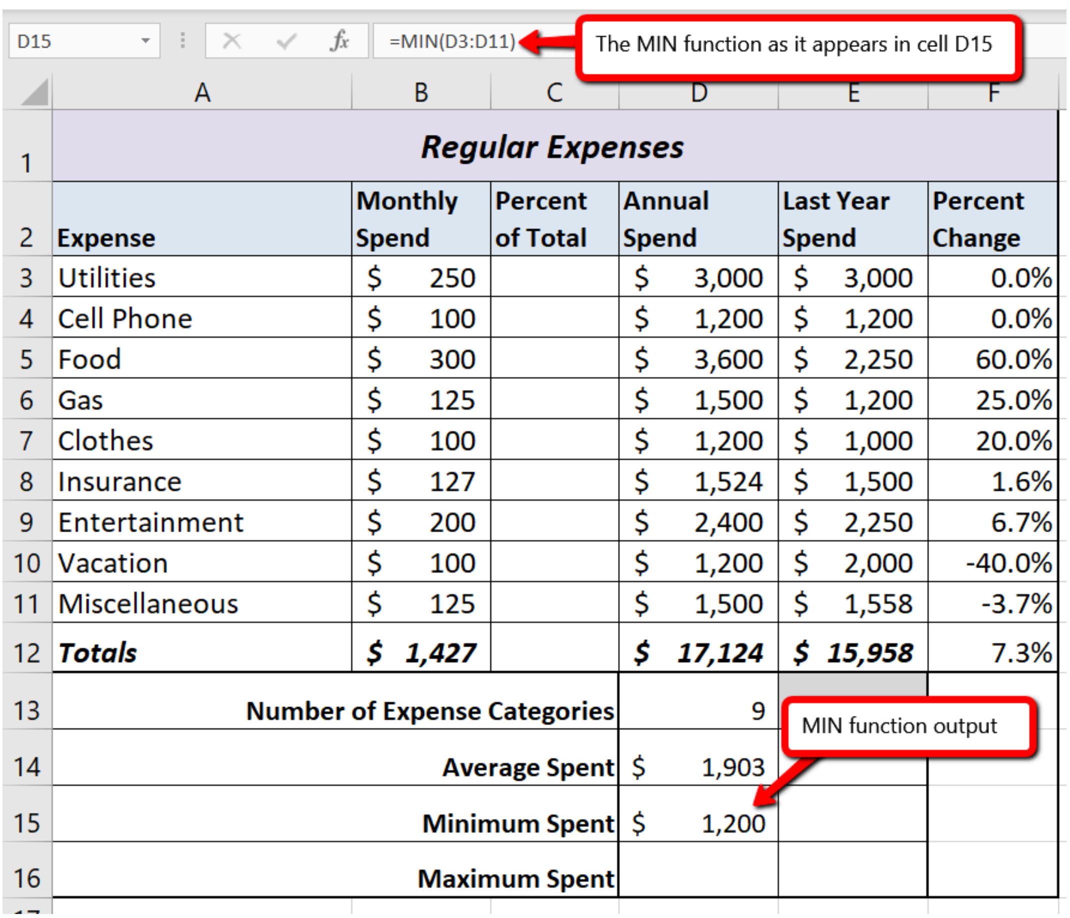
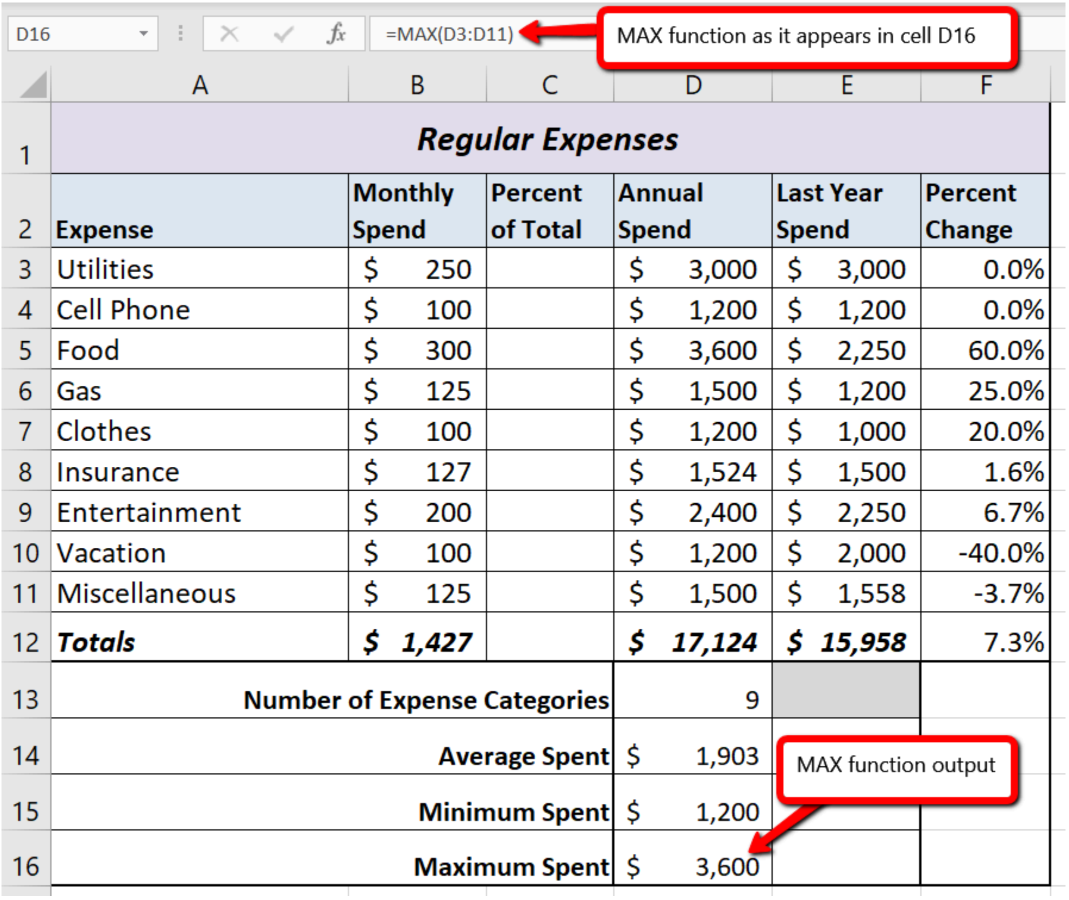


Figure 2.22 Insert Function Dialog Box



	A	B	C	D	E	F
1	***Regular Expenses***					
2	**Expense**	**Monthly Spend**	**Percent of Total**	**Annual Spend**	**Last Year Spend**	**Percent Change**
3	Utilities	$ 250		$ 3,000	$ 3,000	0.0%
4	Cell Phone	$ 100		$ 1,200	$ 1,200	0.0%
5	Food	$ 300		$ 3,600	$ 2,250	60.0%
6	Gas	$ 125		$ 1,500	$ 1,200	25.0%
7	Clothes	$ 100		$ 1,200	$ 1,000	20.0%
8	Insurance	$ 127		$ 1,524	$ 1,500	1.6%
9	Entertainment	$ 200		$ 2,400	$ 2,250	6.7%
10	Vacation	$ 100		$ 1,200	$ 2,000	-40.0%
11	Miscellaneous	$ 125		$ 1,500	$ 1,558	-3.7%
12	***Totals***	***$ 1,427***		***$ 17,124***	***$ 15,958***	7.3%
13	**Number of Expense Categories**			9		
14	**Average Spent**			$ 1,903		
15	**Minimum Spent**			$ 1,200		
16	**Maximum Spent**					

2.23 MIN function added to the Budget Detail worksheet



	A	B	C	D	E	F
1	***Regular Expenses***					
2	**Expense**	**Monthly Spend**	**Percent of Total**	**Annual Spend**	**Last Year Spend**	**Percent Change**
3	Utilities	$ 250		$ 3,000	$ 3,000	0.0%
4	Cell Phone	$ 100		$ 1,200	$ 1,200	0.0%
5	Food	$ 300		$ 3,600	$ 2,250	60.0%
6	Gas	$ 125		$ 1,500	$ 1,200	25.0%
7	Clothes	$ 100		$ 1,200	$ 1,000	20.0%
8	Insurance	$ 127		$ 1,524	$ 1,500	1.6%
9	Entertainment	$ 200		$ 2,400	$ 2,250	6.7%
10	Vacation	$ 100		$ 1,200	$ 2,000	-40.0%
11	Miscellaneous	$ 125		$ 1,500	$ 1,558	-3.7%
12	***Totals***	***$ 1,427***		***$ 17,124***	***$ 15,958***	7.3%
13	**Number of Expense Categories**			9		
14	**Average Spent**			$ 1,903		
15	**Minimum Spent**			$ 1,200		
16	**Maximum Spent**			$ 3,600		

Figure 2.24 MAX function added to the Budget Detail worksheet

Skill Refresher

Typing a function or selecting from the function list

1. Type an equal sign **=**.
2. Type the function name followed by an open parenthesis **(** or double click the function name from the function list.
3. Highlight the range of cells to use or click individual cell locations followed by commas.
4. Type a closing parenthesis **)** and press the ENTER key or press the ENTER key to close the function.

Inserting a function using the ribbon

1. On the Formulas ribbon, select the correct category in the Function Library. Click the desired function in the list.
2. In the Function Dialog box, click the Collapse Dialog button and highlight the range of cells to

use.

3. Click the Expand Dialog button and then click the OK button in the Function Arguments dialog box.

Inserting (and searching for) a function using the Insert Function button

1. On the Formulas ribbon, click the Insert Function button and search for the function to use. Double-click on the desired function in the list.
2. In the Function Dialog box, click the Collapse Dialog button and highlight the range of cells to use.
3. Click the Expand Dialog button and then click the OK button in the Function Arguments dialog box.

Copy and Paste Formulas (Pasting without Formats)

Data file: Continue with CH2 Personal Budget.

As shown in **Figure 2.24**, the COUNT, AVERAGE, MIN, and MAX functions are summarizing the data in the Annual Spend column. You will also notice that there is space to copy and paste these functions under the Last Year Spend column. This allows us to compare what we spent last year and what we are planning to spend this year. Normally, we would simply copy and paste these functions into the range E14:E16. However, you may have noticed the thicker style border that was used around the perimeter of the range D13:E16. If we used the regular Paste command, the thick line on the right side of the range D13:E16 would be replaced with a single line. Therefore, we are going to use one of the Paste Special commands to paste only the functions without any of the formatting treatments. This is accomplished through the following steps:

1. Highlight the range D14:D16 in the **Budget Detail** worksheet.
2. Click the Copy button in the Home tab of the Ribbon.
3. Click cell E14.
4. Click the down arrow below the Paste button in the Home tab of the Ribbon.
5. Click the Formulas option from the drop-down list of buttons (see **Figure 2.25**).

Figure 2.25 shows the list of buttons that appear when you click the down arrow below the Paste button in the Home tab of the Ribbon. One thing to note about these options is that you can preview them before you make a selection by dragging the mouse pointer over the options. When the mouse pointer is placed over the Formulas button, you can see how the functions will appear before making a selection. Notice that the thick line border does not change when this option is previewed. That is why this selection is made instead of the regular Paste option.

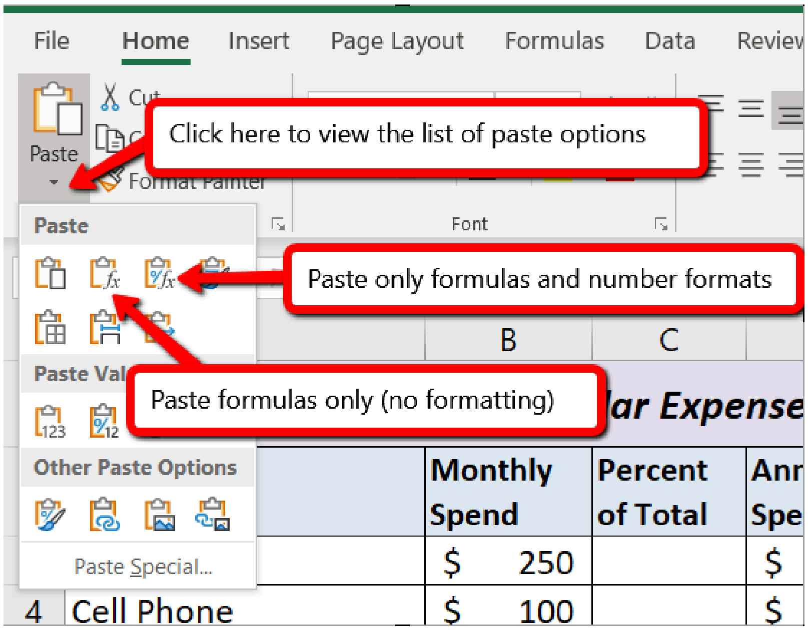


Figure 2.25 Paste Formulas Option

Skill Refresher

Paste Formulas without formatting

1. Click a cell location containing a formula or function.
2. Click the Copy button in the Home tab of the Ribbon.
3. Click the cell location or cell range where the formula or function will be pasted.
4. Click the down arrow below the Paste button in the Home tab of the Ribbon.
5. Click the Formulas button under the Paste group of buttons.

Absolute References (Calculating Percent of Totals)

Data file: Continue with CH2 Personal Budget.

To further analyze your budget, you want to see what percentage of your total monthly spending is spent in each category. Since totals were added to row 12 of the **Budget Detail** worksheet, a percent of total calculation can be added to Column C beginning in cell C3. The percent of total calculation shows the percentage for each value in the Monthly Spend column with respect to the total in cell B12. However, after the formula is created, it will be necessary to turn off Excel's relative referencing feature before copying and pasting the formula to the rest of the cell locations in the column. Turning off Excel's relative referencing feature is accomplished through an absolute reference.

First we will create the formula, which needs needs to divide the amount in B3 by the total monthly spend in B12.

1. Click cell C3 in the **Budget Detail** worksheet.
2. Type an equal sign =.
3. Click cell B3.
4. Type a forward slash /.
5. Click cell B12.
6. Press the ENTER key. You will see that Utilities represent about 17.5% of the Monthly Spend budget (see **Figure 2.26**).

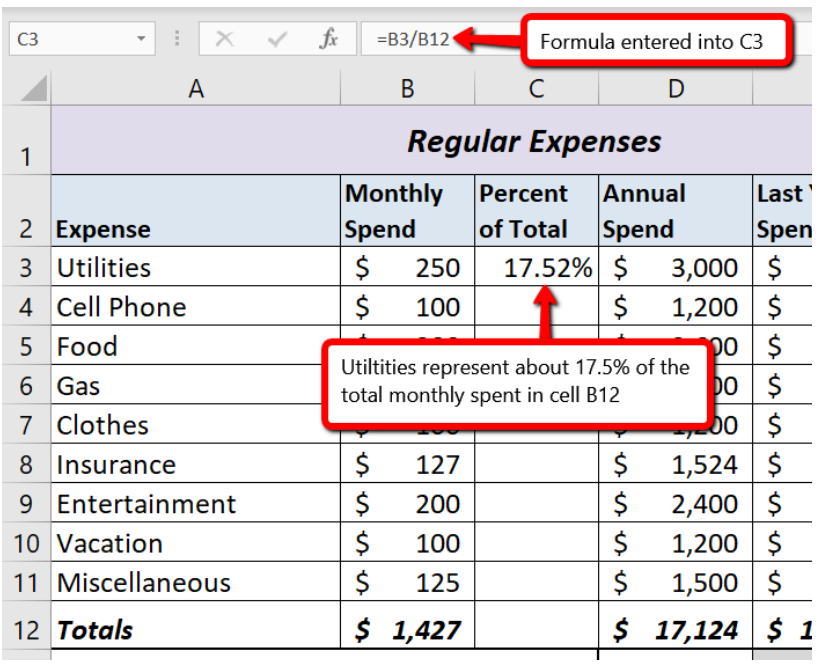


Figure 2.26 Adding a Formula to Calculate the Percent of Total

Figure 2.26 shows the completed formula that is calculating the percentage that Utilities represents to the total Monthly Spend for the budget (see cell C3). Normally, we would copy this formula and paste it into the range C4:C11. However, because of relative referencing, both cell references will increase by one row as the formula is pasted into the cells below C3. This is fine for the first cell reference in the formula (C3) but not for the second cell reference (C12).

Figure 2.27 illustrates what happens if we paste the formula into the range C4:C12 in its current state. Notice that Excel produces the **#DIV/0** error code. This means that Excel is trying to divide a number by zero, which is impossible. Looking at the formula in cell C4, you see that the first cell reference was changed from B3 to B4. This is fine because we now want to divide the Monthly Spend for Cell Phone (cell B4) by the total Monthly Spend in cell B12. However, Excel has also changed the B12 cell reference to B13. Because cell location B13 does not contain a number, the formula produces the #DIV/0 error code.

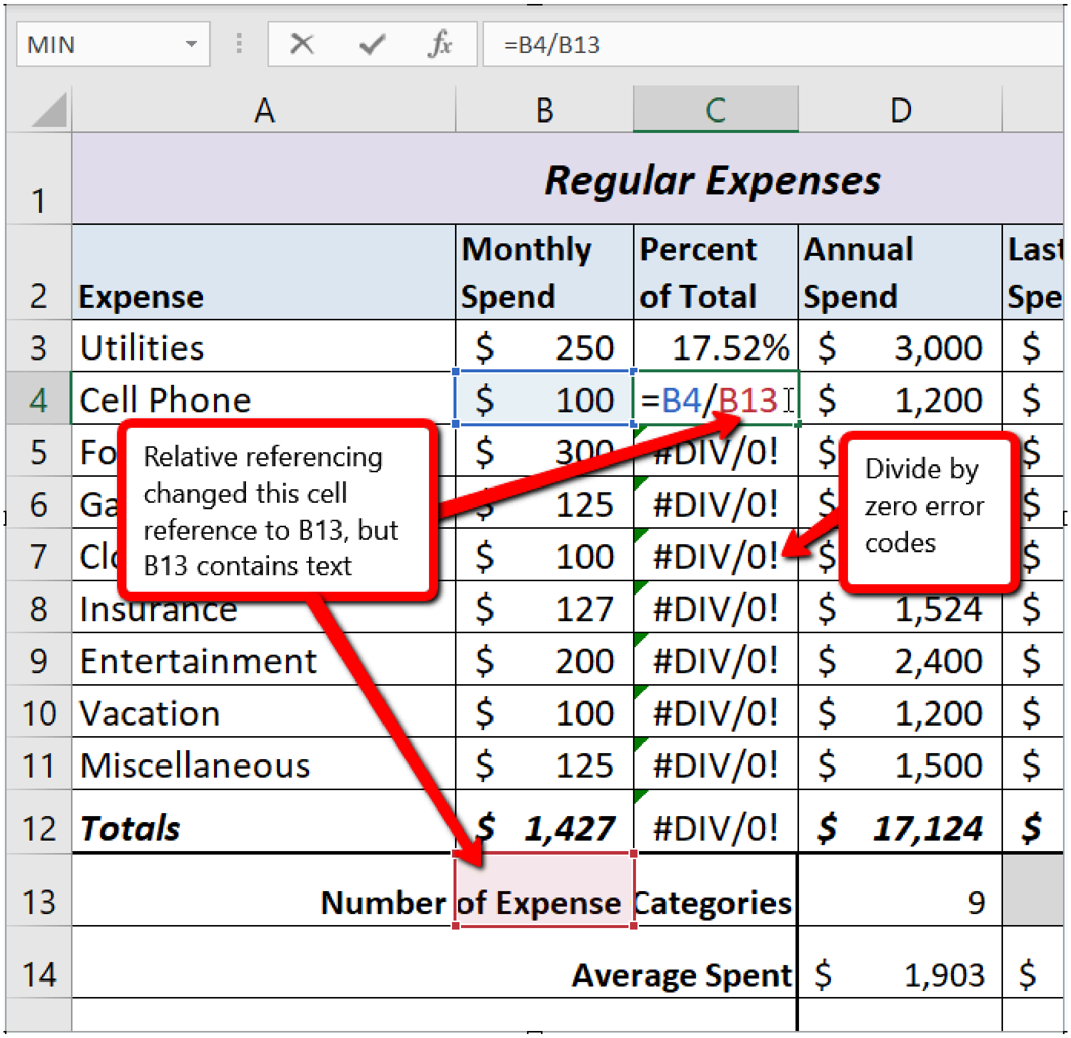


Figure 2.27 #DIV/0 Error from Relative Referencing

To eliminate the divide-by-zero error shown in **Figure 2.27** we must add an absolute reference to cell B12 in the formula. An absolute reference prevents relative referencing from changing a cell reference in a formula. This is also referred to as locking a cell. No matter where you copy a formula with an absolute reference, it will always refer back to the locked cell. An absolute reference is indicated by a $ sign in front of both the column letter and the row number. For example, **$A$15** is an absolute reference to cell A15.

We are going to modify the existing formula in C3 to make the reference to cell B12 an absolute reference. The following explains how this is accomplished:

$A$15 is an example of an absolute reference

1. Double click cell C3.
2. Place the mouse pointer in front of B12 and click. The blinking cursor should be in front of the B in the

cell reference B12.

3. Press the F4 key. You will see a dollar sign ($) added in front of the column letter B and the row number 12. *You can also type the dollar signs in front of the column letter and row number if you prefer. The formula should appear as =B3/$B$12. The* ***F4*** *key is a cool shortcut for adding the dollar signs.* Mac Users: If the **F4** key does not insert the $ symbols, check the keyboard settings: click black Apple icon at top left of the screen, choose "**System Preferences**", click the **Keyboard icon**, make sure the checkbox is checked for the item that says: "Use F1, F2, etc. as standard function keys".
4. Press the ENTER key.
5. Click cell C3.
6. Use the AutoFill Handle or Copy and Paste to copy the formula from C3 to the range C4:C11.

Figure 2.28 shows the percent of total formula with an absolute reference added to B12. Notice that in cell C4, the cell reference remains B12 instead of changing to B13. Also, you will see that the percentages are being calculated in the rest of the cells in the column, and the divide-by-zero error is now eliminated.

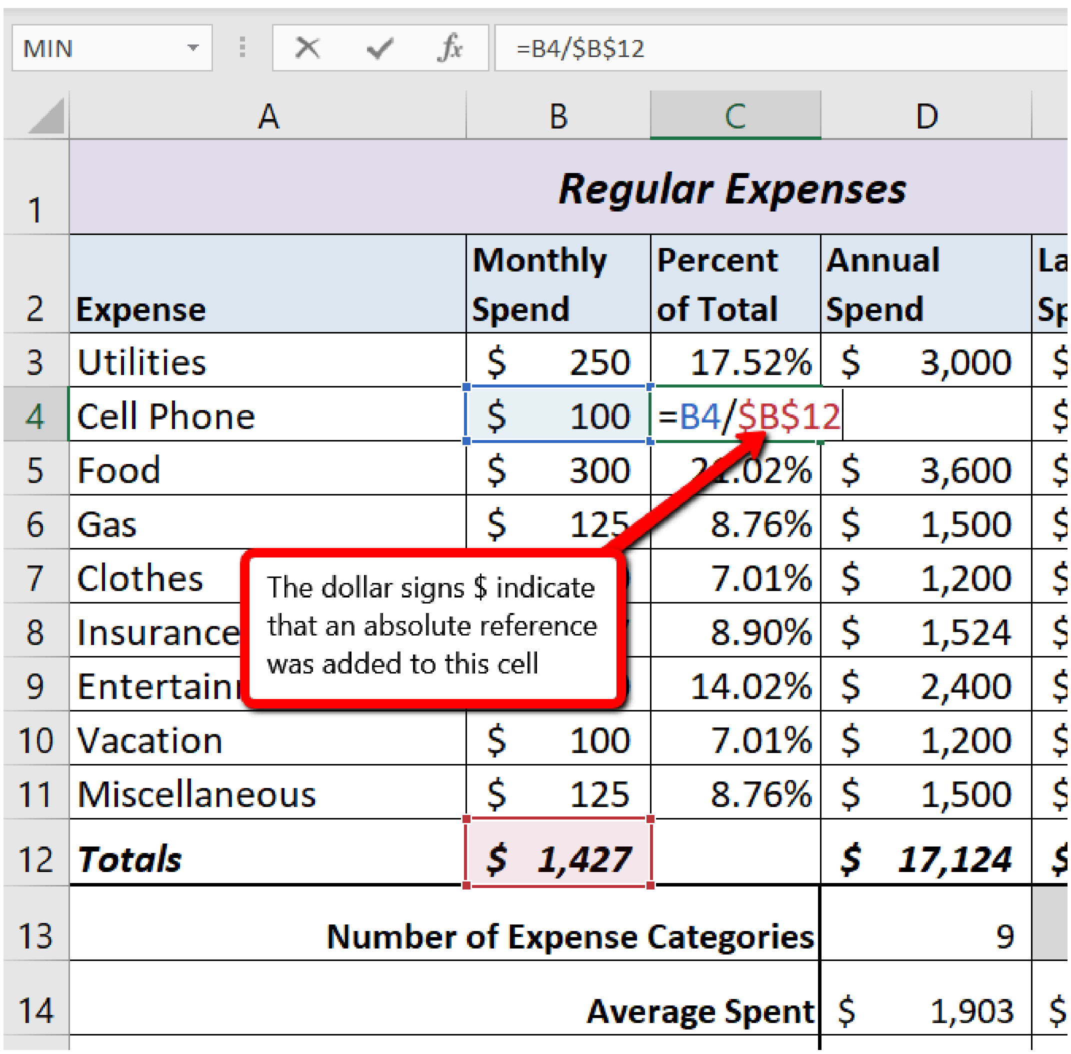


Figure 2.28 Adding an Absolute Reference to a Cell Reference in a Formula

Skill Refresher

Absolute References

1. Click in front of the column letter of a cell reference in a formula or function that you do not want altered when the formula or function is pasted into a new cell location.
2. Press the F4 key or type a dollar sign $ in front of the column letter and row number of the cell reference.

Sorting Data (Multiple Levels)

Data file: Continue with CH2 Personal Budget.

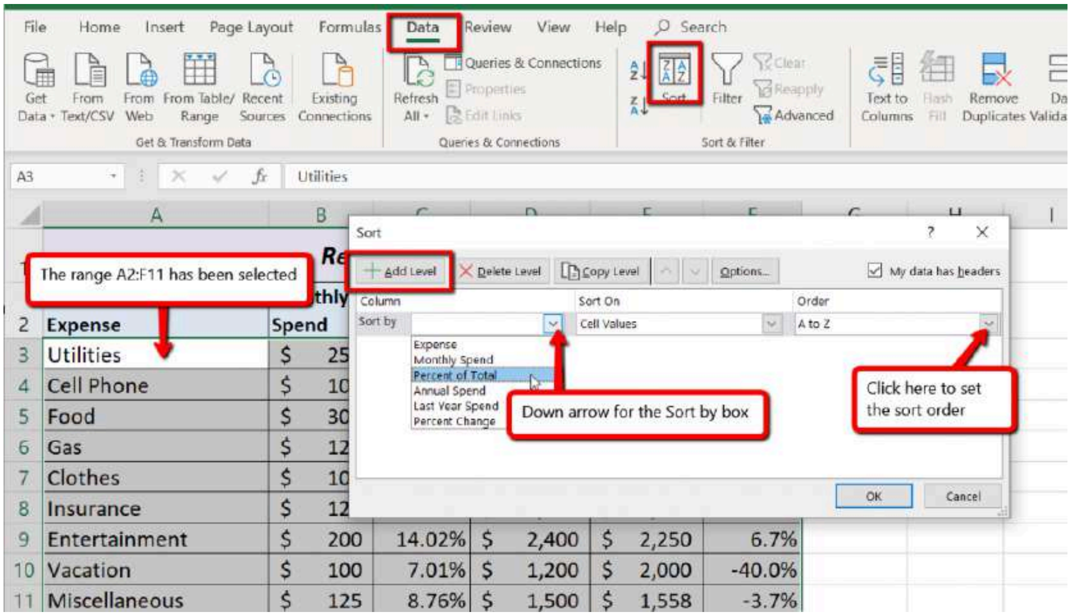
The **Budget Detail** worksheet shown in **Figure 2.28** is now producing several mathematical outputs through formulas and functions. The outputs allow you to analyze the details and identify trends as to how money is being budgeted and spent. Before we draw some conclusions from this worksheet, we will sort the data based on the Percent of Total column. Sorting is a powerful tool that enables you to analyze key trends in any data set. Sorting will be covered thoroughly in a later chapter, but will be briefly introduced here.

For the purposes of the **Budget Detail** worksheet, we want to set multiple levels for the sort order. We are going to sort first by the Percent of Total, and then by the Last Year Spend amount. Excel will first sort the items by the Percent of Total, and any items with the same Percent of Total will then be sorted by Last Year Spend. This is accomplished through the following steps:

1. Highlight the range A2:F11.
2. Click the Data tab in the Ribbon.
3. Click the Sort button in the Sort & Filter group of commands. This opens the Sort dialog box, as shown in **Figure 2.29**.
4. Click the down arrow next to the "Sort by" box.
5. Click the Percent of Total option from the drop-down list.
6. Click the down arrow next to the sort Order box.
7. Click the Largest to Smallest option.
8. Click the Add Level button. This allows you to set a second level for any duplicate values in the Percent of Total column.

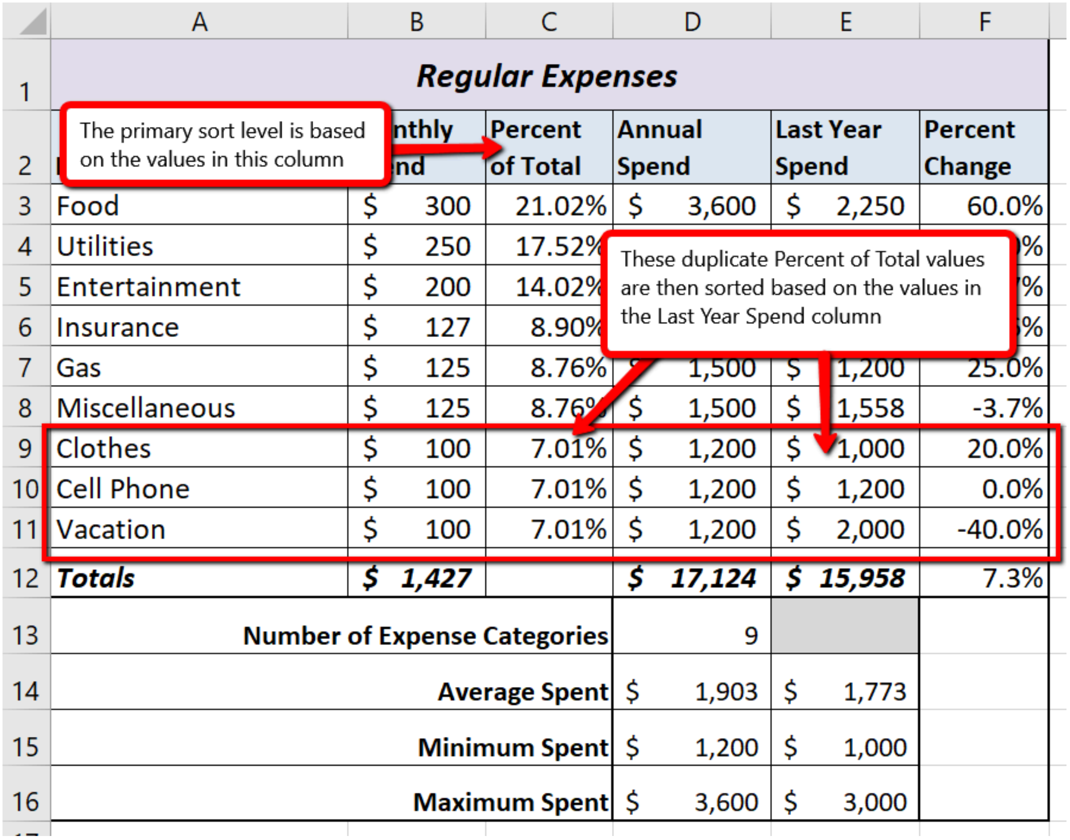
   the + symbol at bottom left corner is the "Add Level" button for Excel for Mac
9. Click the down arrow next to the "Then by" box.

10. Select the Last Year Spend option. Leave the Sort Order as Smallest to Largest
11. Click the OK button at the bottom of the Sort dialog box.
12. Save the **CH2 Personal Budget** file.



2.29 Sort Dialog Box

Figure 2.30 shows the **Budget Detail** worksheet after it has been sorted. Notice that there are three identical values in the Percent of Total column. This is why a second sort level had to be created for this worksheet. The second sort level arranges the values of 7.01% based on the values in the Last Year Spend column in ascending order. Excel gives you the option to set as many sort levels as necessary for the data contained in a worksheet.



	A	B	C	D	E	F
1	***Regular Expenses***					
2		**nthly nd**	**Percent of Total**	**Annual Spend**	**Last Year Spend**	**Percent Change**
3	Food	$ 300	21.02%	$ 3,600	$ 2,250	60.0%
4	Utilities	$ 250	17.52%			%
5	Entertainment	$ 200	14.02%			%
6	Insurance	$ 127	8.90%			%
7	Gas	$ 125	8.76%	1,500	$ 1,200	25.0%
8	Miscellaneous	$ 125	8.76%	$ 1,500	$ 1,558	-3.7%
9	Clothes	$ 100	7.01%	$ 1,200	$ 1,000	20.0%
10	Cell Phone	$ 100	7.01%	$ 1,200	$ 1,200	0.0%
11	Vacation	$ 100	7.01%	$ 1,200	$ 2,000	-40.0%
12	***Totals***	***$ 1,427***		***$ 17,124***	***$ 15,958***	7.3%
13	**Number of Expense Categories**			9		
14	**Average Spent**			$ 1,903	$ 1,773	
15	**Minimum Spent**			$ 1,200	$ 1,000	
16	**Maximum Spent**			$ 3,600	$ 3,000	

Figure 2.30 Budget Detail Worksheet after Sorting

Skill Refresher

Sorting Data (Multiple Levels)

1. Highlight a range of cells to be sorted.
2. Click the Data tab of the Ribbon.
3. Click the Sort button in the Sort & Filter group.
4. Select a column from the “Sort by” drop-down list in the Sort dialog box.
5. Select a sort order from the Order drop-down list in the Sort dialog box.
6. Click the Add Level button in the Sort dialog box.
7. Repeat Steps 4 and 5.
8. Click the OK button on the Sort dialog box.

Key Takeaways

- Statistical functions are used when a mathematical process is required for a range of cells, such as summing the values in several cell locations. For these computations, functions are preferable to formulas because adding many cell locations one at a time to a formula can be very time-consuming.
- Statistical functions can be created using cell ranges or selected cell locations separated by commas. Make sure you use a cell range (two cell locations separated by a colon) when applying a statistical function to a contiguous range of cells.
- To prevent Excel from changing the cell references in a formula or function when they are pasted to a new cell location, you must use an absolute reference. You can do this by placing a dollar sign ($) in front of the column letter and row number of a cell reference.
- The #DIV/0 error appears if you create a formula that attempts to divide a constant or the value in a cell reference by zero.
- The Paste Formulas option is used when you need to paste formulas without any formatting treatments into cell locations that have already been formatted.
- You need to set multiple levels, or columns, in the Sort dialog box when sorting data that contains several duplicate values.

Attribution

Adapted by Mary Schatz from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

4.XLSX.3 FUNCTIONS FOR PERSONAL FINANCE

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Learning Objectives

1. Understand the fundamentals of loans.
2. Use the PMT function to calculate monthly car loan payments.
3. Use the PMT function to calculate monthly mortgage payments on a house using a down payment.
4. Learn how to summarize data in a workbook by using worksheet links to create a summary worksheet.

In this section, we continue to develop the **Personal Budget** workbook. Notable items that are missing from the **Budget Detail** worksheet are the payments you might make for a car or a home. This section demonstrates Excel functions used to calculate loan payments for a car and to calculate mortgage payments for a house.

The Fundamentals of Loans and Leases

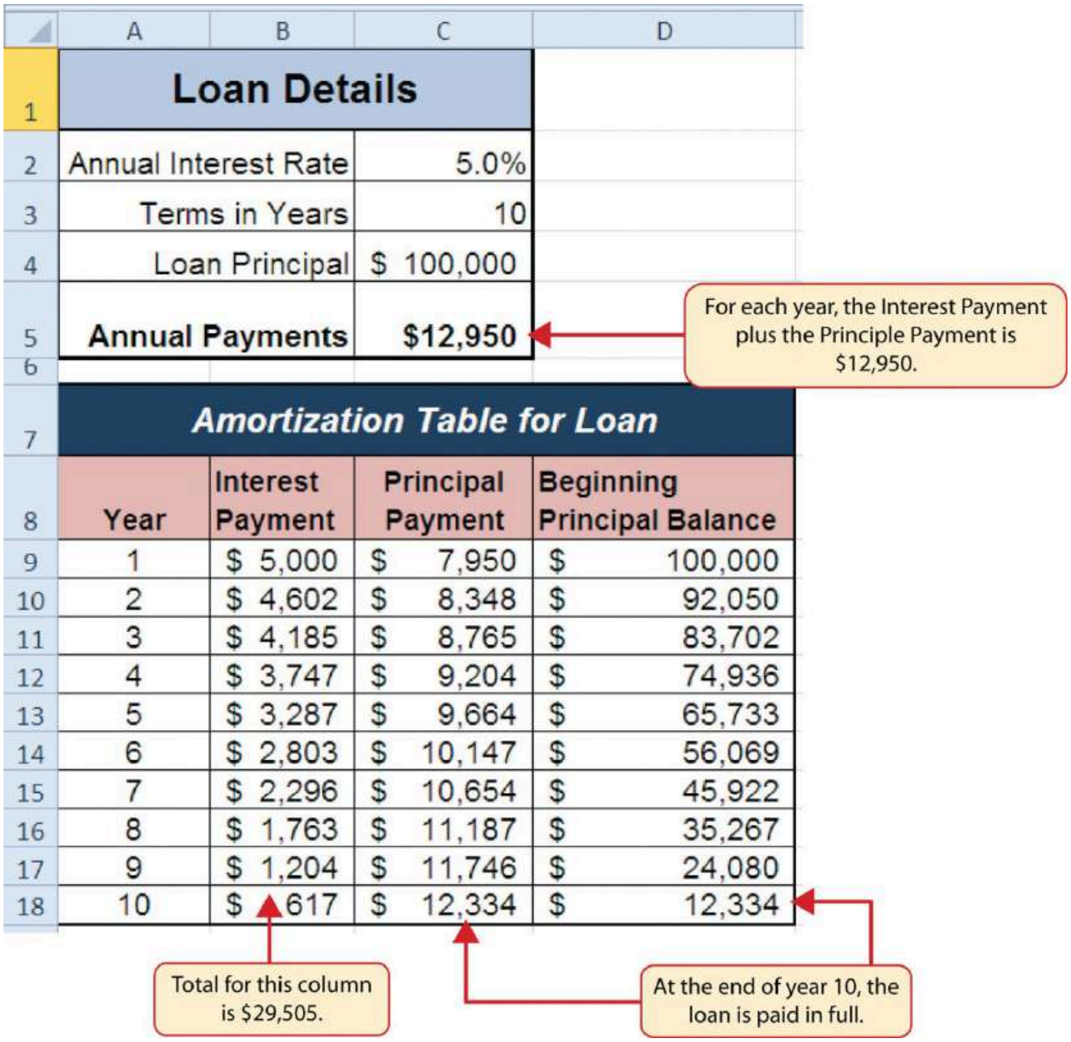
One of the functions we will add to the Personal Budget workbook is the PMT function. This function calculates the payments required for loan repayment. However, before demonstrating this function, it is important to cover a few fundamental concepts on loans.

A loan is a contractual agreement in which money is borrowed from a lender and paid back over a specific period of time. The amount of money that is borrowed from the lender is called the **principal** of the loan. The borrower is usually required to pay the principal of the loan plus interest. When you borrow money to buy a house, the loan is referred to as a mortgage. This is because the house being purchased also serves as collateral to ensure payment. In other words, the bank can take possession of your house if you fail to make loan payments. As shown in **Table 2.5**, there are several key terms related to loans.

Table 2.5 Key Terms for Loans

Term	Definition
Collateral	Any item of value that is used to secure a loan to ensure payments to the lender
Down Payment	The amount of cash paid toward the purchase of a house. If you are paying 20% down, you are paying 20% of the cost of the house in cash and are borrowing the rest from a lender.
Interest Rate	The interest that is charged to the borrower as a cost for borrowing money
Mortgage	A loan where property is put up for collateral
Principal	The amount of money that has been borrowed
Residual Value	The estimated selling price of a vehicle at a future point in time
Length	The amount of time you have to repay a loan

Figure 2.31 shows an example of an amortization table for a loan. A lender is required by law to provide borrowers with an amortization table when a loan contract is offered. The table in the figure shows how the payments of a loan would work if you borrowed $100,000 from a lender and agreed to pay it back over 10 years at an interest rate of 5%. You will notice that each time you make a payment, you are paying the bank an interest fee plus some of the loan principal. Each year the amount of interest paid to the bank decreases and the amount of money used to pay off the principal increases. This is because the bank is charging you interest on the amount of principal that has not been paid. As you pay off the principal, the interest rate is applied to a lower number, which reduces your interest charges. Finally, the figure shows that the sum of the values in the Interest Payment column is $29,505. This is how much it costs you to borrow this money over 10 years. Indeed, borrowing money is not free. It is important to note that to simplify this example, the payments were calculated on an annual basis. However, most loan payments are made on a monthly basis.



	A	B	C	D
1	Loan Details			
2	Annual Interest Rate		5.0%	
3	Terms in Years		10	
4	Loan Principal		$ 100,000	
5	Annual Payments		$12,950	
6				
7	Amortization Table for Loan			
8	Year	Interest Payment	Principal Payment	Beginning Principal Balance
9	1	$ 5,000	$ 7,950	$ 100,000
10	2	$ 4,602	$ 8,348	$ 92,050
11	3	$ 4,185	$ 8,765	$ 83,702
12	4	$ 3,747	$ 9,204	$ 74,936
13	5	$ 3,287	$ 9,664	$ 65,733
14	6	$ 2,803	$ 10,147	$ 56,069
15	7	$ 2,296	$ 10,654	$ 45,922
16	8	$ 1,763	$ 11,187	$ 35,267
17	9	$ 1,204	$ 11,746	$ 24,080
18	10	$ 617	$ 12,334	$ 12,334

Figure 2.31 Example of an Amortization Table

The PMT (Payment) Function for Loans

Data file: Continue with CH2 Personal Budget.

If you own a home, your mortgage payments are a major component of your household budget. If you are planning to buy a home, having a clear understanding of your monthly payments is critical for maintaining strong financial health. In Excel, mortgage payments are conveniently calculated through the PMT (payment) function. This function is more complex than the statistical functions covered in Section 2.2 “Statistical Functions”. With statistical functions, you are required to add only a range of cells or selected cells within the parentheses of the function, also known as the argument. With the PMT function, you must accurately define a series of arguments in order for the function to produce a reliable output. **Table 2.6** lists the arguments for the PMT function. It is helpful to review the key loan terms in **Table 2.5** before reviewing the PMT function arguments.

Table 2.6 Arguments for the PMT Function

Argument	Definition
Rate	This is the interest rate the lender is charging the borrower. The interest rate is usually quoted in annual terms, so you have to divide this rate by 12 if you are calculating monthly payments.
Nper	The argument letters stand for number of periods. This is the term of the loan, which is the amount of time you have to repay the bank. This is usually quoted in years, so you have to multiply the years by 12 if you are calculating monthly payments.
Pv	The argument letters stand for present value. This is the principal of the loan or the amount of money that is borrowed.
[Fv]	The argument letters stand for future value. The brackets around the argument indicate that it is not always necessary to define it. It is used if there is a lump-sum payment that will be made at the end of the loan terms. This is also used for the residual value of a lease. If it is not defined, Excel will assume that it is zero.
[Type]	This argument can be defined with either a 1 or a 0. The number 1 is used if payments are made at the beginning of each period. A 0 is used if payments are made at the end of each period. The argument is in brackets because it does not have to be defined if payments are made at the end of each period. Excel assumes that this argument is 0 if it is not defined.

By default, the result of the PMT function in Excel is shown as a negative number. This is because it represents an outgoing payment. When making a mortgage or car payment, you are paying money out of your pocket or bank account. Depending on the type of work that you do, your employer may want you to leave your payments negative or they may ask you to format them as positive numbers. In the following assignments, the payments calculated using the PMT function will be made positive to make them easier to work with. To do this, you will place a negative sign between the equal sign and the function name PMT.

We will first use the PMT function in the Personal Budget workbook to calculate the monthly loan payments for a car. These calculations will be made in the **Loan Payments** worksheet and then displayed in the **Budget Summary** worksheet through a cell reference link. So far we have demonstrated several methods for adding functions to a worksheet. When working with more complex functions such as the PMT, it is easiest to use the Function Dialog box.

Remember to use cell references for the arguments of the PMT function whenever possible. This will allow you the flexibility to change aspects of the loan, such as a lower interest rate or more expensive car, and have the payment automatically recalculate.

The following steps use the Insert Function command covered in Section 2.2 to add the PMT function:

Using cell references for the arguments provides greater flexibility in trying different scenarios.

1. Switch to the **Loan Payments** worksheet.
2. Click cell B5.
3. Click the Formulas tab on the Ribbon.

4. Click the Insert Function button to bring up the Insert Function dialog box.
5. Type loan **payment** in the search box and click Go.
   the Excel for Mac search box does is not the same as the "Search for a function: input box". Mac Users must type: **PMT** in the search box instead. Then press Enter.
6. Double-click the PMT option in the "Select a function:" box. This will open the Function Arguments dialog box.
7. Drag the Function Arguments dialog box out of the way so that you can see the worksheet cells you want to use in the function. Refer to **Figure 2.31** for the completed Function Arguments dialog box as you complete the next steps.
8. Click in the Rate argument box in the dialog box, then click cell B3 in the worksheet. This will add B3 (the annual interest rate) to the Rate argument.
9. Type a forward slash / for division.
10. Type the number **12**. *Since our goal is to calculate the monthly payments for the loan, we need to divide the rate, which is stated in annual terms, by 12. This converts the annual rate to a monthly rate.*
11. Click the Nper argument box (or use the Tab key) and then click cell B4 in the worksheet. This will add B4 (the number of years to repay the loan) to the Nper argument.
12. Type an **asterisk** * for multiplication.
13. Type the number **12**. *Since our goal is to calculate the monthly payments for the loan, we need to multiply the terms of the loan by 12. This converts the terms of the loan from years to months.*
14. Click the Pv argument box (or use the Tab key) and then click cell B2 in the worksheet. This will add B2 (the amount of the loan) to the Pv argument.
15. You will now see the Rate, Nper, and Pv arguments defined for the function. (**see Figure 2.31**)
16. Click the OK button at the bottom of the Function Arguments dialog box. The function will now be placed into the worksheet. Since we are not paying any lump sums of money at the end of the loan, there is no need to define the Fv argument. Also, we will assume that the monthly payments will be made at the end of each month. Therefore, there is no need to define the Type argument.
17. Notice that the result of the formula in cell B5 is showing as a negative number (see **Figure 2.32).** To fix this, double-click on cell B5 and type a negative sign between the equal sign and the letters PMT in the formula (see **Figure 2.33**).
18. The finished formula in cell B5 should be =**-PMT(B3/12,B4*12,B2)**

Figure 2.31 shows the completed Function Arguments dialog box for the PMT function. Notice that the dialog box shows the values for the Rate and Nper arguments. The Rate is divided by 12 to convert the annual interest rate to a monthly interest rate. The Nper argument is multiplied by 12 to convert the terms of the loan from years to months. Finally, the dialog box provides you with a definition for each argument. The definition appears when you click in the input box for the argument.

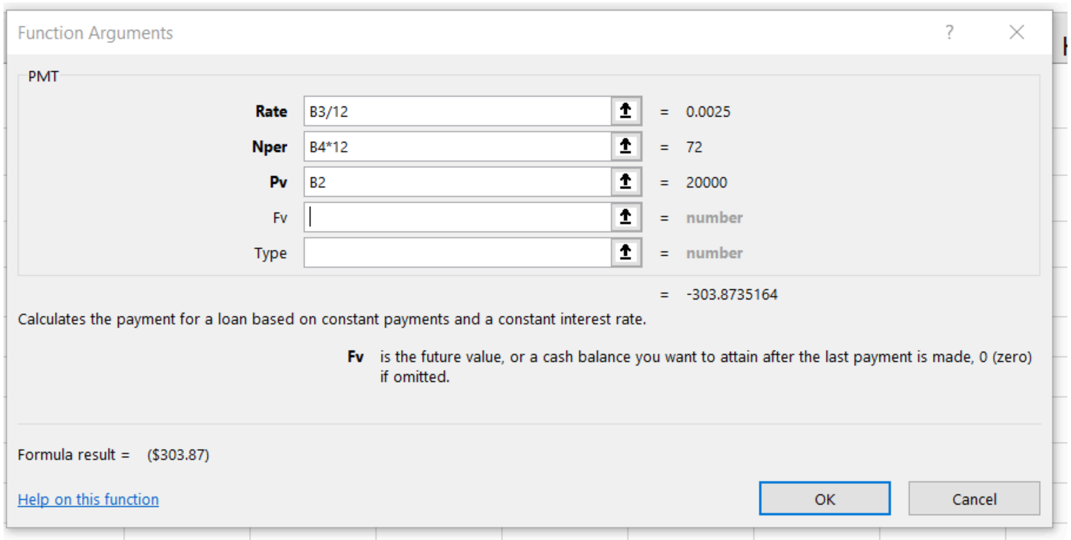


Figure 2.31 Function Arguments Dialog Box for the PMT function

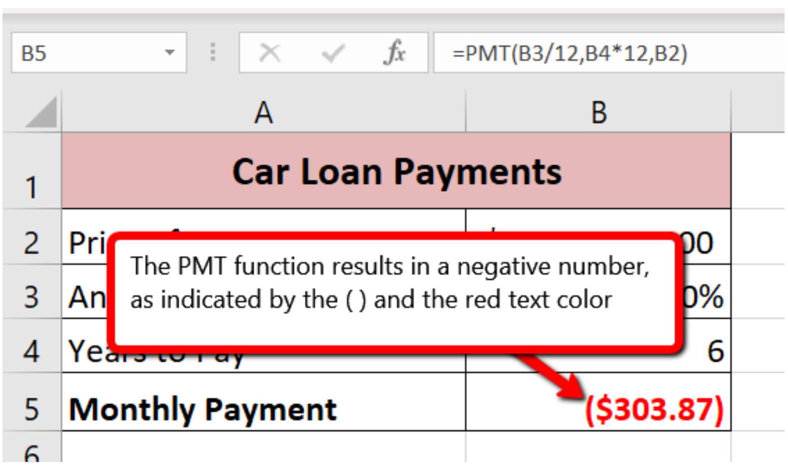


Figure 2.32 Result of the PMT Function as a Negative Number

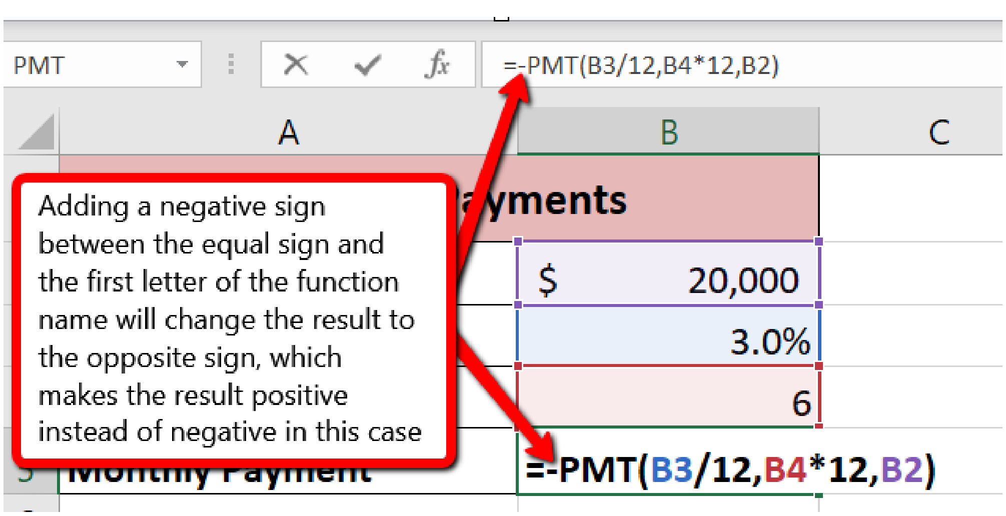


Figure 2.33 The PMT Function Modified to Result in a Positive Number

Keyboard Shortcuts

Insert Function

- Hold the SHIFT key while pressing the F3 key.

Function Arguments Dialog Box

- After the equal sign **=** and function name are typed into cell a location, hold down the CTRL key and press the letter A on your keyboard.

Integrity Check

Comparable Arguments for PMT Function

When using functions such as PMT, make sure the arguments are defined in comparable terms. For example, if you are calculating the monthly payments of a loan, make sure both the Rate and Nper argument are expressed in terms of months. The function will produce an erroneous result if one argument is expressed in years while the other is expressed in months.

The PMT Function when there is a down payment

In addition to calculating the loan payments for a car, the PMT function will be used in the Personal Budget workbook to calculate the mortgage payments for a home. The details for the mortgage payments are also found in the **Loan Payments** worksheet. Unlike the car loan, there is a down payment with the mortgage. A down payment on a mortgage is usually a percentage of the price of the home, which is paid up front and reduces the amount of the loan itself. The down payment amount and amount of the loan will both need to be calculated using formulas. While we did not use a down payment in the car loan example, it is fairly common to have a down payment when purchasing a car too.

Write the formulas to calculate the Down Payment Amount and new Loan Amount by following these steps:

1. Click cell B11.
2. Write the formula =**B9*B10**. *This will calculate 20% of the price of the house.*
3. Click cell B12. Write the formula =**B9-B11**. *This will subtract the down payment amount from the price of the house* (see **Figure 2.34** for the Show Formulas View and **Figure 2.35** for the formula results).

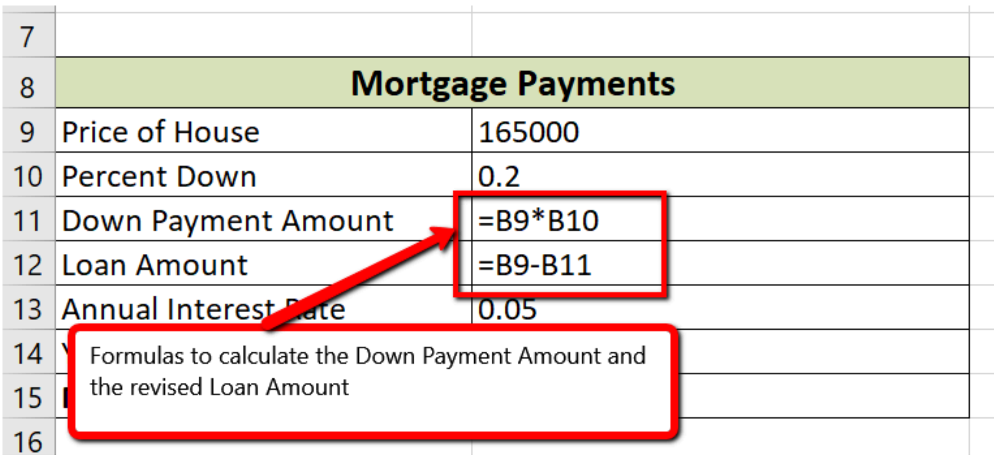


Figure 2.34 Show Formulas View

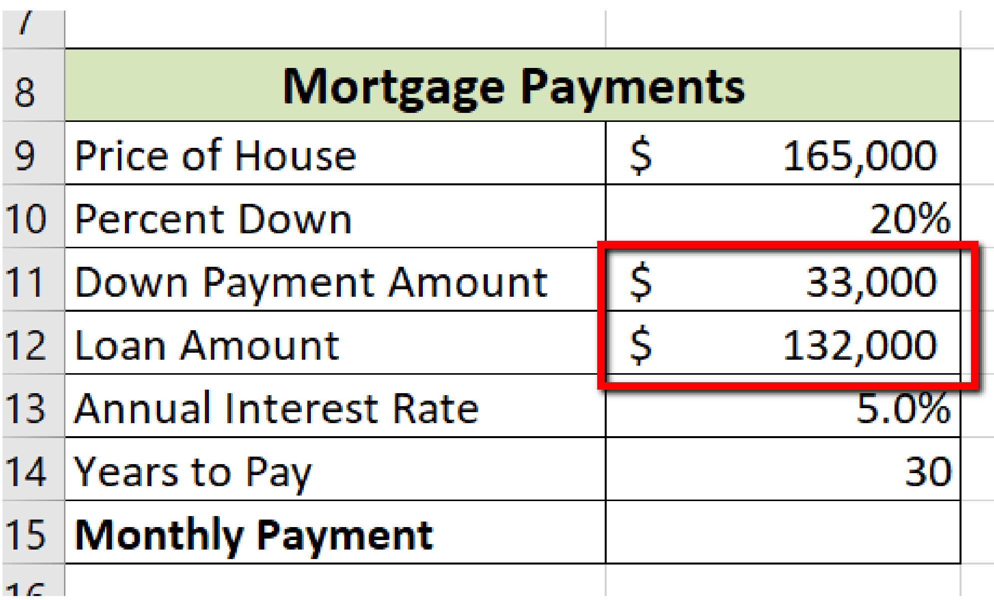


Figure 2.35 Results of the Down Payment Amount and Revised Loan Amount Formulas

Now that we have the revised Loan Amount in cell B12, we can write the PMT function following the same process we did for the car loan.

1. Click cell B15.
2. Click the Formulas tab on the Ribbon.

3. Click the Insert Function button to bring up the Insert Function dialog box.
4. Type **PMT** in the search box and click Go.
5. Double-click the PMT option in the "Select a function:" box. This will open the Function Arguments dialog box.
6. Enter the following arguments (see **Figure 2.36**)

    - Rate: **B13/12** –> divide by 12 to convert the annual rate to a monthly one
    - Nper: **B14*12** –> multiply by 12 to convert the number of years into number of months
    - Pv: **B12** –> this is the cell with the actual loan amount, not the price of the house

7. Click OK in the Function Arguments dialog box.
8. Modify the formula in cell B15 to display the result as a positive number. *Remember to type a negative sign between the equal sign and the letters PMT.*
9. Cell B15 should contain the function: **=-PMT(B13/12,B14*12,B12)** and the result should be $708.60 (see **Figure 2.37**).

Figure 2.36 shows how the the completed Function Arguments dialog box for the PMT function for the mortgage should appear before pressing the OK button.

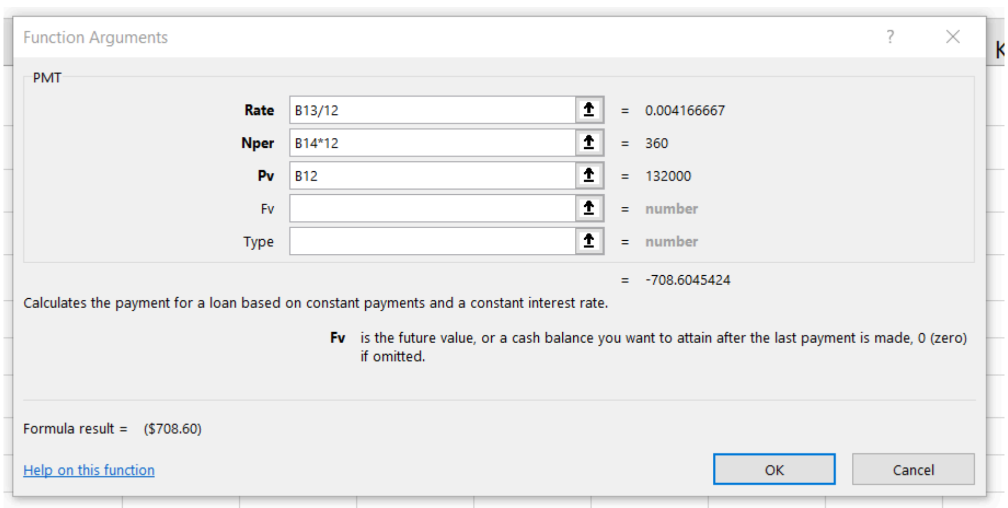


Figure 2.36 Function Arguments Dialog Box for the Mortgage Payment PMT Function

Figure 2.37 shows the result of the PMT function for the mortgage. The monthly payments for this mortgage are $708.60. This monthly payment will be displayed in the **Budget Summary** worksheet.

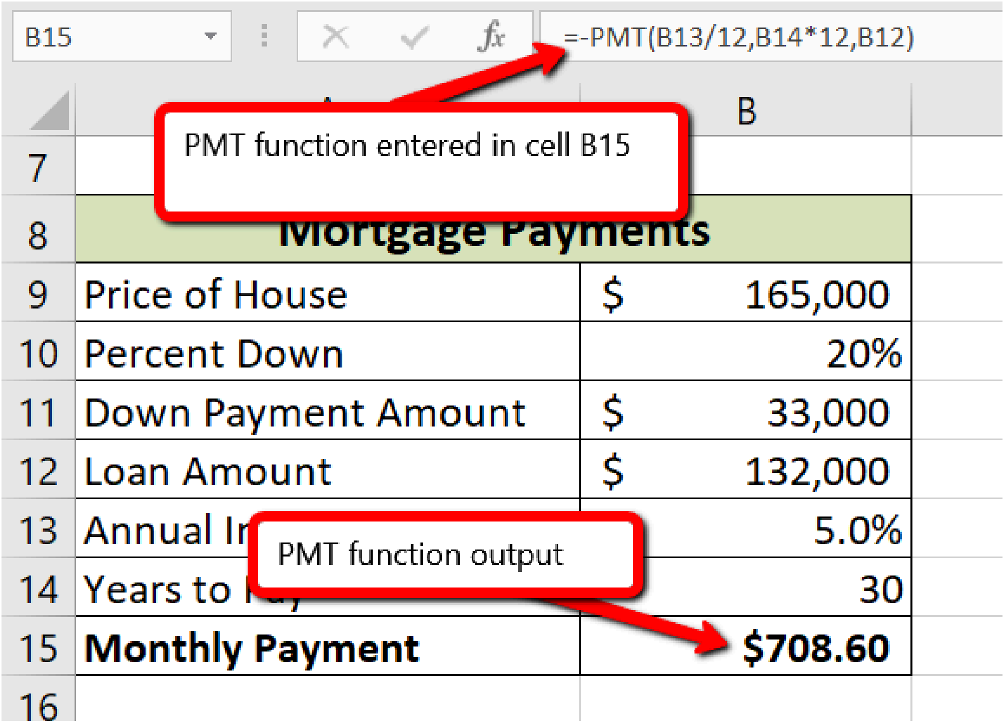


Figure 2.37 Mortgage Monthly Payment Calculation

Skill Refresher

PMT Function

1. Type an equal sign **=**.
2. Type the letters **PMT** followed by an open parenthesis, or double click the function name from the function list.
3. Define the Rate argument with a cell location that contains the rate being charged by the lender for the loan or lease. If the interest rate given is an annual rate, divide it by 12 to convert it to a monthly rate.
4. Define the Nper argument with a cell location that contains the amount of time to repay the loan or lease. If the amount of time is in years, multiply it by 12 to convert it to number of months.
5. Define the Pv argument with a cell location that contains the principal of the loan or the price of the item being leased.
6. Type a closing parenthesis **)**.
7. Press the ENTER key.
8. If the result needs to be shown as a positive number, add a negative sign between the equal sign and the letters PMT.

Linking Worksheets (Creating a Summary Worksheet)

So far we have used cell references in formulas and functions, which allow Excel to produce new outputs when the values in the cell references are changed. Cell references can also be used to display values or the outputs of formulas and functions in cell locations on other worksheets. This is how we will complete the Budget Summary worksheet using values from both the Budget Detail and Loan Payments worksheets.

Outputs from the formulas and functions that were entered into the **Budget Detail** will be displayed on the **Budget Summary** worksheet through the use of cell references.

1. Switch to the Budget Summary worksheet and select cell B4. *This cell needs to reference the Total Annual Spend (D12) from the Budget Detail worksheet.*
2. Type an =
3. Click the **Budget Detail** worksheet tab.
4. Click cell D12.
5. Press the ENTER key on your keyboard.
6. The formula bar will display the formula =**'Budget Detail'!D12** and the cell will display $17,124. (**see Figure 2.38)**

Figure 2.38 shows how the cell reference appears in the **Budget Summary** worksheet. Notice that the cell reference D12 is preceded by the **Budget Detail** worksheet name enclosed in apostrophes followed by an exclamation point ('Budget Detail'!) This indicates that the value displayed in the cell is referencing a cell location in the **Budget Detail** worksheet.

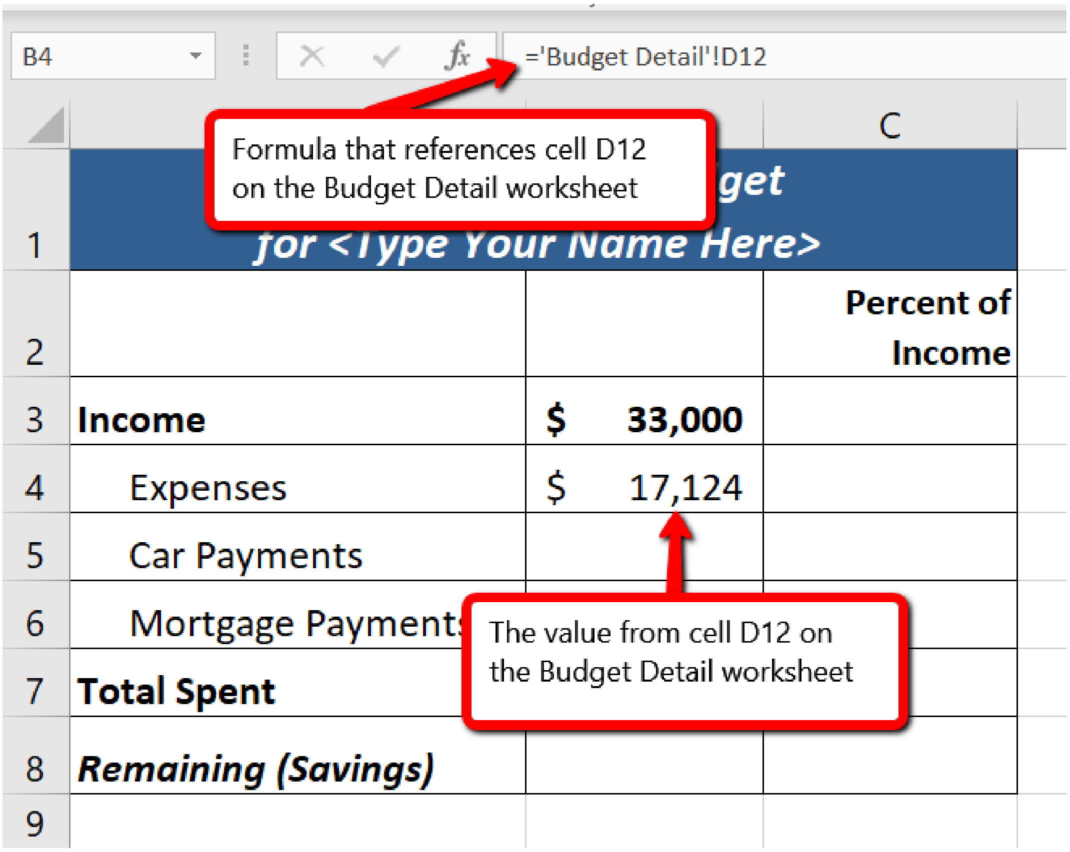


Figure 2.38 Cell Reference Showing the Total Annual Spend from the Budget Detail Worksheet

We will use a similar process to enter in the annual car payments and mortgage payments from the Loan Payments worksheet. The payments on the Loan Payments worksheet are monthly payments though, so we will need to multiply each one by 12 to get the **annual** amount to display in the Budget Summary worksheet.

1. Click on cell B5. *This cell needs to contain a formula that references the monthly car payment cell (B5) on the Loan Payments worksheet and multiplies by 12.*
2. Type an =
3. Click the **Loan Payments** worksheet tab.
4. Click cell B5 on the Loan Payments worksheet.
5. The formula bar will display the formula =**'Loan Payments'!B5**
6. Type an asterisk * for multiplication.
7. Type the number **12**. The formula in the formula bar should read: =**'Loan Payments'!B5*12**
8. Press the ENTER key on your keyboard.
9. Click on cell B6. *This cell needs to contain a formula that references the monthly mortgage payment cell (B15) on the Loan Payments worksheet and multiplies by 12.*
10. Type an =
11. Click the **Loan Payments** worksheet tab.
12. Click cell B15 on the Loan Payments worksheet.

13. The formula bar will display the formula =**'Loan Payments'!B15**
14. Type an asterisk * for multiplication.
15. Type the number **12**. The formula in the formula bar should read: =**'Loan Payments'!B15*12**
16. Press the ENTER key on your keyboard.

Figure 2.39 shows the results of creating formulas that reference cell locations in the **Loan Payments** worksheet.

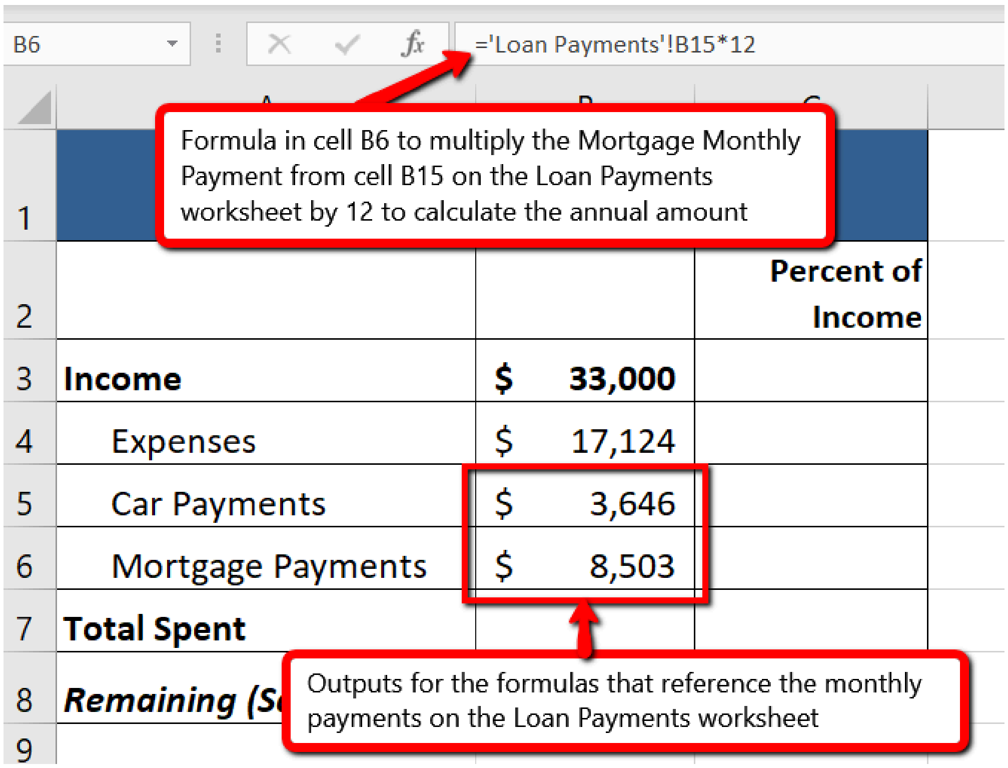


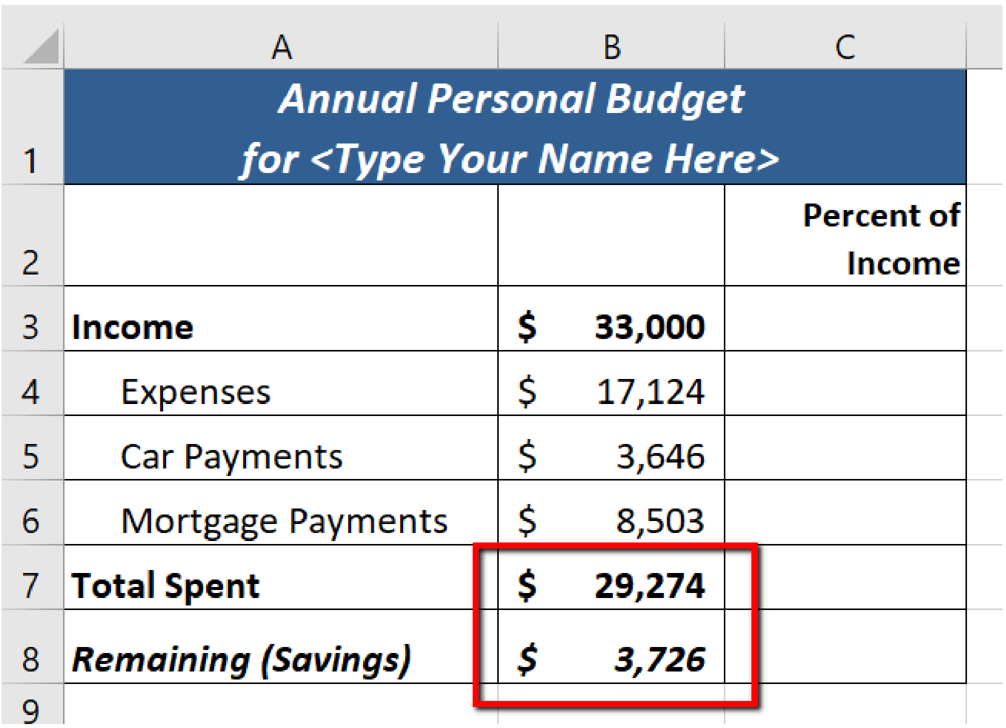
Figure 2.39 Results of the Formulas for the Annual Loan Payments

We can now add other formulas and functions to the **Budget Summary** worksheet that can calculate the difference between the total spend dollars vs. the total net income in cell B3. The following steps explain how this is accomplished:

1. Click cell B7 in the **Budget Summary** worksheet.
2. Type an equal sign =.
3. Type the function name **SUM** followed by an open parenthesis **(**.
4. Highlight the range B4:B6.
5. Type a closing parenthesis **)** and press the ENTER key on your keyboard or simply press the ENTER key to close the function. *The total for all annual expenses now appears on the worksheet.*
6. Click cell B8 on the **Budget Summary** worksheet. You will enter a formula to calculate Remaining (Savings) amount in this cell.

7. Type an equal sign =.
8. Click cell B3.
9. Type a minus sign – and then click cell B7.
10. Press the ENTER key on your keyboard. This formula produces a positive number, indicating our income is greater than our total expenses.

Figure 2.40 shows the results of the formulas that were added to the **Budget Summary** worksheet. Overall, having your income exceed your total expenses is a good thing because it allows you to save money for future spending needs or unexpected events.



	A	B	C
1	***Annual Personal Budget for <Type Your Name Here>***		
2			**Percent of Income**
3	**Income**	**$ 33,000**	
4	Expenses	$ 17,124	
5	Car Payments	$ 3,646	
6	Mortgage Payments	$ 8,503	
7	**Total Spent**	**$ 29,274**	
8	***Remaining (Savings)***	***$ 3,726***	
9			

Figure 2.40 Formulas Added to Calculate Amount Remaining for Savings

We can now add a few formulas that calculate both the spending rate and the savings rate as a percentage of net income. These formulas require the use of absolute references, which we covered earlier in this chapter. The following steps explain how to add these formulas:

1. Click cell C7 in the **Budget Summary** worksheet.
2. Type an equal sign =.
3. Click cell B7.
4. Type a forward slash / for division and then click B3.
5. Press the F4 key on your keyboard. This adds an absolute reference to cell B3.
6. Press the ENTER key. The result of the formula shows that total expenses consume 89% of our net income.
7. Click cell C7.

8. Place the mouse pointer over the Auto Fill Handle.
9. When the mouse pointer turns to a black plus sign, left click and drag down to cell C8. This copies and pastes the formula into cell C8.
10. Compare your worksheets with Figures 2.41a-c below. Make any necessary changes before moving on to the next section.
11. Save the **CH2 Personal Budget** file.

Figure 2.41a shows the completed **Budget Summary** worksheet

	A	B	C	D
1	***Annual Personal Budget for <Type Your Name Here>***			
2			**Percent of Income**	
3	**Income**	**$ 33,000**		
4	Expenses	$ 17,124		
5	Car Payments	$ 3,646		
6	Mortgage Payments	$ 8,503		
7	**Total Spent**	**$ 29,274**	**89%**	
8	***Remaining (Savings)***	***$ 3,726***	**11%**	
9				

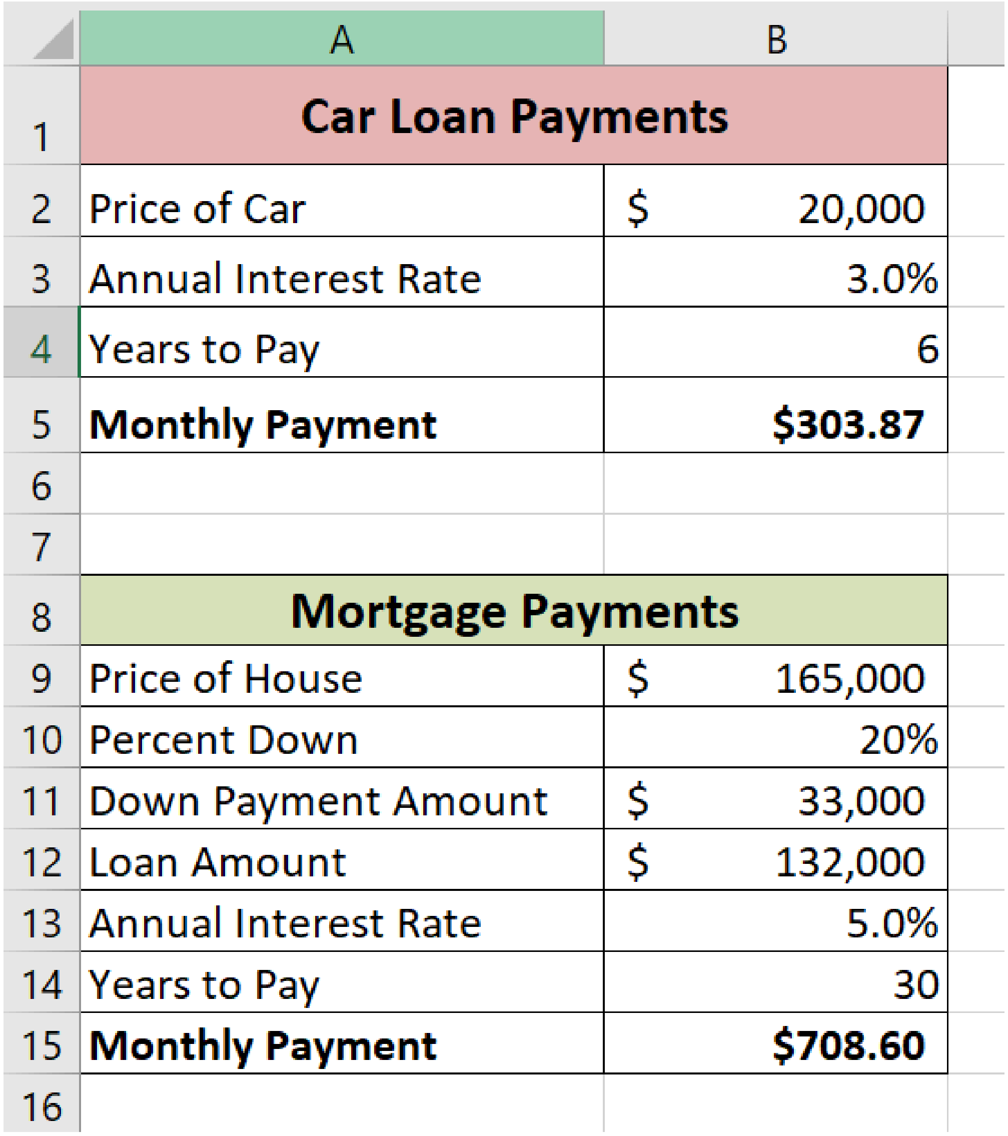
Figure 2.41a Completed Budget Summary worksheet

Figure 2.41b shows the completed **Budget Detail** worksheet

	A	B	C	D	E	F
1	***Regular Expenses***					
2	**Expense**	**Monthly Spend**	**Percent of Total**	**Annual Spend**	**Last Year Spend**	**Percent Change**
3	Food	$ 300	21.02%	$ 3,600	$ 2,250	60.0%
4	Utilities	$ 250	17.52%	$ 3,000	$ 3,000	0.0%
5	Entertainment	$ 200	14.02%	$ 2,400	$ 2,250	6.7%
6	Insurance	$ 127	8.90%	$ 1,524	$ 1,500	1.6%
7	Gas	$ 125	8.76%	$ 1,500	$ 1,200	25.0%
8	Miscellaneous	$ 125	8.76%	$ 1,500	$ 1,558	-3.7%
9	Clothes	$ 100	7.01%	$ 1,200	$ 1,000	20.0%
10	Cell Phone	$ 100	7.01%	$ 1,200	$ 1,200	0.0%
11	Vacation	$ 100	7.01%	$ 1,200	$ 2,000	-40.0%
12	***Totals***	***$ 1,427***		***$ 17,124***	***$ 15,958***	7.3%
13	**Number of Expense Categories**			9		
14	**Average Spent**			$ 1,903	$ 1,773	
15	**Minimum Spent**			$ 1,200	$ 1,000	
16	**Maximum Spent**			$ 3,600	$ 3,000	
17						

Figure 2.41b Completed Budget Detail worksheet

Figure 2.41c shows the completed **Loan Payments** worksheet



	A	B
1	**Car Loan Payments**	
2	Price of Car	$ 20,000
3	Annual Interest Rate	3.0%
4	Years to Pay	6
5	**Monthly Payment**	**$303.87**
6		
7		
8	**Mortgage Payments**	
9	Price of House	$ 165,000
10	Percent Down	20%
11	Down Payment Amount	$ 33,000
12	Loan Amount	$ 132,000
13	Annual Interest Rate	5.0%
14	Years to Pay	30
15	**Monthly Payment**	**$708.60**
16		

Figure 2.41c Completed Loan Payments worksheet

Key Takeaways

- The PMT function can be used to calculate the monthly mortgage payments for a house or the monthly lease payments for a car.
- When using the PMT function, each argument must be separated by a comma.
- To calculate the monthly payment for a loan using the PMT function, the Rate and Nper arguments must be defined in terms of months. The Rate should be divided by 12 to convert it from an annual rate to a monthly rate. The Nper should be multiplied by 12 to convert the term of the loan from years to months.

- The PMT function produces a negative output if the Pv argument is not preceded by a minus sign. For the purposes of this textbook, a minus sign will be entered before the PV argument in the PMT dialog box.

Attribution

Adapted by Mary Schatz from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

4.XLSX.4 PREPARING TO PRINT

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Learning Objectives

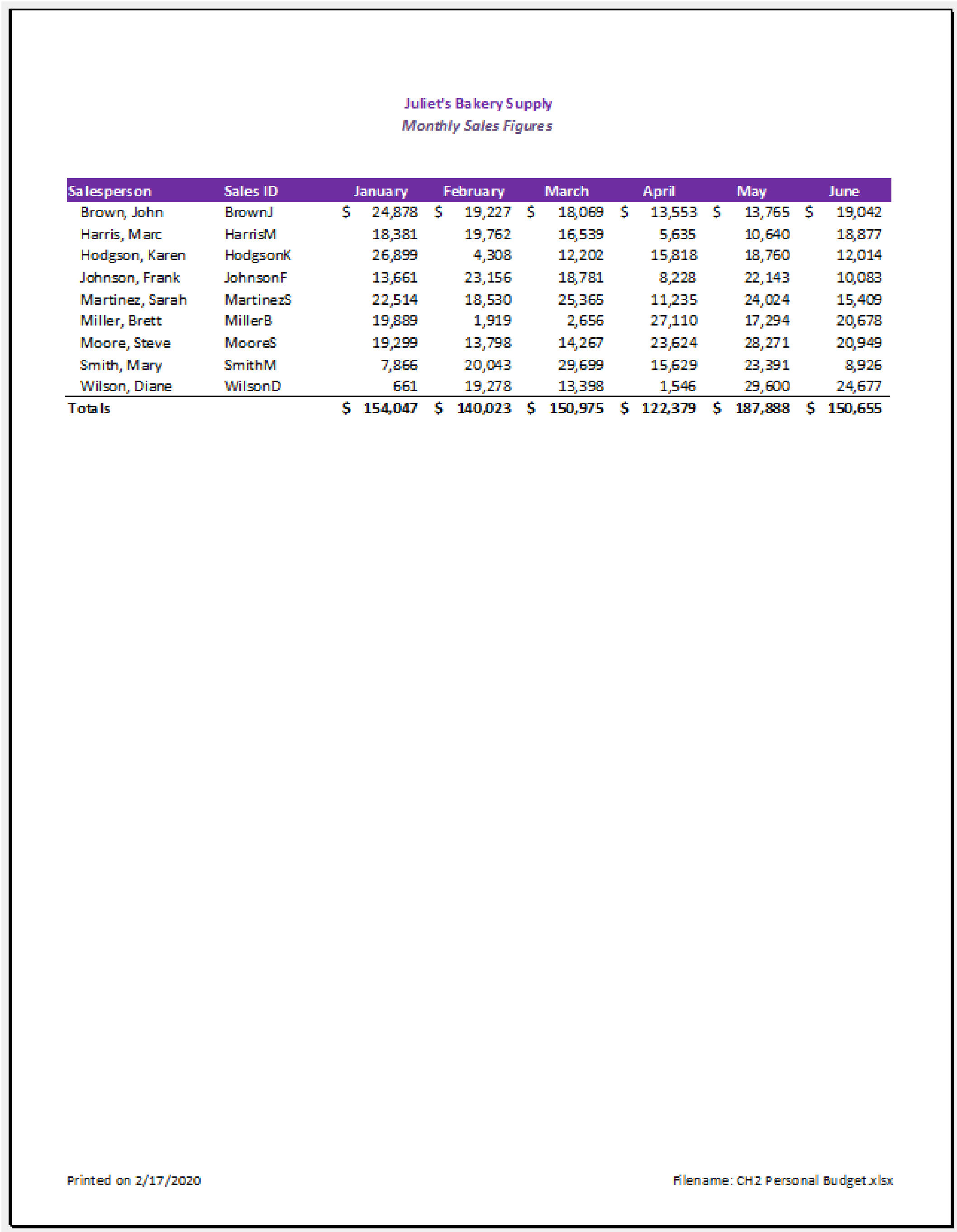
1. Review and learn new cell formatting techniques.
2. Understand how to modify page scaling and margins.
3. Create custom headers and footers to automatically update information.

In this section, we will review some of the formatting techniques covered in Chapter 1, as well as learn some new techniques. We will also preview a two-page worksheet and set page setup options to present the data in a professional manner. A new data file will be used in this section.

Formatting Worksheet Data

Data File: Continue working with CH2 Personal Budget

You have been given sales data that needs to be formatted in a professional manner. This worksheet will be printed and presented to investors, so it needs to be prepared for printing as well. **Figure 2.42** shows how the finished worksheet will appear in Print Preview.



Juliet's Bakery Supply

Monthly Sales Figures

Salesperson	Sales ID	January	February	March	April	May	June
Brown, John	BrownJ	$ 24,878	$ 19,227	$ 18,069	$ 13,553	$ 13,765	$ 19,042
Harris, Marc	HarrisM	18,381	19,762	16,539	5,635	10,640	18,877
Hodgson, Karen	HodgsonK	26,899	4,308	12,202	15,818	18,760	12,014
Johnson, Frank	JohnsonF	13,661	23,156	18,781	8,228	22,143	10,083
Martinez, Sarah	MartinezS	22,514	18,530	25,365	11,235	24,024	15,409
Miller, Brett	MillerB	19,889	1,919	2,656	27,110	17,294	20,678
Moore, Steve	MooreS	19,299	13,798	14,267	23,624	28,271	20,949
Smith, Mary	SmithM	7,866	20,043	29,699	15,629	23,391	8,926
Wilson, Diane	WilsonD	661	19,278	13,398	1,546	29,600	24,677
Totals		**$ 154,047**	**$ 140,023**	**$ 150,975**	**$ 122,379**	**$ 187,888**	**$ 150,655**

Printed on 2/17/2020

Filename: CH2 Personal Budget.xlsx

Figure 2.42 Completed Prepare to Print worksheet

1. Switch to the **Prepare to Print** worksheet.

1. To change the font of the entire worksheet, click the Select All button in the top left corner of the worksheet grid (see **Figure 2.43**).

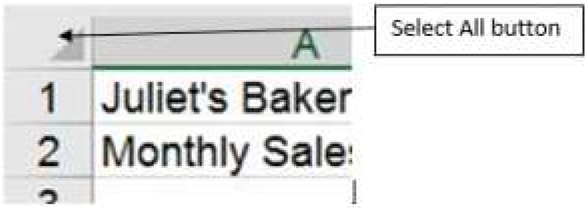


Figure 2.43
Select All button

1. Change the font to Calibri, Size 12.
2. Using the skills learned in Chapter 1, make the following formatting changes:
   1. A1:H1 – Merge and Center; format text as bold and apply a font color and size of your choice
   2. A2:H2 – Merge and Center; format text as bold and italic, apply a font color of your choice
   3. A5:H5 – Apply a dark fill color; format text as white and bold
   4. C5:H5 – Center align
   5. A15:H15 – Apply Top Border to the cells; format text as bold
   6. C6:H6 and C15:H15 – Apply Accounting Number format with 0 decimal places
   7. C7:H14 – Apply Comma style with 0 decimal places
   8. Highlight A6:A14 (salespeople's names) and click the Increase Indent button in the Alignment group on the Home ribbon (see **Figure 2.44**). This will indent the text from the cell border.

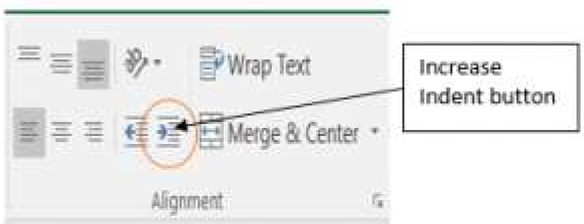


Figure 2.44
Increase Indent button

Using Page Setup Options

Once the worksheet is professionally formatted, you need to look in Print Preview to see how the pages will print.

1. Go to Backstage View by clicking the File tab on the ribbon. Select Print from the menu. Notice that the worksheet is currently printing on two pages, with the page breaking between the April and May columns. To fix this problem, you will first change the left and right margins while still in Print Preview

   Mac Users should click the **File menu option** and select **Print** from the menu
2. Click the Margins drop-down arrow in the Settings section (see **Figure 2.45**)
3. Select Custom Margins... at the bottom of the list.

   Mac Users should select **"Manage Custom Margins"**

4. Type in **0.5** for the Left Margin and **0.5** for the Right Margin.
5. Click OK. Changing the margins brought the May column onto the same page, but the June column is still on a separate page. Next you will use Page Scaling to fix this while still in Print Preview.
6. Click the Scaling drop-down arrow in the Settings section (**Figure 2.46**).

   Mac Users: there is no “Scaling drop-down arrow”. Just click the checkbox for **“Scale to fit”**
7. Select Fit All Columns on One Page.
8. Exit Backstage View.

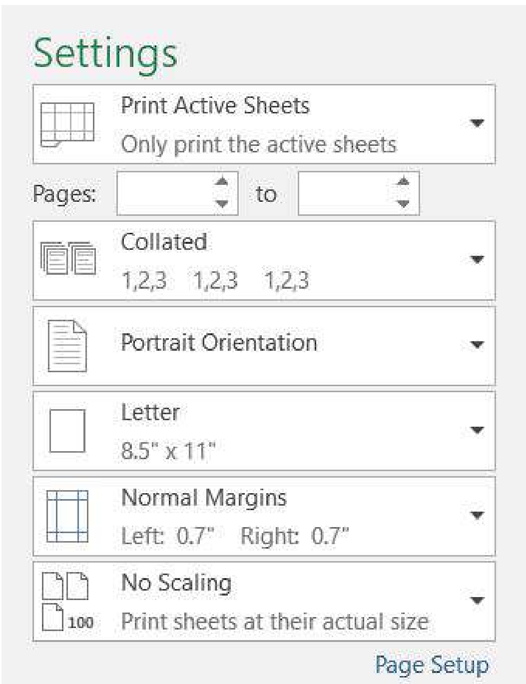


Figure 2.46
Settings section of
Print Preview

Creating a Footer using Page Setup

Now that the entire worksheet is printing on one page, you need to add a footer with information about the date the file was printed along with the filename. In Chapter 1 you learned how to create headers and footers using the Insert ribbon. You can also create headers and footers using the Custom Header/Footer dialog box.

1. Click the **Page Layout** tab on the ribbon.
2. Click the dialog box launcher in the Page Setup group. A window similar to **Figure 2.47** should appear.

   Mac Users: there is no “dialog box launcher”. Just click the **Page Setup button** 

   and continue with Step 3 below.

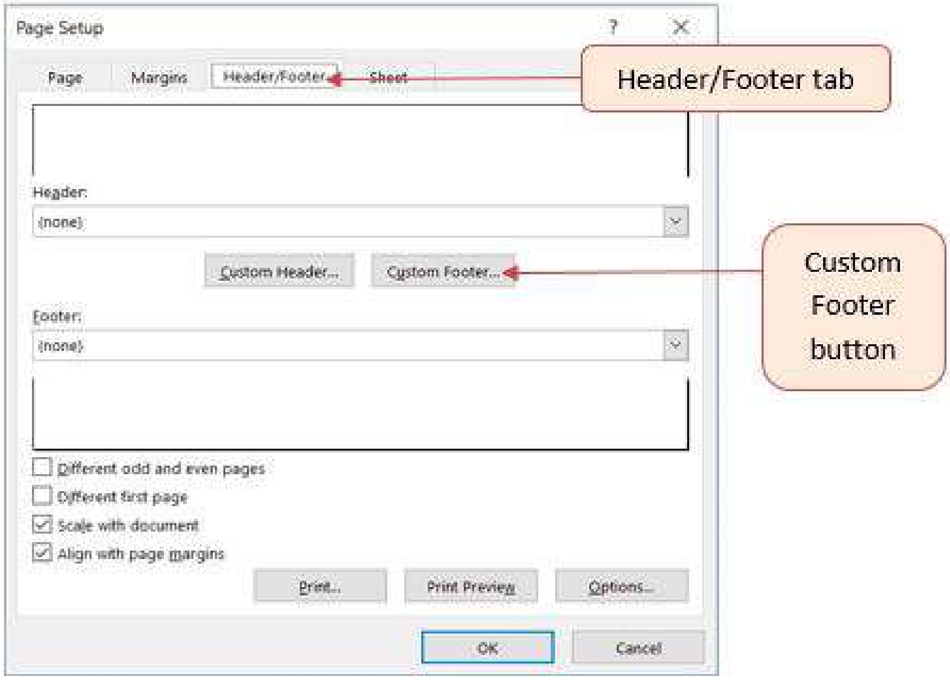


Figure 2.47 Page Setup Dialog Box

3. Click the Header/Footer tab in the Page Setup dialog box.
4. Click the Custom Footer button. The Footer dialog box should appear (see **Figure 2.48**).

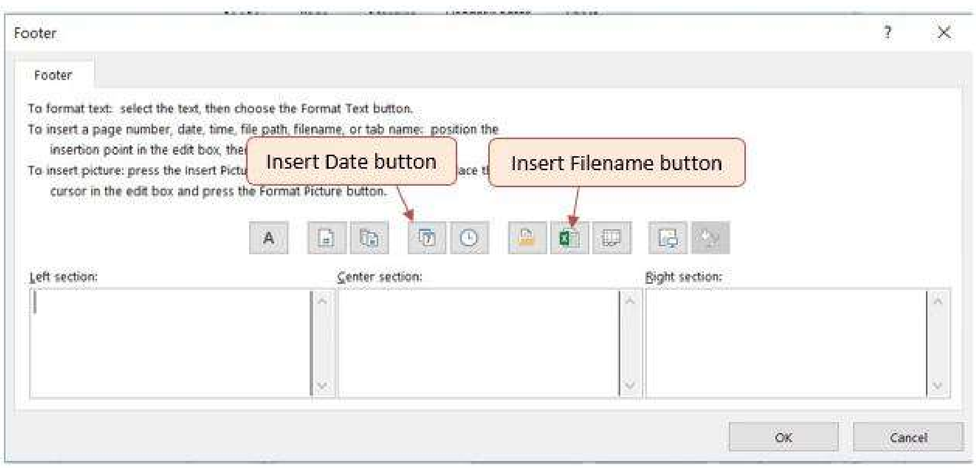


Figure 2.48 Footer Dialog Box

5. Click in the Left section: box and type **Printed on.**
6. Making sure to leave a space after the word on, click the Insert Date button.
7. Click in the Right section: box and type **Filename:.**
8. Making sure to leave a space after the colon, click the Insert File Name button.
9. The Footer dialog box should look like **Figure 2.49**.
10. Click the OK button. Click OK again to close the Page Setup dialog box.
11. Go to Print Preview to see that the current date and file name are displayed in the footer.
12. Exit Backstage View. Check the spelling on all of the worksheets and make any necessary changes.

13. Save the **CH2 Personal Budget** file.
14. Compare your work with the completed worksheet shown in **Figure 2.42** and then submit the **CH2 Personal Budget** workbook as directed by your instructor.

Figure 2.49 Completed Custom Footer Dialog Box

Key Takeaways

- It is important to always check your workbooks in Print Preview to ensure that the data is printed in a professional and easy to read manner.
- Adjust margins and page scaling as needed to keep columns of data together on one page if possible.
- Use headers and footers to display information in the top and bottom margins of the printed worksheet. Use the Insert buttons to insert changing information, such as dates and file names, instead of typing them in directly.

Attribution

"2.4 Preparing to Print" by Julie Romey, Portland Community College is licensed under CC BY 4.0

4.XLSX.5 CHAPTER PRACTICE

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Financial Plan for a Lawn Care Business

Download Data File: [PR2-Data]

Running your own lawn care business can be an excellent way to make money over the summer while on break from college. It can also be a way to supplement your existing income for the purpose of saving money for retirement or for a college fund. However, managing the costs of the business will be critical in order for it to be a profitable venture. In this exercise you will create a simple financial plan for a lawn care business by using the skills covered in this chapter.

There are two worksheets in the workbook you will be using.

- **Annual Plan** – provides calculations to determine how much money the lawn care business brings in for one year, based on the average price per lawn cut and the number cut per year, as well as the expenses for the year.
- **Equipment Loans** – calculates the monthly payments for the various lawn care equipment loans.

Annual Plan Worksheet

1. Open the file named **PR2 Data** and then Save As **PR2 Lawn Care**.
2. Switch to the Annual Plan worksheet if needed.
3. Enter the following data into cells B14, B15, and B16:
    - Gasoline cost (per cut) = $10
    - Number of customers = 30
    - Annual lawn cuts per customer = 20
4. In cell B3, enter the average price per lawn cut of $50.
5. In cell B4, write a formula that calculates the total number of lawns cut in the year. *This is the number of customers multiplied by the annual lawn cuts per customer.*
6. In cell B5, write a formula that calculates the total annual sales. *This is found by multiplying the average price per lawn by the total number of lawn cuts.*

7. In cell B8, write a formula to calculate the total cost of gasoline for the year. *This is found by multiplying the gasoline cost per cut by the total number of lawns cut.*
8. You will finish the rest of this worksheet after completing the Equipment Loans worksheet.

Equipment Loans Worksheet

1. Switch to the Equipment Loans worksheet.
2. In cell E3, write a PMT function to calculate the monthly payment for the Commercial Lawn Mower. *Don't forget the negative sign in between the equal sign and the PMT! Remember to convert the interest rate and years to monthly terms and to use cell references.* The arguments of the PMT function should be as follows:
   - RATE: B3/12
   - NPER: C3*12
   - PV: D3
3. Copy the PMT function from cell E3 to the other equipment items.
4. In cell E10, use the SUM function to calculate the total for the monthly loan payments. Make sure that the blank rows (7 through 9) were included in the range for the SUM function so that you can add more equipment items later if needed.
5. In cell E11, write a formula that calculates the total **annual** loan payments. *This will be the monthly total multiplied by 12 (the number of months in a year).*
6. If needed, apply Accounting format to all of the monetary values so that the placement of the dollar sign is consistent throughout the worksheet.
7. Sort the data in the range A3:E6 first by **Interest Rate** and then by **Loan Amount** using the following steps:
   - Select the range A3:E6.
   - Click the Sort button in the Data tab of the Ribbon.
   - In the Sort dialog box, select the Interest Rate option in the "Sort by" drop-down box. Select Largest to Smallest for the sort order.
   - Click the Add Level button in the Sort dialog box.
   - Select the Loan Amount option in the "Then by" drop-down box. Select Largest to Smallest for the sort order.
   - Click the OK button in the Sort dialog box.
8. Add a header with the date on the left and the worksheet name on the right. Be sure to insert the date

and worksheet name so that they will automatically update.

9. Check Print Preview and make any other changes necessary for professional printing.

Complete the Annual Plan Worksheet

1. Switch back to the Annual Plan worksheet.
2. In cell B9, write a formula that displays the annual monthly payments total from cell E11 in the Equipment Loans worksheet using the following steps:
    - Type an equal sign =
    - Click the **Equipment Loans** worksheet
    - Click cell E11
    - Press the ENTER key
3. In cell B10, calculate the Total Expenses by adding the Gasoline Cost and the Annual Equipment Payments.
4. In cell B12, calculate the Annual Profit by finding the difference between the Total Annual Sales and the Total Expenses. *Hint: This will hopefully be a positive number which shows that your business is making money instead of losing money.*
5. Format all cells that contain monetary amounts in the Annual Plan worksheet for Accounting Number Format ($) with no decimal places. Format all other numerical values as Comma format with no decimal places.
6. Add a header with the date on the left and the worksheet name on the right. Be sure to insert the date and worksheet name so that they will automatically update.
7. Check Print Preview and make any changes necessary for professional printing.
8. Check the spelling on all of the worksheets and make any necessary changes.

Compare both worksheets with the answer keys below.

10/3/2020 Annual Plan

Lawn Care Annual Financial Plan	
Annual Sales Plan	
Average Price per Lawn Cut	$ 50
Total Number of Lawns Cut	600
Total Annual Sales	$ 30,000
Annual Expenses	
Gasoline	$ 6,000
Annual Equipment Payments	$ 2,321
Total Expenses	$ 8,321
Annual Profit	$ 21,679
Gasoline Cost (per cut)	$ 10
Number of Customers	30
Annual Lawn Cuts per Customer	20

1/30/2020 Equipment Loans

Equipment Purchasing Plan				
Item	Interest Rate	Years	Loan Amount	Monthly Payment
Blower	6.0%	3	$ 700	$ 21.30
Commercial Lawn Mower	5.5%	4	$ 6,000	$ 139.54
Edger	4.0%	2	$ 400	$ 17.37
Trimmer	4.0%	2	$ 350	$ 15.20
			Monthly Total	$ 193.40
			Annual Total	$ 2,320.83

Attribution

Adapted by Mary Schatz from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

4.XLSX.6 CHAPTER SCORED

Noreen Brown; Barbara Lave; Hallie Puncochar; Julie Romey; Mary Schatz; Art Schneider; Diane Shingledecker; and Jennifer Evans

Hotel Occupancy and Expenses

Download Data File: [SC2-Data]

The hotel management industry presents a wide variety of career opportunities. These range from running a bed and breakfast to a management position at a large hotel. No matter what hotel management career you choose to pursue, understanding hotel occupancy and costs are critical to running a successful operation. This exercise examines the occupancy rate and expenses of a small hotel.

There are three worksheets in the workbook for this assignment.

- **Occupancy** – calculates and displays the maximum **hotel capacity** for each month (based on the number of rooms, the capacity of each room, and the number of days in the specific month), the **actual occupancy** (how many actually stayed in the hotel that month), and the **occupancy percentage** (what percentage of capacity was the hotel each month).
- **Statistics** – calculates and displays the highest, lowest, and average actual occupancy and occupancy percentages form the Occupancy worksheet.
- **Shuttle Purchase** – calculates three different down payment options for a loan to purchase a shuttle for the hotel.

Occupancy Worksheet

1. Open the file named **SC2 Data** and then Save As **SC2 Hotel**.
2. Switch to the Occupancy worksheet if needed.
3. Replace the [Insert Year] in A1 with the year number for last year.
4. You need to calculate the January capacity for the hotel in C5. The capacity shows how many people the hotel can hold during the month. It is calculated by first multiplying the *occupants per room* by the *number of rooms* in the hotel. This **result** is then multiplied by the number of days in the month (cell B5 for January). Create this formula using **absolute references** so that the appropriate cells do not change when the formula is pasted throughout column C. *Hint: two of the cells in the formula need to be absolute references.*
5. Copy the formula in cell C5 and paste it into the range C6:C16. Use a paste method that does not

remove the border at the bottom of cell C16.

6. Format the numbers in columns C and D for comma format with zero decimal places.
7. In cell C17, enter a function that finds the sum of the monthly hotel capacity values. Do the same in cell D17 to find the sum of the monthly actual occupancy values.
8. Enter a formula in cell E5 to calculate the Percent Occupied of the hotel (this statistic shows what percentage of the hotel is full or occupied). Your formula should divide the Actual Occupancy by the Hotel Capacity. Then copy and paste the formula into the range E6:E17. Use a paste method that does not remove the borders at the bottom of cell E16 and E17. Format the results in E5:E17 as *percentages with two decimal places*.
9. Format the Totals (C17:E17) as bold.
10. Apply any number formatting that aids in the readability and professionalism of the worksheet.
11. Check Print Preview and make any changes necessary for professional printing.

Statistics Worksheet

1. Replace the [Insert Year] in A1 with the year number for last year.
2. Enter a **function** in cell B3 that finds the highest value in the Actual Occupancy column from the Occupancy worksheet.
3. Enter a **function** in cell B4 that finds the lowest value in the Actual Occupancy column from the Occupancy worksheet.
4. Enter a **function** in cell B5 that shows the average value of the Actual Occupancy column on the Occupancy worksheet.
5. Use the Auto Fill handle to copy the formulas in the range B3:B5 to the range C3:C5.
6. Apply any number formatting that aids in the readability and professionalism of the worksheet. The numbers should be formatted similarly to the Occupancy worksheet.
7. Check Print Preview and make any changes necessary for professional printing.

Shuttle Purchase Worksheet

The hotel is considering buying a car to shuttle customers to and from the airport. You need to decide how much of a down payment to make, so you are going to calculate the monthly payment based on three different down payment percentages. The number of years to pay off the loan will vary for each of the down payment percentage options. Remember, the down payment amount is found by multiplying the price of the car by the down payment percentage. This amount is then subtracted from the price of the car to find the amount of the loan.

1. In cell B5 write a formula that will calculate the **amount of the down payment**. *Be sure to use cell references as much as possible.* Copy the formula to the other down payment options.

2. In cell B6 write a formula to calculate the **amount of the loan**. *Be sure to use cell references as much as possible.* Copy the formula to the other down payment options.
3. In cell B9 create a PMT function to calculate the **monthly payment.** Make sure the arguments in the PMT function are converted into months and that the monthly payment is a positive number. *Be sure to use cell references as much as possible.* Copy the function to the other down payment options.
4. In cell B10 create a formula that calculates **how much will be paid in total** for the vehicle, including the down payment and the total amount paid on the loan in the given number of years. Copy the formula to the other down payment options.
5. In cell A12, write an explanation of which down payment option is the best and why.
6. Apply any number formatting that aids in the readability and professionalism of the worksheet.
7. Make the following page setup changes to the worksheet:
    - Center horizontally on the page
    - Create a footer with the date on the left and the file name on the right. Make sure that both the date and the file name will update automatically.
8. Check the spelling on all of the worksheets and make any necessary changes.
9. Save the **SC2 Hotel** workbook and submit the workbook as directed by your instructor.

Attribution

Adapted by Mary Schatz from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

SECTION VIII

5. DESCRIBING, EXPLORING, AND COMPARING DATA