Achieve

Statistics

Study Guide

2nd Edition
1/5/2019

Acknowledgements

We would like to thank the authors for their patience, support, and expertise in contributing to this study guide; and the editors for their invaluable efforts in reading and editing the text. We would also like to thank those at Achieve Test Prep whose hard work and dedication to fulfilling this project did not go unnoticed. Lastly, we would like to thank the Achieve Test prep students who have contributed to the growth of these materials over the years.

Printed in the United States of America

First Printing, 2012

Achieve
PO Box 10188 #29831
Newark, NJ 07101-3188

Tel: 888.900.8380

Visit the Achieve Test Prep website at http://www.achievetestprep.com/student

Table of Contents

Statistics

Chapter 1: An Introduction to Statistics

Statistics is used in many applications. Statistical methods are often used to describe and study a population, drug therapies, research, economics, and ecosystems, just to name a few of the many areas that encompass statistics and statistical applications. As a student in statistics, you will learn how to organize, define, describe, and interpret data. This section will begin with an overview of statistics and an explanation of commonly used statistical terms, calculations, and applications.

Learning Objectives

After reading Chapter 1 and completing the workbook, you should be able to:

1. Identify the difference between quantitative and qualitative statistics.
2. Identify the difference between differential and inferential statistics.
3. Define basic statistical terms.
4. Define mean, median, mode, and range.
5. Calculate mean, median, mode, and range.
6. Apply the basic statistical concepts to data interpretation.

Study Clues

A clear understanding of the basic statistical terms and concepts presented in Chapter 1 will help prepare you to advance in this course and learn more complex statistical calculations and specific statistical tests. As you study, you should pay particular attention to the definitions and how to calculate each of the following: mean, median, mode, and range. Your exam will be multiple choice, so you must make sure you can correctly calculate accurately as no partial points are given.

1.1 Basic Math Review

The purpose of this pre-chapter is to offer a basic review of many important mathematical functions that you will be required to know for the statistics exam.

Signs and Symbols

On your exam, basic mathematical functions will be represented by symbols. Here we will cover the basic symbols you will encounter on the exam.

Symbol	√	*, ×, ·	÷, /	<	>	≠
Meaning	Square Roots	Multiplication	Division	Greater Than	Less Than	Not Equal

Order of Operations

In math, order is everything! There is a unique order of operations we use to solve all mathematical equations. The order of operations (sometimes called operator precedence) is a rule used to clarify which procedures should be performed first in a given mathematical expression.

The order of operations--or precedence--is used throughout mathematics, science, technology, and computer programming, and is expressed here. It states the order in which problems should be solved:

1. Terms inside parentheses or brackets
2. Exponents and roots
3. Multiplication and division
4. Addition and subtraction

This means that if a mathematical expression is preceded by one operator and followed by another, the operator higher on the list should be applied first.

Examples

- $(1 - 3) + 7 = - 2 + 7 = 5$
- $(2\times3) + (4\times1) = 6 + 4 = 10$
- $(4\div2)-(3-2)\times2 = 2-1\times2 = 2$

Always remember to perform the functions inside parenthesis first then read the problem left to right to complete it.

1.2 Algebra

For your exam, you will have to apply the concepts of elementary algebra. This is the most basic form of algebra. In arithmetic, only numbers and their arithmetical operations (such as $+, -, \times, \div$) occur. In algebra, numbers are often denoted by symbols (such as a, n, x, y, or z). This is useful for the following reasons:

- It allows the general formulation of arithmetical laws (such as $a + b = b + a$ for all a and b)
- It allows the reference to "unknown" numbers, the formulation of equations and the study of how to solve these (for instance, "Find a number x such that $3x + 1 = 10$" or going a bit further "Find a number x such that $ax + b = c$")
- It allows the formulation of functional relationships. (For instance, "If you sell x tickets, then your profit will be $3x - 10$ dollars, or $f(x) = 3x - 10$, where f is the function, and x is the number to which the function is applied.")

Solving a Linear Equation

$3x - 6 = 0$	Given equation
$3x - 6 + 6 = 0 + 6$ $3x = 6$	Add 6 to both sides Combine like terms (-6+6) on left side and (0+6) on right side
$\frac{3x}{3} = \frac{6}{3}$	Divide both sides by 3
$x = 2$	

After solving an equation, you should check each solution in the original equation. In the above example, check that 2 is a solution by substituting 2 for x in the original equation.

Evaluating Expressions

Evaluate $2x + 3$ if $x = 3$.

$2(3) + 3$ $6 + 3$ 9	Replace the value of x with 3 then evaluate the expression according to the order of operations.

1.3 Exponents

Repeated multiplications can be written in exponential form.

Repeated Multiplication	Exponential Form
$2\times2\times2$	2^3
$(5)(5)(5)(5)$	5^4
$(-4)(-4)(-4)$	$(-4)^3$
$(2x)(2x)(2x)(2x)$	$(2x)^4$

Properties of Exponents

Let a and b be real numbers, variables, or algebraic expressions, and let m and n be integers. (Assume all denominators and bases are nonzero.)

Property	Example
Add exponents when multiplying.	$5^2 \times 5^4 = 5^{2+4} = 5^6 = 15626$
Subtract exponents when dividing.	$5^4 \div 5^2 = 5^{4-2} = 5^2 = 25$

1.4 Basic Terms

- **Statistician**: Is someone who specializes in the field of statistics. It is often the job of the statistician to develop experimental designs, organize and analyze data, and generate graphical interpretations of the data. Statisticians are often hired by hospitals, pharmaceuticals, universities, insurance companies, and government agencies.

- **Quantitative**: An objective measurement based on numerical values of a given data set or population. Examples of quantitative data are the average age of a population or the number of male students in a given class.

- **Qualitative**: A subjective measurement based on opinion and non-numerical values. Examples of qualitative measurements would be color preference. "I prefer the color red to the color blue" is a qualitative assessment based on an individual's opinion or preference. It does not hold any numerical value.

There are two types of statistics: descriptive and inferential statistics. Each plays an important unique role in the final interpretation of the data set.

- **Descriptive statistics**: Uses quantitative measurements to objectively describe a data set. For example, descriptive statistics can use numerical values collected to describe the average age of a given population; or the success rate of a trial clinical therapy.

- **Inferential statistics**: Uses qualitative measurements to make subjective interpretations about a given group. For example, if you were to ask a single class of college freshmen what their favorite color was and 75% of the class responded that their favorite color was blue, we could infer that the majority of college freshmen prefer the color blue. We are inferring this single observation to an entire population but do not have data for the entire population. Therefore, our inferential statistics is subjective and may change as we survey more college freshmen in additional classes.

**Tip: Descriptive statistics tends to yield a more solid interpretation of the entire population. Inferential statistics tend to be subjective to change.*

1.5 Measurements

Mean: The most general definition of mean is the calculated average of the given population or set of values. The mean, or average, can yield useful information about a population or given data set. From this calculation, we can determine the average age or average response. The formula for calculating the mean is:

$$Mean = \frac{Sum\ of\ observations}{Number\ of\ observations}$$

Let us look at an example! We are given the following data set of student ages in the general statistics class. You are asked to calculate the mean of our given population.

Student ages: 31, 33, 30, 31, 35, 33, 36, 28, 42, 37, 33

For the 1st step, we need to find the sum of the observations. To do this, we simply add all the ages together.

$$31 + 33 + 30 + 31 + 35 + 33 + 36 + 28 + 42 + 37 + 33 = 369$$

369 is the sum of all the observations, or the sum total of all of the student's ages in our example population. Right now this number does not tell us a lot of information, so we need to move on to step 2.

For the 2nd step, we need to determine the number of observations in the data set. For our example, we have 11 students in the class, so 11 is our number of observations. Now, we can solve for the mean by taking 369 / 11 which gives us an average class age of 33.5 years you could round this number up to 34 years.

**Important Points!*

- Many times, you will be asked to calculate the average. Remember, the average is the same as the mean.
- The number of observations can also be written as n. The letter n is just shorthand to denote the number of observations in a given data set or population.
- The / symbol stands for division.

**Questions to think about: Why is it important to study the mean of a population? What useful information can the mean give use about a given population?*

Median: The median is the middle value or number to a given set of numbers placed in order from smallest to largest. The median can be used to separate the data set into lower and upper values. The median is easily identified in number sets with an un-even amount of values. For example, in a number set with 15 values, the median can be identified by counting equally from each end. The median would be the 8th value. In even number sets, the median is calculated by adding the two middle values and dividing by two.

Let us look at our previous data set to determine the median.

Student ages: 31, 33, 30, 31, 35, 33, 36, 28, 42, 37, 33

The 1st step is to place the data set in order from smallest value to the largest value.

28, 30, 31, 31, 33, 33, 33, 35, 36, 37, 42

**Tip: Sometimes it is helpful to cross out the numbers as you place them in order – always go back and count to make sure you have included all of the values!*

The 2nd step is to determine which value is the middle value. You can easily do this by counting evenly from both ends.

----------------------- -----------------------
28, 30, 31, 31, 33, 33, 33, 35, 36, 37, 42

For our data set, we have 11 values so we can count five from each end. Our middle value, or median, is 33. The works for all data sets with an odd number of observations. But, what if we have an even number of observations? Let us look at the following data set:

28, 30, 31, 31, 33, 33, 33, 35, 36, 37, 38, 42

Now we have 12 observations, so there is no one middle value. In this example, we need to take the average of the 2 middle values. As in the number set above, we count evenly from both sides.

----------------------- -----------------------
28, 30, 31, 31, 33, 33, 33, 35, 36, 37, 38, 42

For this data set, we count five from each end. We then add the two middle values and divide by 2. This calculates the average of the middle values. For this example, the median would be $\frac{33+33}{2} = 33$.

Mode: The mode is the most commonly occurring number in a given data set. The mode can give important information about the randomness of a given data set, and therefore the strength of the experimental design, which we will cover latter in the text. It is important to note that a data set can have more than one mode. Let us look at our original data set:

Student ages: 31, 33, 30, 31, 35, 33, 36, 28, 42, 37, 33

The 1st step is to order the data set, just like we did for the median. This allows you to better visualize and identify repeating numbers. Our ordered data set is:

28, 30, 31, 31, 33, 33, 33, 35, 36, 37, 42

The 2nd step is to identify repeating numbers.

28, 30, 31, 31, 33, 33, 33, 35, 36, 37, 42

From the above data set, we have two sets of repeating numbers; the age 31 occurs twice and the age 33 occurs three times. So, the mode for the given data set is 33, because it occurs most often. We can interpret this as saying 3 out of 11 students is 33 years old.

But what if the data set changes to the following?

28, 30, 31, 31, 31, 33, 33, 33, 35, 36, 37, 42

Now we have three students who are 31 and three who are 33. Therefore, we have two modes and we would say that the mode of our population is 31 and 33. Remember, you can always have more than one mode if two sets of numbers appear an equal time in a given data set.

Let us look at one more example:

28, 30, 31, 33, 35, 36, 37, 42

In this example, there are no repeating ages. Therefore, this data set does not have a mode. A data set can only have a mode when repeating numbers are observed.

**Tip: A data set can have 1 mode, more than 1 mode, or no modes!*

Range: The range is the interval between the lowest and highest values, of a given number set place in numerical order. The range can give us useful information about a given population or data set.

Let's look at our original data set:

Student ages: 31, 33, 30, 31, 35, 33, 36, 28, 42, 37, 33

The 1st step is to order the data set, just like we did previously. This allows you to better visualize and identify the lowest and highest values.

Our ordered data set is: 28, 30, 31, 31, 33, 33, 33, 35, 36, 37, 42

From the above example, we can identify the lowest or youngest age is 28 and the highest, or oldest age, is 42. Therefore, we would say that our population range is 28 years to 42 years. The rest of the class ages fall between 28 and 42 years.

But what if we have the following age of students?

Student ages: 28, 28, 28, 28, 28, 28, 28, 28, 28, 28, 28

Now all 11 students are 28 years old; therefore, we do not have a range in this given population example.

Let us look at another example of student ages for a given class:

Student ages: 20, 21, 22, 22, 23, 24, 24, 59

We identify the lowest or youngest age to be 20 and the highest or oldest age to be 59. Therefore, our class age range is 20 to 59 years. This appears to give us a large range in student ages. However, we can clearly see that the majority of the students are in their early twenties, with a single student being 59 year old. In this case, the range alone does not yield a true representation of the population. We can say that the range is skewed due to an outlier, the 59 year old student. We will discuss outliers latter on in the text, but keep this concept in mind as you progress to data interpretation.

**Questions to think about: Why is it important to study the range of a population? What useful information can the range give use about a given population? How can the results skew the actual population sample?*

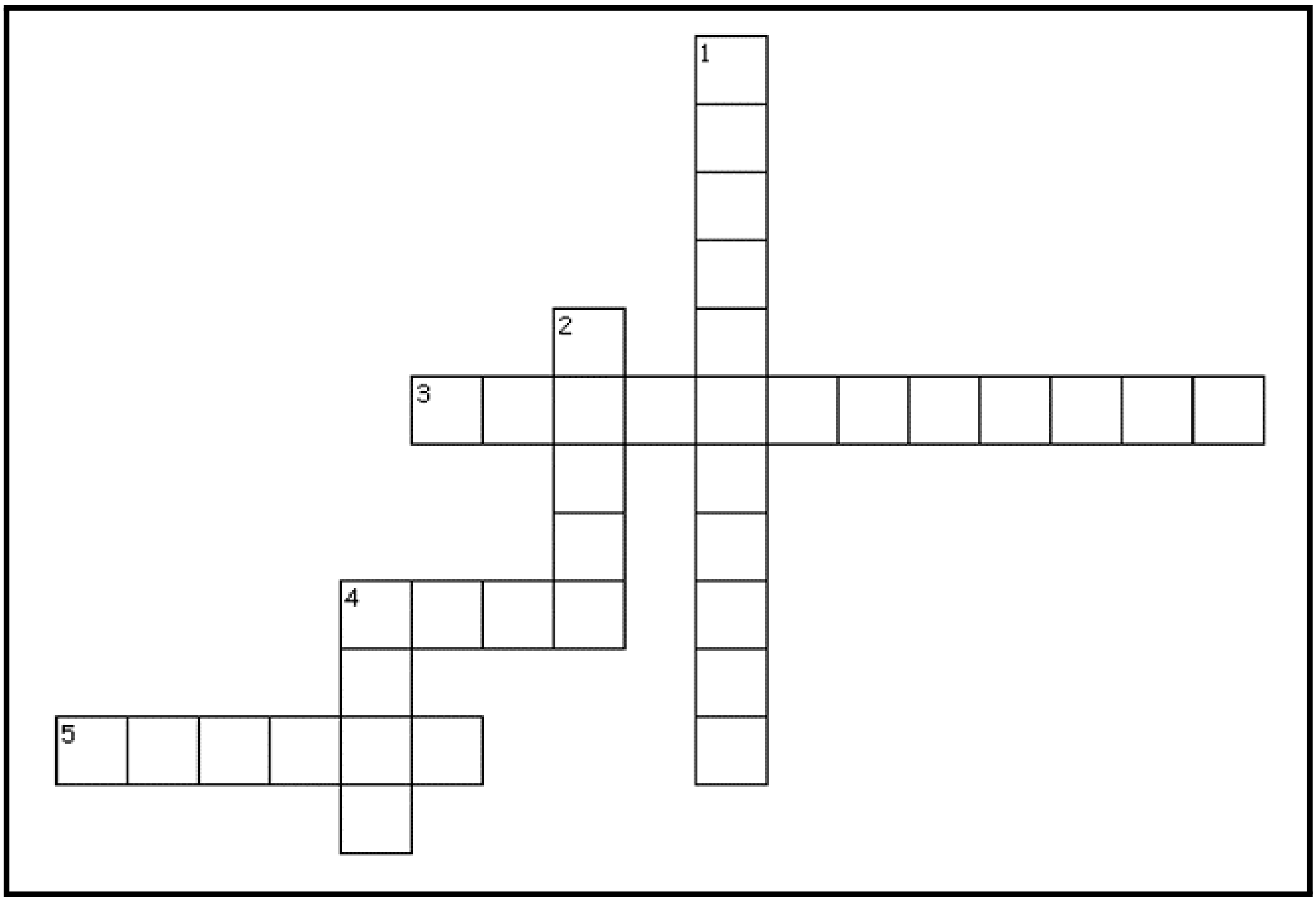
Chapter 1 Review

Below is an outline of the major points covered in Chapter 1. You will need to clearly understand the concepts, terms and calculation presented in Chapter 1. You may need to refer back to Chapter 1 as we progress through the text. All concepts build from each other.

You should be able to …

Define	Calculate	Understand
Quantitative	Mean	What the mean, median, mode,
Qualitative	Median	and range are used for, and any
Descriptive Statistics	Mode	special circumstances
Inferential Statistics	Range	discussed.

Ch. 1 Crossword Puzzle



ACROSS
3. An objective measurement
4. The most frequently occurring number
5. The middle number

DOWN
1. A subjective measurement
2. Separates the data set into high and low
4. Known as the average

Chapter 1 Practice Problems

Complete the following practice problems to test your understanding. Check your work with the solutions provided.

You are given the following data set of patient weights in pounds from a clinical trial:

180, 155, 202, 169, 135, 189, 220, 179, 122, 169

Use this data set to answer the following questions:

1. What is the mean weight of the patients?

2. What is the median weight of the patients?

3. Does this data set have a mode? If so, what is the mode?

4. What is the range of the patient's weight?

Answer key is found in the Answer Keys section.

Chapter 1 Quiz

1. John is 22 years old, Stacy is 33 years old, and Jeff is 42 years old. What is the average age of John, Stacy, and Jeff?
    a. 97
    b. 33
    c. 32
    d. 25

2. Collected numerical values of a classes average exam score is an example of:
    a. Descriptive Statistics
    b. Inferential Statistics
    c. Descriptive Statistics and Quantitative Measurements
    d. Descriptive Statistics and Qualitative Measurements

3. Given the following exam scores, find the median: 75, 69, 82, 93, 97, 85, 79.
    a. 82
    b. 79
    c. 84
    d. 91

4. Using the same data set, find the mode: 75, 69, 82, 93, 97, 85, 79.
    a. The mode would be the same as the median
    b. The data set does not have a mode
    c. All values
    d. The mode would be the same as the average

5. Using the same data set above, find the range: 75, 69, 82, 93, 97, 85, 79.
    a. 75 to 79
    b. 69 to 97
    c. 97 to 69
    d. 82 to 97

Answer key is found in the Answer Keys section.

Chapter 2: Summarizing, Organizing, and Describing Data

Often, you will encounter statistical data that has been summarized and organized into graphs or tables. This is done to allow for the proper description of a given data set. As a student in statistics, you will encounter several different types of data organizational methods and will have to apply your statistical knowledge to interpret the given data set. This section will begin with a description of common ways to organize and summarize statistical data.

Learning Objectives

After reading Chapter 2 and completing the workbook, you should be able to:

1. Know the two types of data.
2. Know how to organize data into a graphical, chart, and table.
3. Know how to interpret graphs, charts, and tables.
4. Know how to apply basic statistical calculations to the data found in graphs, charts, and tables.
5. Apply the basic statistical concepts and understanding of graphs, charts, and tables to data interpretation.

Study Clues

A clear understanding of the basic statistical terms and concepts presented in Chapter 2 will help to prepare you to advance in this course and learn more complex statistical calculations and specific statistical tests. As you study, you should pay particular attention to the types of graphs and charts presented. You should pay particular attention to the stem-and-leaf plot and understand how to develop and interpret the data found in the stem-and-leaf plot. You should also understand the basic concepts of a histogram and how to interpret graphically represented data.

2.1 Basic Terms

Types of Data

- Quantitative: An objective measurement based real numbers as discussed in Chapter 1. Quantitative data can be calculated.
- Categorical: Often referred to as qualitative and the data represents specific categories that are not associated with real numbers. For example, male versus female; tall versus short – these are categories that tell us important information about the data set. However, it does not affiliate or assign any numerical value.

Types of Plots, Graphs, and Charts

- Stem-and-leaf plots: Organize the data based on the properties of real-numbers.

Let's say you have the weights of 20 individuals, you need an easy way to represent the data.

Given the following weights (in pounds):

110, 234, 101, 100, 245, 198, 173, 165, 210, 205, 166, 167, 188, 182, 183, 185, 145, 122, 222, 155.

Our first step is to organize the data based on the power of 10 to spate the data into a stem and leaf portion.

The 1st step is to organize the data from least to greatest:

100, 101, 110, 122, 145, 155, 165, 166, 167, 173, 182, 183, 188, 198, 205, 210, 234, 245

The 2nd step is to divide the data into a stem and a leaf. Our leaf should be a single value, which is the last digit in a number. If that does not make sense, a good way to visualize this is to think of your numbers as having two parts; for example, 122: Our stem would be 12 and our leaf would be 2. There are two main rules to remember: 1) the leaf can only be a single number, which is the last digit; and 2) every number must be represented.

Let us organize our data set. A normal stem-and-leaf plot will only have a stem and a leaf portion. But for visualization, we have added a 3rd column for the original value.

Original Value	1 0 0	1 0 1	1 1 0	1 2 2	1 4 5	1 5 5	1 6 5	1 6 6	1 6 7	1 7 3	1 8 2	1 8 3	1 8 8	1 9 8	2 0 5	2 1 0	2 3 4	2 4 5
Stem	10	10	11	12	14	15	16	16	16	17	18	18	18	19	20	21	23	24
Leaf	0	1	0	2	5	5	5	6	7	3	2	3	8	8	5	0	4	5

Above, we broke down the data set to represent the stem and leaf portions. Now, we can combine like stems and finish a completed stem-and-leaf plot.

Stem	Leaf
10	0 1
11	0
12	2
13	
14	5
15	5
16	5 6 7
17	3
18	2 3 8
19	8
20	5
21	0
22	
23	4
24	5

Now we have a completed plot. Remember, like stems can be combined. Look at 165, 166, and 167. In the stem-and-leaf plot, it is written as 16 for the stem with 5, 6, and 7 for the leaf. This will give us 165, 166, and 167.

Let us look at a slightly different data set. You are given the exam grades for the statistics class: 88, 88, 82, 93, 94, 94, 94, 99, 100, 71.

First, put the data in order: 71, 82, 88, 88, 93, 94, 94, 99, 100

Now you are asked to represent this data using a stem-to-leaf plot. Remember our rules; the leaf is only a single value which is the last digit in the number and every number must be represented.

7	1
8	2 8 8
9	3 4 4 9
10	0

From the example above, we had two 88s on the exam. Therefore, our stem-and-leaf plot will have two 8s in the leaf portion. Likewise, we have two 94s on the exam and our leaf portion has two 4s after the value 9.

During your exam, you may be given a stem-and-leaf plot and be asked to determine the mean, median, mode, and range. For this example; you are given the following stem-and-leaf plot and you are asked to determine the mean.

5	0 1 1 2
6	6
7	3 4 4 9
8	0

The 1st step is to write out the data set. For this, the stem goes with each of the leaf values.

50, 51, 51, 52, 66, 73, 74, 74, 79, 80.

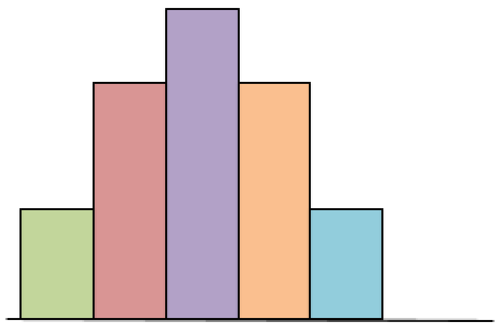
Now we can calculate the mean, this concept was covered in Chapter 1, by adding the sum of the values and dividing by 10.

Our mean is 50+51+51+52+66+73+74+74+79+80 / 10 = 65

**Can you determine the median, mode and range? Be sure to revisit concepts in Chapter 1.*
Remember a stem-and-leaf plot use the entire data set; it is best used for smaller data sets.

- Histograms: A common method to graphically represent the frequencies data set. A histogram can be used for a larger data set. For example, if you have a large population and are gathering statistical exam scores from numerous area colleges, you may have 100 students out of 500 who receive a 94 on an exam. That is a large data set and it is not practical to generate a stem-and-leaf plot. Histograms are usually generated using graphing or statistical software. You can also generate them by hand!

Let us take a look at a histogram:

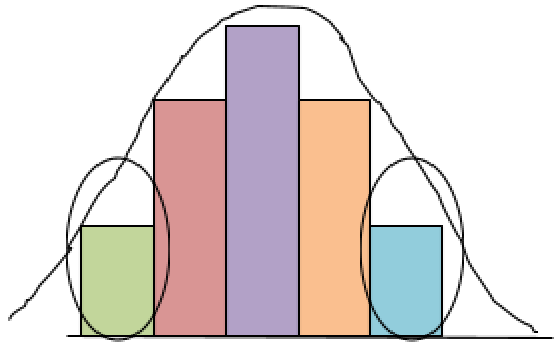


This is the typical structure of a histogram. What type of information can we obtain from a histogram? Although there are not any actual data or numerical values assigned, we can see that there are several parts to the histogram.

- The purple box seems to be the highest. This represents the highest frequency for the data set. Therefore, if we were to look at a data set, the majority of the values would fall within the region of the purple box.
- Now looking to the right and left of our highest frequency, we can see that the histogram is divided into two sides. The green and red boxes are to the left of our highest frequencies. These represent higher values than our frequency. If that is confusing, think of it this way: we will say the data represents the amount of time it takes students to complete an exam. Most of the students take 65 minutes to complete the exam (purple box), the red and green boxes represent the number of students who take less than 65 minutes to complete the exam. We will say red represents students who can complete the exam in 55 minutes and green represents students who can complete the exam in 45 minutes.
- We also have data to the right of our highest frequency, the orange and blue boxes. In our example, these boxes would represent students who took longer than 65 minutes to complete the exam. Let's say the orange box represents students who took 75 minutes while the blue box represents students who took 80 minutes to complete the exam.

How can we interpret this data? We can say that the majority of the students can complete the exam in 65 minutes while a small portion of the student population can complete the exam in 45 minutes and 80 minutes. This information can give us an upper and lower range. 45 minutes would be the fastest completion time, whereas 80 minutes would be the maximum upper time limit for the exam.

When we look at our original histogram:



It has a bell shape to it. The bell shape is drawn in, and as students, you are probably familiar with bell curves for grading! Our histogram has equal distribution and a perfect bell. You can see that the circled green and blue boxes values fall under the bell. Remember they were the lowest frequencies observed in our data set. It is common to refer to these values as the 5% above the curve (green box) and the 5% below the curve (blue box).

It is important to note that histograms can be skewed. Meaning there may not be a lower or upper 5% region under the bell curve. Think to a real life example, maybe you were in a class where one person received an 80 and everyone else had a grade below 80. The opposite is also true. What would the histogram look like if one person received an 80 and everyone else received a 90 or greater. We will talk more about distributions in Chapter 5.

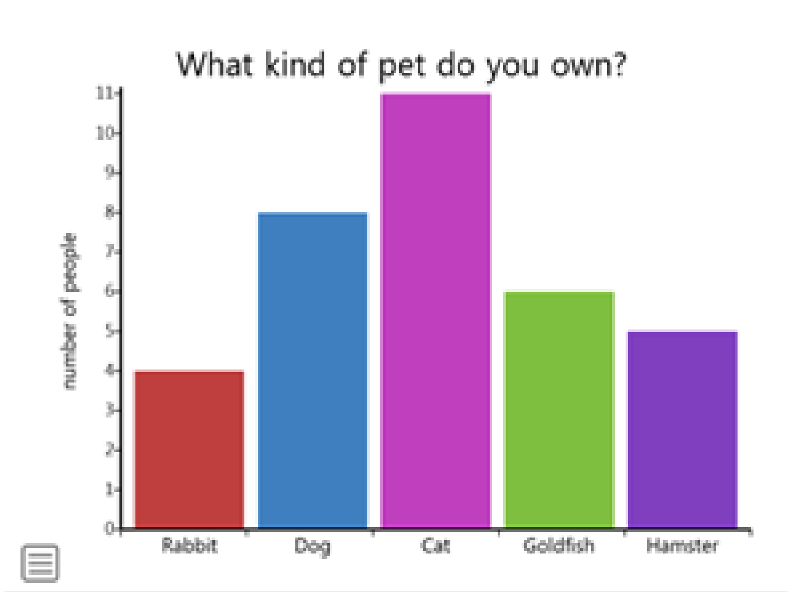
- Table: It is also common to represent data in the form of a table. A table is created as a quick way to visualize quantitative data to qualitative categories.

Students	Number of Students	Average Exam Score
Male	12	97
Female	15	98

Above is a very simple table. What are we comparing? The table shows male student versus female student exam scores. We are also given the number of male and female students. What is our qualitative data? Gender is the qualitative or category data, male versus female. What are the quantitative measurements? The number of students and average exam score is the quantitative values. Remember, quantitative data is numerical. Qualitative data, by itself, is not numerical.

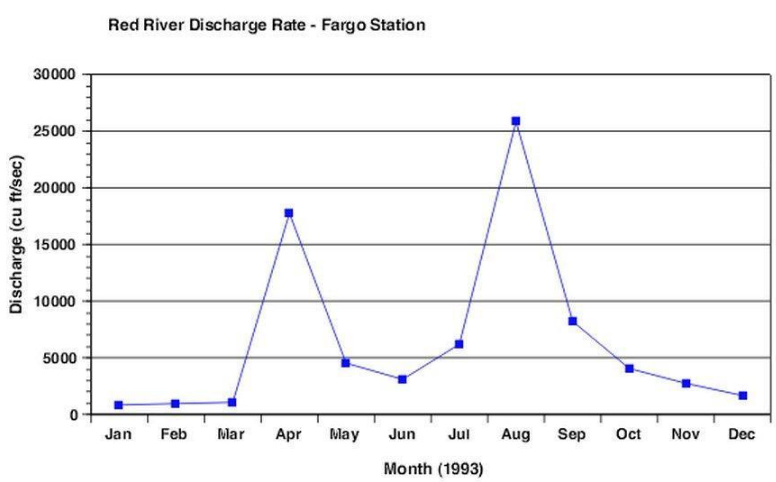
Additional Types of Graphs

- Bar Graph: Very similar to the histogram, it uses bars to represent numerical values. A bar graph does not have to follow a bell shape curve and may only have an upper or lower region. Below represents a typical bar graph.



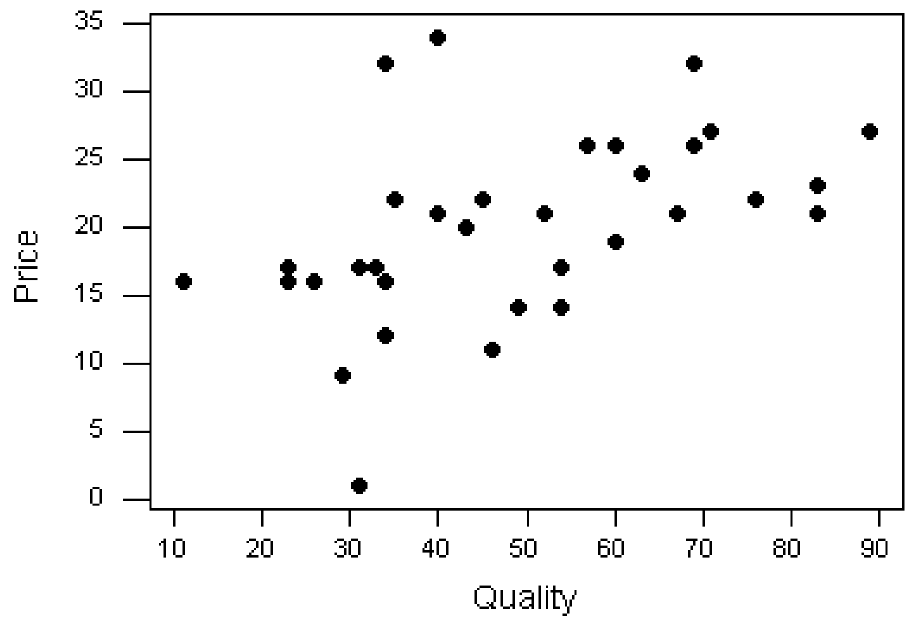
- Line Graph: Unlike the bar graph, a line graph uses a line to connect dots or points for a given data set; typically to represent changes over time.

The below line graph represents changes in river discharge rate over time, which is measured in months.

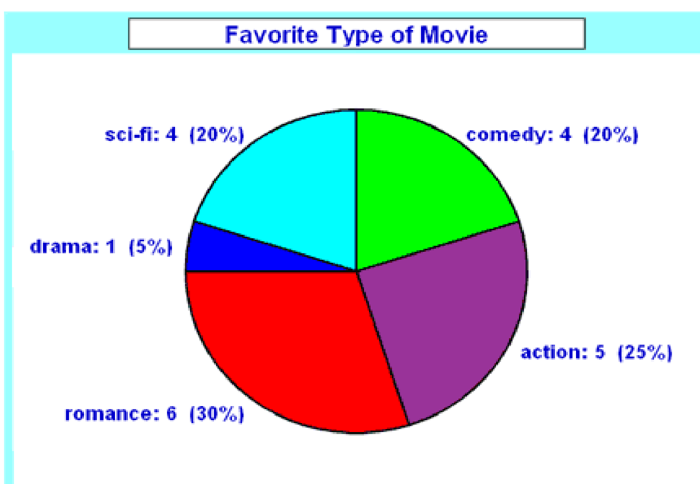


- Scatter Plot: Use specific points on a data set to mark specific coordinates or frequencies. However, unlike the line plot, the points are not connected. The independent variable is on the x-axis (or horizontal). The independent variable can be controlled or manipulated. In this case, the independent variable is the quality of the peanut butter. The dependent variable is the variable in the regression that cannot be controlled or manipulated, which is on the y-axis, or vertical. In this case, it is the price of the peanut butter.

Peanut Butter



- Pie Chart: A circle, or "pie," that is separated to represent a portion of the whole. Usually this is based on 100%. From this we can see that romance has the largest slice, which represents the most favorite type of movie.



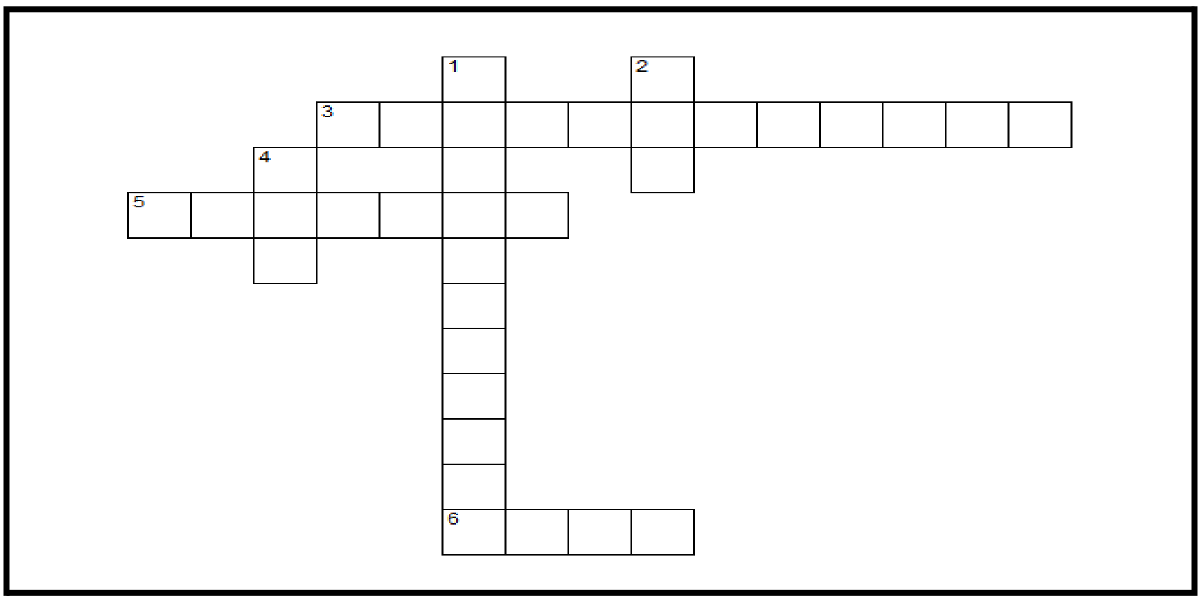
Chapter 2 Review

Below is an outline of the major points covered in Chapter 2. You will need to clearly understand the concepts, terms and graphs presented in Chapter 2. You may need to refer back to Chapters 1 and 2 as we progress through the text. All concepts build from each other. You should be able to define and interpret:

- Quantitative variables
- Categorical variables
- Stem-and-Leaf plots
- Histograms
- Tables
- Bar graphs
- Line graphs
- Scatter plots
- Pie chart

Ch. 2 Crossword Puzzle

Complete the following crossword using definitions from Chapter 2.



ACROSS

3. An objective measurement based on real numbers
5. Plot that uses specific points on a data set to mark specific coordinates or frequencies
6. Type of graph using a line to connect dots or points for a given data set; typically to represent changes over time

DOWN

1. Data represents specific categories that are not associated with real numbers
2. Chart that makes a circle, which is separated to represent a portion of the whole
4. Type of graph similar to the histogram; uses bars to represent numerical values

Chapter 2 Practice Problems

Complete the following practice problems to test your understanding. Check your work with the solutions provided.

You are given the following data of patient weight in pounds from a clinical trial:

180, 181, 182, 155, 150, 151, 167, 178, 135, 189, 220, 179, 122, 169

Use this data set to answer the following questions:

1. Make a stem-and-leaf plot of the data.

2. What is the mean weight of the patients?

3. Does this data set have a mode? If so, what is the mode?

4. What is the range of the patient's weight?

Answer key is found in the Answer Keys section.

Chapter 2 Quiz

Use the graph below to answer questions 1-2.

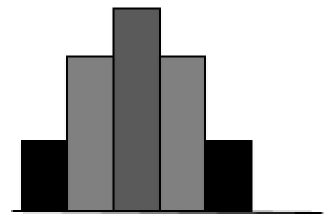
2	0 1 1 2
3	6
4	3 4 4 4 4 9
5	0

1. Given the following data, calculate the mean.
   a. 21
   b. 36.5
   c. 44
   d. 50

2. From the above data set, what is the mode?
   a. 44
   b. 21
   c. 36
   d. 29

3. Stem-and-leaf plots organize the data based on the properties of ___________.
   a. Odd numbers
   b. Even numbers
   c. Real numbers
   d. None of the above

4. The following graph is an example of?



   a. Bar graph
   b. Histogram
   c. Scatter plot
   d. Line graph

5. A ______ is created as a quick way to visualize quantitative data to qualitative categories.
   a. Bar graph
   b. Scatter plot
   c. Pie chart
   d. Table

Answer key is found in the Answer Keys section.

Chapter 3: Regression and Correlation

Regression and correlation analysis are used to determine the relationship between two quantitative variables. This section will begin with a description of common terms followed by an in-depth explanation of each concept, regression and correlation. As we progress, do not forget to refer back to Chapters 1 and 2.

Learning Objectives

After reading Chapter 3 and completing the workbook, you should be able to:

- Know the two types variables.
- Know how to calculate slope, intercept and regression.
- Know how to interpret graphs and determine correlations.
- Know the definitions of regression and correlation
- Apply the basic statistical concepts and understanding of data to draw conclusions and interpretations.

Study Clues

A clear understanding of the basic statistical terms and concepts presented in Chapter 3 will help prepare you to advance in this course and learn more complex statistical calculations and specific statistical tests. You should refer back to Chapters 1 and 2 and understand all concepts presented thus far. As you study, you should pay particular attention to the definitions and you should have an understanding of how a graph is laid out. You should be able to locate the x-axis and y-axis and know which represents the dependent and independent variable. You should pay particular attention to the equations and calculations for regression analysis.

3.1 Basic Terms

- Variable: A mathematical function that may change with time. It is the item or set of items being investigated and compared in the data set. Examples of variable include: effect, time, days, scores, weights, and grades.
- Independent variable: This is a variable that stands alone and is not subject to change. Example of independent variable would be time and gender.
- Dependent variable: This variable is dependent upon the independent variable. The dependent variable most often changes in response to the independent variable. Examples of a dependent variable include exam scores or weight. Both may be subject to change based on an independent variable such as time or gender.
- Normal distribution: The distribution of several random variables, it is most often seen as a symmetrical bell-shaped graph; recall the histogram described in Chapter 2.
- Random variable: As the name suggests, it is a randomly assigned quantity that has a numerical value for each member of a group. Its value has an equal number of opportunities

to be chosen. Think of putting names in a hat—you have to draw ten names, and each name has an equal chance of being drawn

**Most statistical analyses must have randomization. This is a very important concept when you are applying the information to a given population.*

3.2 Regression Analysis

Regression analysis is use to determine how the dependent variable changes when the independent variable is altered. For example, how do student exam scores (dependent variable) change over the semester or time (independent variable)? From this example, we could compare early scores from the first exam taken at the start of the semester to those scores from exams taken at the end of the semester.

We can use regression analysis to make a quantitative prediction. From our example, how do scores change over time? We may hypothesize, and hope, that scores will increase over time.

Formula and Calculations for Regression Analysis

The formula for regression analysis is written as: $y = a + bx$

We have several components to this equation. Let us look at each.

- Both x and y are always the variables.
- b is the slope of the line.
    - Slope can be calculated with the following formula:

$$b = \frac{n\sum xy - \left[\left(\sum x\right)\left(\sum y\right)\right]}{n\sum x^2 - \left(\sum x\right)^2}$$

- a is the intercept point of the line or slope at the y-axis.
    - The intercept can be calculated with the following formula:

$$a = \frac{\sum y - b\left(\sum x\right)}{n}$$

- n = Number of values or observations
- x = First variable
- y = Second variable

- $\sum xy$ = Sum of the product of first and second variable
    - Σ is the capital Greek letter, "Sigma", in mathematics Σ is an operator meaning summation

- $\sum x$ = Sum of first variables
- $\sum y$ = Sum of second variables
- $\sum x^2$ = Sum of square first variables

That is a lot of information. It is vital that you understand the components to the formulas and understand how to calculate each part.

Let us look at an example. You are given the following data set and asked to perform a regression analysis to make a prediction about the data.

x	10	11	12	13	14
y	2	4	6	8	10

The 1st step is to determine the slope, $b = \frac{n\sum xy - \left[\left(\sum x\right)\left(\sum y\right)\right]}{n\sum x^2 - \left(\sum x\right)^2}$. So we need to find all the values for our formula.

To do this, we need to know the number of observations or n. In this problem $n = 5$.

Now we need to find the other values: xy and x^2. We can easily calculate those by creating a table:

x Value	y Value	xy	x^2
10	2	20	100
11	4	44	121
12	6	72	144
13	8	104	169
14	10	140	196

**For review: xy is the product of x and y, for example,* $10\times2 = 20$*; and* x^2 *is the product of x and x, for example,* $10\times10 = 100$*.*

To calculate slope, we need to find $\sum x$, $\sum y$, $\sum xy$, and $\sum x^2$ (remember, Σ just stands for sum).

- $\sum x = 60$
- $\sum y = 30$
- $\sum xy = 380$
- $\sum x^2 = 730$

Now we are ready to solve for the slope. Substitute in the above values into the slope formula.

$$b = \frac{n\sum xy - \left[\left(\sum x\right)\left(\sum y\right)\right]}{n\sum x^2 - \left(\sum x\right)^2} = \frac{5(380)-60(30)}{5(730)-(60)^2} = \frac{100}{50} = 2$$

The 2nd step in regression analysis is to determine the intercept: We can substitute the calculated slope into the intercept formula.

$$a = \frac{\sum y - b\left(\sum x\right)}{n} = \frac{30-2(60)}{5} = -\frac{90}{5} = -18$$

The 3rd step, now that we know the slope and intercept, we can use these values in regression equation formula.

$$y = a + bx = -18 + 2x$$

If we want to know the approximate y value is we are given $x = 15$. Then we can substitute the value in the above equation and solve for y.

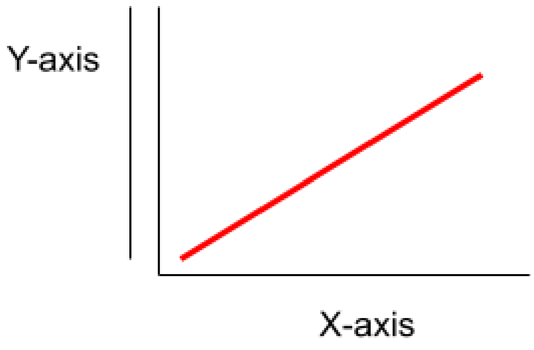
$$y = -18 + 2(15) = -18 + 30 = 12$$

**Tip: Remember to use order of operations when solving for slope and intercept.*

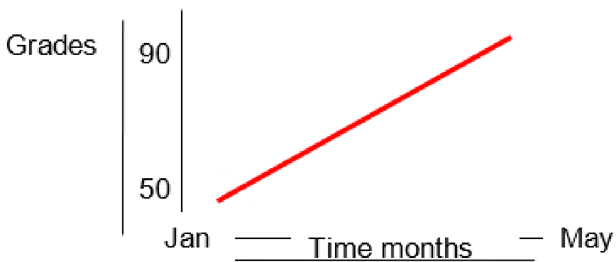
3.3 Correlation Analysis

Correlation is used to determine the effect. For example, how does y change with x; or, how does the dependent variable change when the independent variable changes. From this information, we can determine relationships.

Let's look at an example.



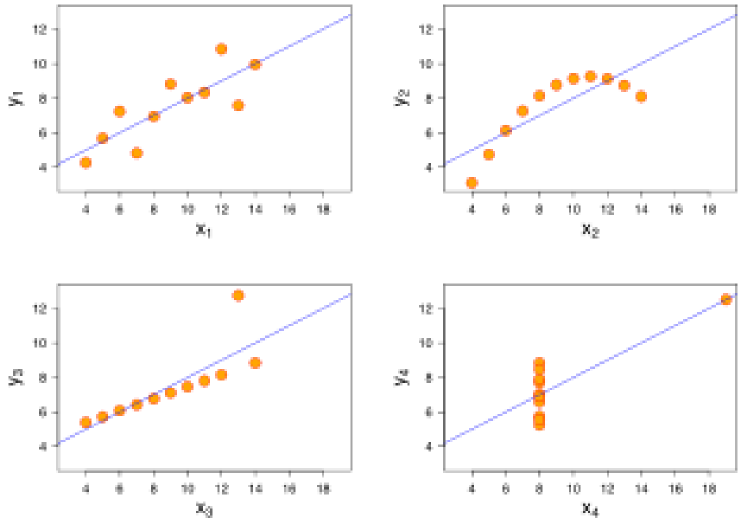
From this linear example, we are looking into the correlation of x (independent variable) on y (dependent variable). *This is true for all graphs, and you should remember the location of the x-axis and y-axis.* We can see that as x increases, y increases. Let's put some values to our graph:



From the above graph, student scores increase as time in months increases. We could formulate a relationship between student grades and time. We would be able to state that there is a positive correlation between improved grades during the course of a semester.

Linearity

The Pearson correlation coefficient indicates the strength of a linear relationship between two variables, but its value generally does not completely characterize their relationship. In particular, if the conditional mean of Y given X, denoted E(Y|X), is not linear in X, the correlation coefficient will not fully determine the form of E(Y|X).



From the image above, the four y variables have the same mean (7.5), standard deviation (4.12), correlation (0.816), and regression line ($y = 3 + 0.5x$). However, as can be seen on the plots, the distribution of the variables is very different. The first one (top left) seems to be distributed normally, and corresponds to what one would expect when considering two variables correlated and following the assumption of normality. The second one (top right) is not distributed normally; while an obvious relationship between the two variables can be observed, it is not linear. In this case the Pearson correlation coefficient does not indicate that there is an exact functional relationship, only the extent to which that relationship can be approximated by a linear relationship. In the third case (bottom left), the linear relationship is perfect, except for one outlier which exerts enough influence to lower the correlation coefficient from 1 to 0.816. Finally, the fourth example (bottom right) shows one outlier is enough to produce a high correlation coefficient, even though the relationship between the two variables is not linear.

These examples indicate that the correlation coefficient, as a summary statistic, cannot replace visual examination of the data. Note that the examples are sometimes said to demonstrate that the Pearson correlation assumes that the data follow a normal distribution, but this is not correct.

3.4 Pearson's Correlation

It is obtained by dividing the covariance of the two variables by the product of their standard deviations. Karl Pearson developed the coefficient from a similar but slightly different idea by Francis Galton.

The population correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is defined as:

$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y},$$

where **E** is the expected value operator, *cov* means covariance, and, *corr* a widely used alternative notation for Pearson's correlation. The Pearson correlation is defined only if both of the standard deviations are finite and both of them are nonzero (we will learn more about standard deviation later).

The Pearson correlation is +1 in the case of a perfect positive (increasing) linear relationship (correlation), −1 in the case of a perfect decreasing (negative) linear relationship (anticorrelation), and some value between −1 and 1 in all other cases, indicating the degree of linear dependence between the variables. As it approaches zero there is less of a relationship (closer to uncorrelated). The closer the coefficient is to either −1 or 1, the stronger the correlation between the variables.

If the variables are independent, Pearson's correlation coefficient is 0, but the converse is not true because the correlation coefficient detects only linear dependencies between two variables. For example, suppose the random variable X is symmetrically distributed about zero, and $Y = X^2$. Then Y is completely determined by X, so that X and Y are perfectly dependent, but their correlation is zero; they are uncorrelated. However, in the special case when X and Y are jointly normal, how a set is uncorrelated is equivalent to its degree of independence.

Chapter 3 Review

Below is an outline of the major points covered in Chapter 3. You will need to clearly understand the concepts, terms and graphs presented in Chapter 3. You may need to refer back to Chapters 1 and 2 as we progress through the text. All concepts build from each other.

You should be able to …

Define	Calculate	Understand
Independent variable Dependent variable Random variable	Slope Intercept Regression	Correlation analysis and know how the independent variable can influence the dependent variable.

Ch. 3 Fill in the Blank

Complete the following fill-in-the blank exercises.

1. A __________ is a mathematical function that may change with time.

2. The __________ variable stands alone and is not subject to change.

3. The ___________ variable most often changes in response to the independent variable.

4. The __________ variable value has an equal number of opportunities to be chosen.

Answer key is found in the Answer Keys section.

Chapter 3 Practice Problems

Complete the following practice problems to test your understanding. Check your work with the solutions provided.

Given the following data set, answer the questions.

x	y
2	4
4	3
6	2
8	1

1. Calculate the slope.

2. Calculate the intercept.

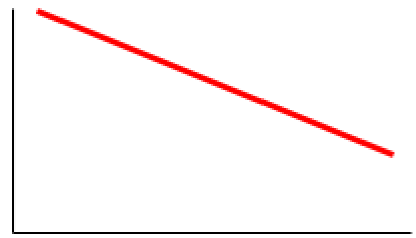
3. Calculate the least squares line of regression.

4. From the above, if $x = 5$ what is y?

Answer key is found in the Answer Keys section.

Chapter 3 Quiz

1. According to the following linear relationship, what happens to y when x increases?



   a. Decreases
   b. Increases
   c. y does not change
   d. None of the above

2. According to the above graph, which axis represents the independent variable?
   a. y
   b. x
   c. x and y
   d. There is not enough information to determine the variable

3. Given the following regression equation, which is true?

$$y = 2 + 3x$$

   a. The correlation between x and y is positive
   b. The value of y is 2 more than that of x
   c. The value of x decreases 3 units per every 1 unit increase in y
   d. The correlation between x and y is negative

4. The symbol Σ stands for
   a. Division
   b. Function
   c. Variable
   d. Sum

5. The regression equation is:
   a. $x = a + by$
   b. $y = a + \frac{b}{x}$
   c. $y = a + bx$
   d. $y = a - bx$

Answer key is found in the Answer Keys section.

Chapter 4: Basic Probability Theory

The probability is a measure that a given outcome will occur. Probability is measured from 0 to 1, with 0 indicating that an outcome will not occur and 1 indicating that an outcome will occur. Very rarely will you see a probability of 1. As statisticians, you will encounter values close to 1; the closer to 1, the more likely the outcome. This section will begin with a description of common terms followed by an in-depth explanation of the basic probability application. As we progress, do not forget to refer back to Chapters 1, 2, and 3.

Learning Objectives

After reading Chapter 4 and completing the workbook, you should be able to:

- Explain the definition of probability.
- Explain the definitions of mutually exclusive and non-mutually exclusive events.
- Explain the definition of conditional probability.
- Explain the definition of randomness.
- Explain probability events and their calculations.
- Explain how to calculate and interpret the probability of events.
- Explain how to apply basic probability.

Study Clues

A clear understanding of the basic statistical terms and concepts presented in Chapter 4 will help prepare you to advance in this course and learn more complex statistical calculations and specific statistical tests. You should refer back to Chapters 1, 2, and 3 and understand all concepts presented thus far. As you study, you should pay particular attention to the definitions and you should have an understanding of how a probability is calculated and applied. You should pay particular attention to the probability events their corresponding calculations.

4.1 Basic Terms

- Mutually exclusive: When two events or outcomes cannot occur at the same time, they are said to be mutually exclusive. However, each event has the same chance or probability of occurring. The classic example of this is tossing a coin. You have a 50% chance of getting heads or tails, but not both at the same time.
- Not mutually exclusive: This occurs when one or more event can occur at the same time. If you are giving a survey, you can expect to have multiple outcomes from the population if you ask the question "How well do you enjoy statistics?" You can also have mutually non-exclusive events if you draw a card from a deck and determine the probability that it will be either black or red and a face card.
- Conditional probability: Some of the outcomes for (A) will occur if some of the outcomes for (B) occur. With this, A is partially dependent on B. If some of the outcomes for B do not

occur, than none of A can occur. With this, there will be some outcomes for B that will not influence A.

- Randomness: All outcomes have the same probability of being chosen or occurring. If you randomly ask people on a street to answer a survey question, given all other factors are equal (meaning that the street is not comprised of a single population, e.g., all male) then you would expect your surveys outcome to represent a random sample of the population.
- Population: A population is an entire group, collection or space of objects that we want to characterize.
- Sample: A sample is a collection of observations on which we measure one or more characteristics. Frequently, we use (small) samples of (large) populations to characterize the properties and affinities within the space of objects in the population of interest. For example, if we want to characterize the US population, we can take a sample (poll or survey) and the summaries that we obtain from the sample (e.g., mean age, race, income, body-weight, etc.) may be used to study the properties of the population, in general.
- Variable: A variable is a characteristic of an observation that can be assigned a number or a category. For instance, the year in college (variable) for a student (observational unit).

4.2 Types of Variables

Appropriate classification of process and variable types are important because they directly influence our decision on how to collect, explore, analyze and interpret data and results. For example, we can carry arithmetic (e.g., average) on quantitative variables, but we need to analyze frequencies of occurrence for qualitative variables.

There are two types of variables: categorical and quantitative. These types of variables can be split further.

- Categorical: Categorical variables are qualitative measurements of samples or populations that are classified into groups:
  - Ordinal categorical variables are qualitative descriptions that have a natural arrangement or order of the measurements; e.g., rank in college (freshman, sophomore, junior, senior), size of soda (small, medium, large), etc.
  - Not ordinal (or nominal) variable is a categorical variable that does not have a naturally imposed (or meaningful) order of its values; e.g., gender, race, political affiliation (democrat, republican, independent, green party, other), etc.
- Quantitative: Quantitative variables are measurements that have a meaningful numerical value representation. There are two types of quantitative variables:
  - Continuous variables indicate numerical observations that contain intervals with infinite (uncountable) possible values; e.g., weight, height, time, speed, etc.

  - Discrete variables are also numerical measurements, but they are sparse in space and any interval will contain at most many possible values; e.g., number of students in a school, number of rational numbers in a given interval [a; b], age, etc.

4.3 Probability

There are a few rules for probability:

1. Probabilities can be represented with classical formulas

2. P always stands for probability

3. The range for probability is always 0 (not occurring) to 1 (occurs)

4. You will always have at least two events: A and/or B

- The probability of event A only occurring or not occurring is written as:
  - $P(A) = [0, 1]$.
- The probability that event A will not occur is written as:
  - $P(A^B) = 1 - P(A)$
    - This is the same as calculating the probability of B
- The probability of A or B occurring is written as:
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - Or for mutually exclusive: $P(A \cup B) = P(A) + P(B)$
    - The symbol ∪stands for union and that the events are A or B or both
    - The symbol ∩ stands for intersect and states that A and B have events in common
- The probability that A and B will occur is written as:
  - $P(A \cap B) = P(AIB) * P(B) = P(BIA) * P(A)$
  - Or if the events are independent: $P(A \cap B) = P(A) * P(B)$
    - The symbol I means to include
- The probability of A occurring given B is written as:
  - $P(AIB) = P(A \cap B) / P(B)$

Application of Probability

Probabilities are used to predict outcomes and they have many different applications. Probabilities can be used to predict:

- Economic predictions and outcomes

- Consumer reaction to a new product
- Scientific research
- Future business forecast
- Voting trends

**Can you think of more ways probabilities are used in everyday life?*

Interpretation of Probability

If the probability of event A occurring is 0.99, this indicates that there is a very good probability of event A occurring. You can multiply the given probability by 100 to see the percent. If we look at event A occurring is $0.99 \times 100 = 99\%$, then there is a 99% chance that event A will occur.

Remember, the closer your event is to 1, the greater the probability. The probability can never be more than 1 or less than 0.

Examples for calculating probability:

Let's say the probability of Event A occurring is 0.75 and we want to know the probability of B occurring. We will assume A and B are mutually exclusive. For this example, we take the greatest probability 1 minus the probability of A.

We get the following: $1 - 0.75 = 0.25$

So, the probability of B occurring in this case is 0.25 or 25%.

Let's look at another example. If there is a 0.15 probability of Event A occurring and a 0.20 probability of Event B occurring, then what is the probability that Event A or B will occur, given that A and B are mutually exclusive?

For this example, we simply add the probabilities together. $0.15 + 0.20 = 0.35$ or a 35%

In this example, there is a 35% chance that events A or B will occur.

It is important that you note the probabilities of A and B do not have to equal 100%. There may be other events or outcomes outside of A and B that can occur. Additionally, in this example, you may have outcomes where A and B do not occur at all.

Chapter 4 Review

Below is an outline of the major points covered in Chapter 4. You will need to clearly understand the concepts, terms and graphs presented in Chapter 4. You may need to refer back to Chapters 1, 2, and 3 as we progress through the text. All concepts build from each other.

Outline	Chapter 4: Basic Probability Theory
Probability	Define ● Mutually exclusive ● Not mutually exclusive ● Conditional probability ● Randomness
Events	Calculate ● Probabilities based on events
Application	Interpret ● Probability and probability outcome ● Know how to apply probabilities to a population

Ch. 4 Complete the Table

Complete the following table:

Event Occurring	Probability Equation
The probability of event A only occurring or not occurring	
The probability that event A will not occur	
The probability of A or B occurring	
The probability that A and B will occur	
The probability of A occurring given B	

Chapter 4 Practice Problems

Complete the following practice problems to test your understanding. Check your work with the solutions provided.

1. Given the probability of event A (voters who will vote) is 65%, calculate the probability of event B (voters who will not vote).

2. Would these events in the above question be mutually exclusive?

3. The carnival will only occur if the weather is above 75 degrees. Which event is the dependent event and which event is the independent event.

4. How would you calculate probability for the above question? (just show the formula)

Answer key is found in the Answer Key section.

Chapter 4 Quiz

1. The probability that the township will pass a new fence ordinance is 0.0001. What can you say about this probability?
   a. There is a very low probability that the new ordinance will pass
   b. There is a very high probability that the new fence ordinance will pass
   c. There is not enough information given
   d. There is greater than a 99% probability that the new fence ordinance will pass

2. If A is mutually exclusive from B, then:
   a. B depends on A
   b. A depends on B
   c. A and B do not depend on one another
   d. Only B is not dependent on A

3. This word means that all outcomes have the same probability of being chosen or occurring.
   a. Randomness
   b. Conditional probability
   c. Probability
   d. Mutually exclusive

4. Given two mutually exclusive events, if the probability of A occurring is 0.10 then what is the probability of A not occurring?
   a. 1
   b. 0
   c. 0.90
   d. 0.75

5. The probability of A occurring is 0.15, the probability of B occurring is 0.15. What is the probability that neither A or B will occur?
   a. 0.70
   b. 0.15
   c. 0.30
   d. 1

Answer key is found in the Answer Key section.

Chapter 5: Probability Distributions

The probability distribution is defined as the probability of a given or set of given outcomes will occur in a statistical experiment. As discussed in Chapter 4, probability is measured from 0 to 1; with 0 indicating that an outcome will not occur and 1 indicating that an outcome will occur. A probability of 1 would indicate that a given outcome would occur 100% of the time, this is a rare occurrence. This section will begin with a description of common types of probability distributions; pay particular attention to the application of each. As we progress, do not forget to refer back to Chapters 1 through 4. Keep in mind that we will only cover the major probability distributions, you may come across others.

Learning Objectives

After reading Chapter 5 and completing the workbook, you should be able to:

1. Explain the basic definitions of common terms used in probability.
2. Explain the different types of commonly encountered probability distributions.
3. Explain the application of the presented probability distributions.

Study Clues

A clear understanding of the basic statistical terms and concepts presented in Chapter 5 will help prepare you to advance in this course and learn more complex statistical calculations and specific statistical tests. You should refer back to Chapters 1 through 4 and understand all concepts presented thus far. As you study, you should pay particular attention definitions of the common types of probability distributions. You should be able to apply the probability distributions to specific statistical applications.

5.1 Basic Terms

- Discrete: Can represent an individual variable that is separate or acting separately from all other variables.
- Finite: Is a mathematical number or function that has a countable end. If you had 10 coins, you have a finite number of coins (10).
- Infinite: Is a mathematical measurement or function that does not have an end; meaning it can continue indefinitely.

5.2 Types of Distributions

- Bernoulli distribution: This classical probability distribution is easily calculated with the following formula: $q = 1-p$. A common example of this distribution is measuring the probability outcome of a single coin toss. There are two outcomes to this experiment; the coin might come up heads with probability p or it may come up tails with probability 1-p.

- Binomial distribution: This type of distribution is used to describe the number of successes within a given series of events that are independent. A classic example and application of this distribution is the measurement of outcomes in a Yes/No experiment, where the outcomes yes versus no, have the same probability of success.

- Poisson binomial distribution: This is very similar to the above binomial distribution. However, it is the measurement of outcomes in a Yes/No experiment, where the outcomes yes versus no, have a different probability of success.

- Poisson distribution: Not to be confused with Poisson binomial distribution, this measures a large number of individual unlikely events that happen within a certain time interval. Think of two events that are not likely to happen at the same time.

- Normal distribution: This distribution is often times referred to as the Gaussian or bell curve. In this distribution, each variable can be measured as the sum of many smaller independent variables. For this, recall how average class grades are often expressed as a bell curve. Each individual grade becomes part of the bell.

- Chi-squared distribution: This distribution measures the sum of the squares of *N= the number of observations*) of independent Gaussian random variables. Gaussian means normally distributed, and example of this would be a bell curve.

- F- distribution: This is the distribution uses the chi-squared distribution. The F-distribution is the measurement of the ratio of two (normalized) random variables that experience chi-squared distribution. This is commonly used in the analysis of variance.

- Exponential distribution: This distribution is used to describe the time between consecutive events that are random in a random, non-reoccurring process.

- Students t-distribution: This distribution is used in the measurement and estimation of unknown means from Gaussian (normally distributed) populations.

Chapter 5 Review

Below is an outline of the major points covered in Chapter 5. Refer back to Chapters 1 through 4 as we progress through the text. All concepts build from each other.

Outline	Chapter 5: Probability Distributions	
Terms	Define: Discrete, Finite, and Infinite	
Distributions	Define the following distributions:	
	• Bernoulli • Binomial • Poisson Binomial • Normal	• Chi Squared • F • Exponential • Students t
Application	Interpret: When is each type of distribution used	

Ch. 5 Complete the Table

Complete the following table:

Distribution	Definition
Bernoulli	
	This type of distribution is used to describe the number of successes within a given series of events that are independent.
Poisson Binomial	
Poisson	
	In this distribution, each variable can be measured as the sum of many smaller independent variables.
Chi Squared	
F	
	This distribution is used to describe the time between consecutive events that are random in a random, non-reoccurring process.
Students T	

Chapter 5 Quiz

1. This can represent an individual variable that is separate or acting separately from all other variables.
    a. Finite
    b. Discrete
    c. Infinite
    d. Probability

2. This distribution is often times referred to as the Gaussian or bell curve. In this distribution, each variable can be measured as the sum of many smaller independent variables.
    a. Normal
    b. Binomial
    c. Poisson
    d. Chi Squared

3. This classical probability distribution is easily calculated with the following formula: $q = 1-p$.
    a. Students T
    b. F
    c. Binomial
    d. Bernoulli

4. Distribution is the measurement of the ratio of two (normalized) random variables that experience chi-squared distribution.
    a. Students T
    b. F
    c. Binomial
    d. Bernoulli

5. This is a mathematical number or function that has a countable end.
    a. Finite
    b. Discrete
    c. Infinite
    d. Probability

Answer key is found in the Answer Key section.

Chapter 6: Statistical Sampling

The purpose of statistical sampling is to select a population or experimental group in order to test a statistical hypothesis. We will begin by looking at the process of sample selection and discuss common sampling techniques. You may wish to refer to the previous chapters as we progress through Chapter 6.

Learning Objectives

After reading Chapter 6 and completing the workbook, you should be able to:

1. Explain the 6 steps of sampling.
2. Explain the basic definitions of common terms used in sampling.
3. Define and interpret population.
4. Explain the different types of commonly used sampling techniques.
5. Explain the application of the presented sampling techniques.

Study Clues

A clear understanding of the basic statistical terms and concepts presented in Chapter 6 will help prepare you to advance in this course and learn more complex statistical calculations and specific statistical tests. You should refer back to Chapters 1 through 5 and understand all concepts presented thus far. As you study, you should pay particular attention to the common types of sampling techniques. You should be able to apply the sampling techniques to specific populations.

6.1 Steps to Take When Selecting a Sample

- Define the population or experimental group.
- Specify and select what is to be tested or measured.
- Determine the correct method of sampling
- Determine the population size.
- Test the hypothesis.
- Collect the data.

Defining the population to be sampled is the most important of the steps. The population is determined by what the researcher wishes to study. A given population may include all people or items with the characteristics of interest. Studying an entire population is not very practical, so subsets are used, and these subsets then define the sample.

A sample that is selected at random holds the highest level of statistical power and yields the greatest interpretation of the data. Thus, it is important that you always strive for randomness when selecting a sample population!

6.2 Basic Terms

- Sampling frame: The source or population from which the sample is taken.
- Probability sampling: For this sampling technique, every possibility (characteristics) within the given population has an equal opportunity or likelihood of being selected for the sample. This sampling method produces an unbiased representation of the population.
- Non-probability sampling: For this sampling technique, not every possibility (characteristic) within the given population has an equal opportunity or likelihood of being selected for the sample. This sampling method can produce a biased or skewed representation of the population.

6.3 Common Sampling Techniques

- Simple random sample: This is the most common of the sampling techniques. During simple random sample selection, all subsets of the given frame have an equal opportunity of being selected.
- Systematic sampling: This technique arranges the population based on some pre-determined order. The sample is then selected at a determined interval from the list. Systematic sampling begins with a random sample and then implements a non-random selection from a formed list. This allows the population to be clustered. Imagine trying to sample New York City. Instead of sampling each block of the city, you may group the blocks and sections and decide to sample every 20th block.
- Stratified sampling: During this sampling technique, the population experiences and is organized into distinct categories or classifications. The sampling frame is then organized into categories also called strata. The individual stratums are then randomly sampled as independent sub-populations.
- Cluster sampling: With a large target population of interest, sometimes sampling is clustered or grouped by geography or other defining population characteristics (such as generations).
- Panel sampling: This is very similar to systemic sampling. First, a given population is selected from a random list of participants. The sample is then reduced by asking questions that would narrow down the given sample. You may have a list of all freshmen at a given college; you can then narrow the list down by asking questions that might pertain to your characteristic of interest. Let's say you want only freshmen who eat yogurt for breakfast. You may ask the original group if they eat breakfast, then the participants that answer no, will be removed from study consideration.

Chapter 6 Review

Below is an outline of the major points covered in Chapter 6. You will need to clearly understand the concepts, terms and graphs presented in Chapter 6. You may need to refer back to Chapters 1 through 5 as we progress through the text. All concepts build from each other.

Outline	Chapter 6: Statistical Samplings	
Terms	Define	
	• Population • Sample • Random sample	• Sample frame • Probability sampling • Non-probability sampling
Sampling	Know the six steps of samplings	
Types of sampling	Define the type of sampling	
	• Simple • Systematic • Stratified	• Cluster • Panel

Ch. 6 Complete the Table

Complete the following table:

Sampling	Definition
	All subsets of the given frame have an equal opportunity of being selected.
	This technique arranges the population based on some pre-determined order. The sample is then selected at a determined interval from the list.
	During this sampling technique, the population experiences and is organized into distinct categories or classifications. The sampling frame is then organized into categories also called strata.
	Sampling is clustered or grouped by geography or other defining population characteristics (such as generations).
	First, a given population is selected from a random list of participants. The sample is then reduced by asking questions that would narrow down the given sample.

Chapter 6 Practice Questions

1. What are the six steps to sampling in order?

2. Why is sampling used to study a population?

3. What is the difference between probability and non-probability sampling?

Answer key is found in the Answer Key section.

Chapter 6 Quiz

1. All subsets of the given frame have an equal opportunity of being selected.
    a. Simple Random Sampling
    b. Cluster Sampling
    c. Infinite Sampling
    d. Panel Sampling

2. This sampling method produces an unbiased representation of the population.
    a. Probability Sampling
    b. Non-probability Sampling
    c. Cluster
    d. Strata Sampling

3. This sampling method can produce a biased or skewed representation of the population.
    a. Probability Sampling
    b. Non-Probability Sampling
    c. Cluster
    d. Strata Sampling

4. The population is determined by what the researcher wishes to study
    a. True
    b. False
    c. True some of the time

5. A sample that is not selected at random holds the highest level of statistical power and yields the greatest interpretation of the data.
    a. True
    b. False
    c. True some of the time

Answer key is found in the Answer Key section.

Chapter 7: Statistical Estimations

Statistical estimations are the given values for a set of known parameters. It is important to note that statistical estimations assume the data is random and the probability distribution is dependent upon the parameters of interest. The given parameters are derived from the measured data. The data set must include at least one component of randomness (usually in selecting the population or assigning a given treatment). The established parameters ultimately will define the distribution of the population of interest. We can further use our known parameters to identify and measure unknown parameters. This chapter will cover basic components and calculations of the Estimation Theory. It may be necessary to refer to previous chapters for terminology.

Learning Objectives

After reading Chapter 7 and completing the workbook, you should be able to:

1. Explain basic statistical estimation tests and predictions.
2. Calculate and define standard deviation of the sample and population.
3. Define and calculate variance.
4. Calculate the degree of freedom.
5. Define and calculate the standard error of the mean.
6. Define and calculate confidence interval.

Study Clues

A clear understanding of the basic statistical estimations presented in Chapter 7 will help prepare you to advance in this course and learn more complex statistical calculations and specific statistical tests. As you study, you should pay particular attention to the definitions, abbreviations and how to calculate each of the following: SEM, SE, and CI. You should refer to previous chapters as all chapters are designed to build from one another.

7.1 Needed Calculations for Estimates

Statistical estimations rely on several calculations defining the parameters. The Standard Deviation (SD) is also represented by the symbol σ (the lowercase Greek letter, sigma). The SD is a measure of how spread (or deviation from the mean). For SD, we are interested in determining how far our results or population is from the mean.

There are two types of standard deviation, each with their own calculation!

Let's first address the *standard deviation of a population*: This can be calculation by fining the square root of the Variance. Variance is the average of the squared differences from the mean. That might sound confusing at first. There are three basic steps used to calculate variance.

- 1st find the sample mean.
- 2nd find the difference. To do this, subtract the mean from each individual observation or value. Then you square the differences.

- 3rd calculate the mean of the squared differences.

The standard deviation of the population is only used if we know every single value of the population. Often times, the population is too big and in statistics a sample of the population is often times measured.

To calculate the *standard deviation of the sample*, we have n data values (n = number of observations). Therefore, variance is calculated differently. Instead of dividing by the number of observations, as we did for the population, we divide by n-1.

All other calculations stay the same!

Take the following example:

If we are given the following values of weights for a sample: 100, 110, 120, 125, 130

We would calculate the variance by the following:

1. Step 1: Find the sample mean. $\mu = \frac{100+110+120+125+130}{5} = 117$
2. Step 2: Subtract the mean from each value and place the result in column B of the table.
3. Step 3: Square each result and place the squares in column C of the table.

A	B	C
x_i	$x_i - \mu$	$\left(x_i - \mu\right)^2$
100	$100 - 117 =- 17$	$(-17)^2 = 289$
110	$110 - 117 =- 7$	$(-7)^2 = 49$
120	$120 - 117 = 3$	$3^2 = 9$
125	$125 - 117 = 8$	$8^2 = 64$
130	$130 - 117 = 13$	$13^2 = 169$
D	$\sum\left(x_i - \mu\right)^2$	$289 + 49 + 9 + 64 + 169 = 5$

4. Step 4: Find the sum of the squares in Column C (see column D).
5. Step 5: Divide the sum by $n - 1$ to get the variance.

$$Var = s^2 = \frac{\sum\left(x_i - \mu\right)^2}{n-1} = \frac{580}{5-1} = 145$$

6. Step 6: Take the square root to get the standard deviation.

$$s = \sqrt{s^2} = \sqrt{145} \approx 12.04$$

- Standard Error of the Mean (SEM) is the standard deviation of the estimate of the sample mean of a given population mean. To break down that definition, think about finding the error of deviation of the given sample from the entire population. In statistics, we need to know how good our sample is. So, we ask, how well does the sample fit the population? There will always be a certain level of error with any statistical estimate.

We can calculate SEM by (sample standard deviation) divided by the square root of the sample size (also expressed as n). SEM is often represented by the following formula:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

- s is the sample standard deviation
- n is the number of observations for our given sample

Using the previous example from above, we calculated:

- $n = 5$ (number of observations from the previous example)
- Sample Standard Deviation = $s \approx 12.04$
- $SEM = S\overline{E}_x = \frac{12.04}{\sqrt{5}} \approx 5.38$

We can now interpret our sample of the population by saying the standard deviation is 12.04 with a calculated sample error of 5.38. For our sample of weights, the mean was 117; our sample falls within the mean with a deviation of + or – 12.04 points from the mean with an error of 5.38 or just 5 if rounded up.

Let's check the math:

Our mean is 117 +12 (deviation from the mean) = 129; we have an error of 5 points, so if we add 5 = 134 (max weight).

Now let's look at the lowest weight. 117 - 12 (deviation from the mean) = 105 with an error of 4, so if we subtract 4 we get 101 (pretty close to 100!) – note it does not have to equal the upper and lower values, but it should be pretty close.

So, our sample does not deviate too much from the mean! The more deviation from the mean and the higher the error, the less effective the sample is at representing the population.

Confidence interval (CI) is used to indicate the reliability of a statistical estimate; meaning how confident we are in our estimation.

We can calculate the confidence interval boundaries: the mean minus a calculation of uncertainty, to the mean plus that calculation of uncertainty.

$$CI = \bar{x} \pm \left(t^{*} \times \frac{s}{\sqrt{n}}\right)$$

- $\bar{x}$ = sample mean
- SEM = standard error of the mean
  - s = standard deviation of the sample
  - n = sample size
- t^{*} is the critical value and found in a t-table. On an exam, you would be given the t-value or t-table. The critical t-value is dependent on the confidence interval you want to achieve, for example 95% with a P < 0.05 has a t-value of 2.045

If we are given the following example:

Mean = 122

SD = 9

n (number of observations) = 30

Degree of freedom = 30-1 = 29

t (look-up, two-tailed) = 2.045 (for 95% CI, P < 0.05)

SEM = 9/√30 = 1.64

95% CI = 122 - (2.045×1.64) = 119 to 122 + (2.045×1.64) = 125

95% CI = 119 to 125.

We can now say, from our given sample, we are 95% confident that the true population mean lies between 119 and 125.

Chapter 7 Review

Below is an outline of the major points covered in Chapter 7. You will need to clearly understand the concepts, terms and calculation presented in Chapter 7. You may need to refer back to previous chapters as we progress through the text. All concepts build from each other.

Outline	Chapter 7: Statistical Estimations
Estimation Terms	Define: ● Standard deviation ● Variance ● Standard error of the mean ● Confidence interval
Estimation Calculations	Calculate: ● Standard deviation (sample and population) ● Variance ● Standard error of the mean ● Confidence interval
Statistical Application	You should know how to apply the critical t-value and be able to interpret the estimation values calculated.

Ch. 7 Fill in the Blank

1. ______________ is the average of the squared differences from the mean

2. _________________ is used to indicate the reliability of a statistical estimate; meaning how confident our we in our estimation

3. N-1 is also referred to as the _____________________.

4. t is the critical value and found in a ___ table.

5. ______________ is the standard deviation of the estimate of the sample mean of a given population mean.

Chapter 7 Practice Problems

Complete the following practice problems to test your understanding. Check your work with the solutions provided.

You are given the following grades from a sample of the statistics class: 75, 77, 82, 85, 93. Use this data set to answer the following questions:

1. What is the mean?

2. What is the variance?

3. What is the standard deviation?

4. What is the standard error of the mean?

Answer key is found in the Answer Key section.

Chapter 7 Quiz

1. What is the average of the squared differences from the mean?
    a. Variance
    b. Standard deviation
    c. Confidence interval
    d. Standard error of the mean

2. Given the following sample information; SD = 25; mean = 40; n=32; and t = 2.045 (for 95% CI, $P < 0.05$): Calculate the degrees of freedom.
    a. 32
    b. 435
    c. 31
    d. 5.7

3. From the above information, calculate the SEM.
    a. 1
    b. 0.78
    c. 32
    d. 4.4

4. Using the same data set above: find the 95% confidence interval.
    a. 31 to 49
    b. 65 to 109
    c. 25 to 27
    d. 33 to 47

5. If you are given a population variance of 169, what is the standard deviation of the population?
    a. 84
    b. 13
    c. 99
    d. Cannot be determined

Answer key is found in the Answer Key section.

Chapter 8: Hypothesis Testing

In order to make a decision or interpretation about the data, you have to test a hypothesis. Hypothesis testing allows the data to be interpreted. The result of an experiment is considered statistically significant if it is not likely to have occurred by chance alone. That means the result was influenced by the experimental factor or dependent variable. We will focus on specific terms associated with hypothesis testing and how to define and interpret a hypothesis. You may want to refer back to Chapters 1 through 7 as we proceed.

Learning Objectives

After reading Chapter 8 and completing the workbook, you should be able to:

1. Explain the fundamentals of hypothesis testing.
2. Explain the difference between the null and alternative hypothesis.
3. Define p-value.
4. Define basic hypothesis terms.
5. Define basic hypothesis test.
6. Apply hypothesis testing.

Study Clues

A clear understanding of the basic statistical terms and hypothesis test presented in Chapter 8 is important to the overall understanding of the statistics presented in this course. As you study, you should pay particular attention to the definitions and application of the tests. Refer back to the previous chapters for a review of terms and concepts.

8.1 Hypothesis

A hypothesis test includes all possible outcomes. There are two types of hypotheses:

- Null hypothesis
- Alternative hypothesis

8.2 P-value

The p-value is the same as the probability value. We accept or reject the hypothesis based on the outcome of the p-value. If the p-value is less than the required significance level, usually the significance level is 0.05, the null hypothesis is rejected and the alternative hypothesis is accepted. However, if the p-value is not less than the significance level, typically anything above 0.05, then the null hypothesis is not rejected.

8.3 Basic Terms

- Null hypothesis (H_0): Is associated with a contradiction to an assumption to be tested.
- Alternative hypothesis (H_1): Is associated to an assumption to be tested.
- Region of acceptance: The values of the test which fail to reject the null hypothesis.
- Critical region: The values of the test which the null hypothesis is rejected.
- Critical value: The threshold value containing the boundaries of the regions in which the hypothesis will be accepted or rejected.
- Power (1 - β): The probability of correctly rejecting the null hypothesis. The symbol β represents the false negative, or incorrectly rejecting the null hypothesis. The power test determines the sensitivity of the statistical test.
- Significance level (α): The probability of *incorrectly* rejecting the null hypothesis. The symbol (α) represents the false positive (1 – α).
- Statistical significance: Results are considered statistically significant the tested sample is inconsistent with the (null) hypothesis.

8.4 Common Statistical Tests

- One-sample test: Compare the sample to the population from which the hypothesis is drawn.
- Two-sample test: Compare two samples, typically experimental and control samples.
- Paired test: Compare two samples where it is impossible to control all variables that may influence the results.
- Z-test: Compare the means of the data where normality and standard deviation are known; these are more tightly controlled statistical conditions.
- T-test: Compare the means of the data where normality and standard deviation may not be known; these are more relaxed statistical conditions.
- F-test: Also referred to the analysis of variance (ANOVA). F-tests are used to determine whether groupings of data by category are meaningful.

8.5 Types of Hypothesis Errors

- Type I error: Occurs when the null hypothesis is rejected when, in fact, it is true.
  - The probability of committing a Type I error is called the significance level, which is often shown with the alpha symbol, denoted by α.
- Type II error: Occurs when one fails to reject a null hypothesis, when it in fact, is false.
  - The probability of not committing a Type II error is called the power of the test; this is often shown with the symbol Beta, denoted by β.

Statistics

The table below outlines the formulas used during hypothesis testing.

Test	Formula
One-sample z-test	$z = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$
Two-sample z-test	$z = \frac{(\overline{x}_1 - \overline{x}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n1} + \frac{\sigma_2^2}{n2}}}$
One-sample t-test	$t = \frac{\overline{x} - \mu_0}{(s/\sqrt{n})},$
Paired t-test	$t = \frac{\overline{d} - d_0}{(s_d/\sqrt{n})},$
Two-sample pooled t-test, equal variances	$t = \frac{(\overline{x}_1 - \overline{x}_2) - d_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$ $df = n_1 + n_2 - 2$
Chi-squared test for variance	$\chi^2 = (n-1)\frac{s^2}{\sigma_0^2}$
Two-sample F test for equality of variances	$F = \frac{s_1^2}{s_2^2}$

Explanation of symbols used in the formulas:

- α, the probability of Type I error - rejecting a null hypothesis incorrectly
- n = sample size
- $\overline{x}$ = sample mean
- μ_0 = hypothesized population mean
- σ = population standard deviation
- s = sample standard deviation
- df = degrees of freedom
- $\overline{d}$ = sample mean of differences
- d_0 =hypothesized population mean difference
- χ^2 = Chi-squared statistic

Examples of Hypothesis Testing

Example 1: A gardener's flowers

- A few flowers in this vase are white
- Most of the flowers in the flower bed are white

Therefore, there is a probability that the flowers in the vase came from another flower bed.

The flowers in the flower bed are the population. The flowers in the vase represent the sample. The null hypothesis is that the observed sample originated from the observed population. We reject the null hypothesis because of the differences in appearance of the flowers.

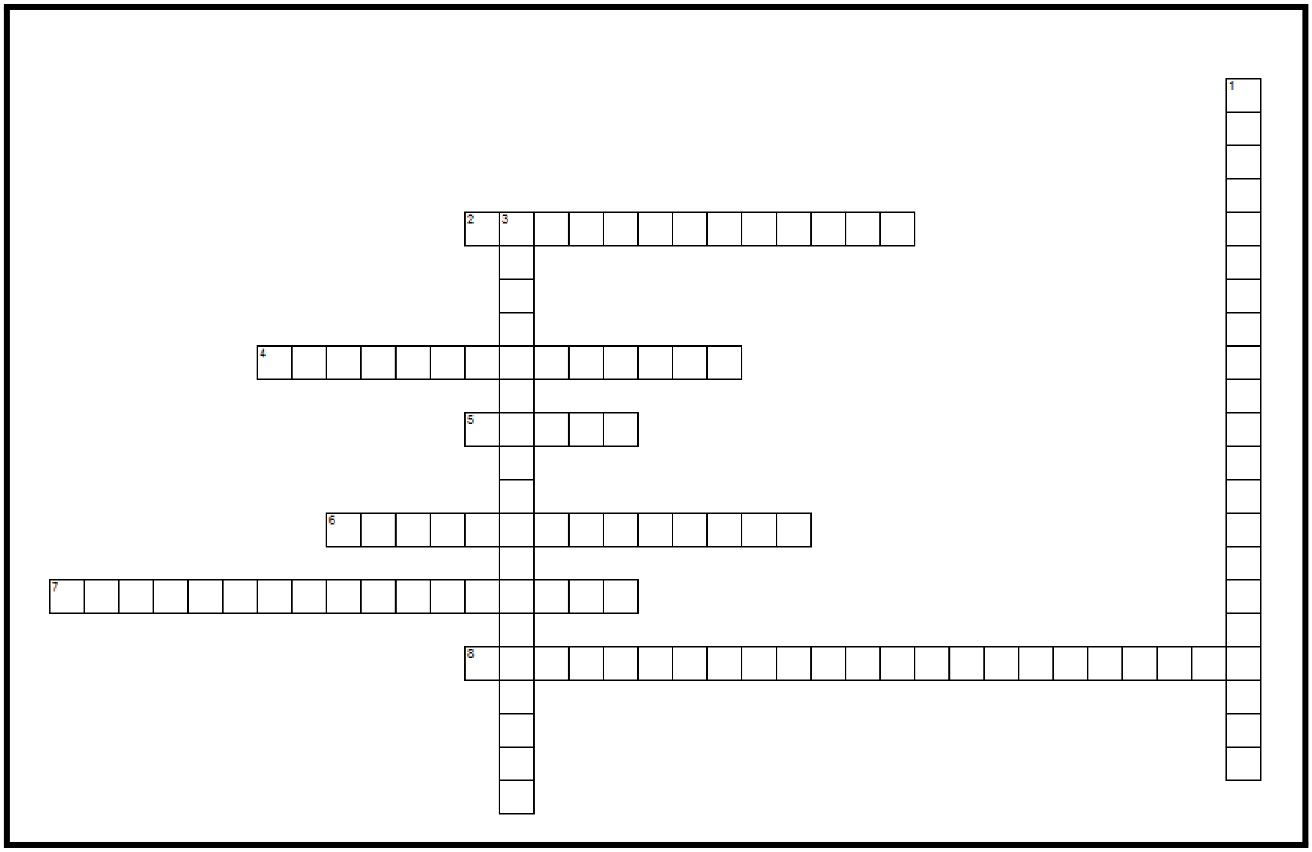
Chapter 8 Review

Below is an outline of the major points covered in Chapter 8. You will need to clearly understand the concepts, terms and calculation presented in Chapter 8. You may need to refer back to previous chapters.

Outline	Chapter 8: Hypothesis Testing	
Basic Terms	Hypothesis and P-value	
Hypothesis Terms	Define:	
	• Null hypothesis • Alternative hypothesis • Region of acceptance • Critical region	• Critical value • Power • Significance level • Statistical significance
Hypothesis Tests	Know the following tests:	
	• One-sampled • Two-sampled • Paired	• Z-tests • T-tests • F-tests
Errors	Know the hypothesis errors	
	• Type I • Type II	
Application	• Understand how to apply statistical hypothesis testing	

Ch. 8 Crossword Puzzle

Complete the following crossword using definitions from Chapter 8.



ACROSS

2. The threshold value containing the boundaries of the regions in which the hypothesis will be accepted or rejected
4. Associated with a contradiction to an assumption to be tested
5. The probability of correctly rejecting the null hypothesis
6. The values of the test which the null hypothesis rejected
7. The probability of incorrectly rejecting the null hypothesis
8. Results are considered statistically significant if the tested sample is inconsistent with the (null) hypothesis

DOWN

1. Associated to an assumption to be tested
3. The values of the test which fail to reject the null hypothesis

Ch. 8 Matching

Match the correct test with the definition:

- One-sample test
- Two-sample test
- Paired test
- Z-test
- T-test
- F-test

	Compare the sample to the population from which the hypothesis is drawn.
	Also referred to the analysis of variance (ANOVA). F-tests are used to determine whether groupings of data by category are meaningful.
	Compare two samples, typically experimental and control samples.
	Compare the means of the data where normality and standard deviation may not be known; these are more relaxed statistical conditions.
	Compare two samples where it is impossible to control all variables that may influence the results.
	Compare the means of the data where normality and standard deviation are known; these are more tightly controlled statistical conditions.

Chapter 8 Practice Problems

Complete the following practice problems to test your understanding. Check your work with the solutions provided.

You have been selected to sit on the jury of an individual accused of murder. Answer the following questions:

1. What is the null hypothesis?

2. What is the alternative hypothesis?

3. What is the outcome if the null is incorrect?

4. What is the outcome if the alternative is correct?

Answer key is found in the Answer Key section.

Chapter 8 Quiz

1. Occurs when the null hypothesis is rejected when, in fact, it is true.
    a. Type I error
    b. Type II error
    c. Power test
    d. Significance level

2. Compare the sample to the population from which the hypothesis is drawn.
    a. One sample test
    b. T-test
    c. Paired test
    d. Z-test

3. Determines the sensitivity of the statistical test.
    a. Significance
    b. Power
    c. T-test
    d. F-test

4. We accept or reject the hypothesis based on the outcome of the p-value.
    a. Null hypothesis
    b. Error type
    c. P-value
    d. Significance level

5. The threshold value containing the boundaries of the regions in which the hypothesis will be accepted or rejected.
    a. Critical region
    b. P-value
    c. Critical value
    d. Null hypothesis

Answer key is found in the Answer Key section.

Appendices

Appendix A: Homework Sets

A.1 Homework Set: Summarizing, Organizing, and Describing Data

Z-score Practice

1. If an IQ test has a mean of 100 and a standard deviation of 15, find the corresponding z score for each IQ.
    a. a. 115 b. 122 c. 93 d. 100 e. 85

2. Which of the following exam grades has a better relative position, (a) or (b)?
    a. A grade of 43 on a test with mean = 40 and s =3.
    b. A grade of 75 on a test with mean= 72 and s = 5.

Percentile Practice

3. Data Set: 5, 12, 15, 16, 20, and 21
    a. Find the percentile rank for each test score in the data set.
    b. What test score corresponds to the 33rd percentile?

Scatter Plots

4. A researcher wishes to determine if a person's age is related to the number of hours he or she exercises per week. The data for the sample are shown here.

Age, x	18	26	32	38	52	59
Hours, y	10	5	2	3	1.5	1

    a. Draw a scatter plot (Graph paper is good, but if you don't have any just do a simple x-axis and y-axis and that will work). Intervals of 1 on the x-axis. Intervals of 5 on the y-axis. Just suggestions.
    b. Describe the data. Is it positive or negative, or no correlation?

5. Give examples of two variables that are positively correlated and two that are negatively correlated.

Order of Operations

Evaluate the expressions for problems 6-10.

6. 3- 4+5
7. 100-(-50)-75
8. 10–(-5)

9. -15+5
10. -16+5(2)

A.2 Homework Set: Regression & Counting Principles

Scatter Plot and Regression Analysis

Subject	Hours, x	Amount, y	xy	x^2	y^2
A	3	48			
B	0	8			
C	2	32			
D	5	64			
E	8	10			
F	5	32			
G	10	56			
H	2	72			
I	1	48			
Totals	36	370			

The table above is data for the number of hours a person exercises and the amount of milk a person consumes per week.

1. Find the values of xy, x^2, and y^2 then find the summation of these values.
2. Make a scatter plot of the values.
3. Calculate the value of r, using the formula for the correlation coefficient
4. What is the relationship between the variables?
5. Write the equation for the line, $y = a + bx$
6. What is the value of y when $x = 15$ hours?

Multiplication Rules for Counting. Find the total number of outcomes in a sequence of events using the multiplication rules.

7. There are four blood types: A, B, AB, and O. Blood can also be Rh+ and Rh-. Finally, a blood donor can be classified as either male or female. How many different ways can a donor have his or her blood labeled?

8. Employees of a large corporation are to be issued special coded identification cards. The code consists of 4 letters of the alphabet. Each letter can be used up to 4 times in the code. How many different ID cards can be issued?

9. A paint manufacturer wishes several different paints. The categories include:

Color	Red, blue, white, black, green, brown, yellow
Type	Latex, Oil
Texture	Flat, semi-gloss, high gloss
Use	Outdoor, indoor

How many different kinds of paint can be made if a person can select one color, one type, one texture, and one use?

10. A stockbroker purchases four different stocks. During the next month, the stock values will either rise, remain the same or decline. How many different possibilities are there?

11. A security analyst would like to have an ID card with two letters followed by two digits. How many different ID cards could be made?

Permutations

12. Suppose a business owner has a choice of five locations in which to establish her business. She decides to rank each location according to certain criteria, such as price of the store and parking facilities. How many different ways can she rank the five cities?

13. What if in problem 13 above the business owner only wants to rank the top three locations. How many different ways can she rank them?

14. How many different ways can a chairperson and an assistant chairperson be selected for a research project if there are seven scientists available?

15. How many different arrangements of three boxcars can be selected from eight boxcars for a train? The order is important since each boxcar is to be delivered to a different location.

16. How many different permutations of the letters in the word statistics are there?

Combination Rule

17. In order to survey the opinions of customers at local malls, a researcher decides to select 5 malls from a total of 12 malls in a specific geographic area. How many different ways can the selection be made?

A.3 Homework Set: Probability & Probability Distributions

1. Determine whether each distribution is a probability distribution.
   a.

x	0	5	10	15	20
$P(x)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

   b.

x	0	2	4	6
$P(x)$	-1	1.5	0.3	0.2

   c.

x	1	2	3	4
$P(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{9}{16}$

   d.

x	2	3	7
$P(x)$	0.5	0.3	0.4

2. State whether the variable is discrete or continuous.
   a. The speed of the car
   b. The number of cups of coffee a fast-food restaurant serves each day
   c. The number of people who play the state lottery each day
   d. The weight of an elephant
   e. The time it takes to complete an exercise session

3. If three coins are tossed, find the mean number of heads that will occur. The probability distribution is as follows: $\mu = E(x) = \sum xP(x)$

No. of Heads, x	0	1	2	3
Probability, $P(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

4. The probability that 0, 1, 2, 3, or 4 people will be placed on hold when they call a radio talk show is shown in the distribution. Find the mean, variance, and standard deviation for the data. The radio station has four phone lines. Remember: $\mu=E(x)=\sum [xP(x)]$

x	0	1	2	3	4
$P(x)$	0.18	0.34	0.23	0.21	0.04

5. If a student randomly guesses at five multiple choice questions, find the probability that the student gets exactly three correct. Each question has five possible choices.
**Hint:* $n = 5, x = 3,$ *and* $p = \frac{1}{5}$ *+ use the binomial*

6. A coin is tossed four times. Find the mean, variance, and standard deviation of the number of heads that will be obtained.

$$n = 4, p = \frac{1}{2}, q = \frac{1}{2}$$

   a. Mean:

   b. Variance:

   c. Standard deviation:

A.4 Homework Set: Normal Probability Distributions

1. Find the area under the normal distribution curve:

2. $P(0 \leq z < 0.56)$

3. $P(z > 0.23)$

4. $P(z \leq -1.92)$

5. $P(-1.43 > z)$

6. What percentage of the area under the normal distribution curve falls within one standard deviation above and below the mean? Two standard deviations? Three standard deviations?

7. A survey found that people keep their television sets an average of 4.8 years. The standard deviation is 0.89 year. If a person decides to buy a new TV set, find the probability that he or she has owned the old set for the following amount of time. Assume the variable is normally distributed.
   a. Less than 2.5 years

   b. Between 3 and 4 years

   c. More than 4.2 years

A.5 Homework Set: Probability Distributions

1. A study of 32 English composition professors showed that they spent, on average, 74.22 minutes correcting a student's term paper. Assume the standard deviation is 23.44 and the margin of error is E=8.12.

    a. Find the 95% confidence interval of the mean time for all composition papers.

    b. If a professor stated that he spent, on average 130 minutes correcting a term paper, what would be your reaction?

2. A study found that 8- to 12- year olds spend an average of $2.86 per day on in-app cell phone purchases. If a sample of 25 children was used, find the 90% confidence interval of the mean. Assume the standard deviation of the sample is $0.15

3. In a sample of 20 customers at the movie theatre, the mean waiting time to order popcorn is 95 seconds. The standard deviation is 21 seconds. Assume that the wait times are normally distributed and construct a 90% confidence interval for the mean wait time of all customers.

4. A sample of 500 nursing applications included 60 from men. Find the 90% confidence interval of the true proportion of men who applied to the nursing program.

A.6 Homework Set: Hypothesis Testing

A survey claims that the average cost of a hotel room in Atlanta is $69.21. In order to see if this is correct, a researcher selects a sample of 30 hotel rooms and finds that the average cost is $68.43. The standard deviation of the population is $3.72. Also, is there enough evidence to reject the claim?

1. State the hypothesis and identify the claim (alternate hypothesis)

    - H_0:

    - H_A:

2. Find the critical value. *Hint: this is two-tailed.* You will have a positive and negative critical value. Draw the normal curve to help you. Alpha is divided by 2. What is the area inside?

3. Compute the test value

4. Make the decision

Answer key is found in the Answer Key section.

Appendix B: Practice Test (Cumulative)

1. Choose the correct order of operations.
   a. Exponents/roots, multiplication/division, parentheses, addition/subtraction
   b. Multiplication/division, parentheses, exponents/roots, addition/subtraction
   c. Parentheses, exponents/roots, multiplication/division, addition/subtraction

2. All of the following are correct except:
   a. $(1 - 3) + 7^2 = 47$
   b. $(2 * 3) + (4 * 1) = 10$
   c. $(2(3 + 4 * 2) = 20$

3. Evaluate $2x + 3$ if $x = 3$.
   a. 0
   b. 8
   c. 9

4. Objective measurements based on numeric values are ______.
   a. Quantitative
   b. Qualitative
   c. Descriptive

5. Subjective measurements based non-numerical values are ________.
   a. Quantitative
   b. Qualitative
   c. Descriptive

6. Quantitative measures objectively describing a data set are ____ statistics.
   a. Quantitative
   b. Inferential
   c. Descriptive

7. Subjective interpretations about a given group are ____ statistics.
   a. Inferential
   b. Qualitative
   c. Descriptive

8. The sum of observations divided by the number of observations equals the _____.
   a. Mean
   b. Median
   c. Mode
   d. Range

9. The sum of observations divided by the number of observations equals ____.
   a. Mean
   b. Median
   c. Mode

10. The middle value of a set of numbers placed in order from smallest to largest is called what?
    a. Mean
    b. Median
    c. Mode
    d. Range

11. The interval between the lowest and highest value is the _____.
    a. Mean
    b. Median
    c. Mode
    d. Range

12. Choose the correct mean for the data set (1, 1, 3, 4, 9, 2, 6, 3, 10, 1).
    a. 1
    b. 4
    c. 3
    d. 9

13. Choose the correct mode for the data set (1, 1, 3, 4, 9, 2, 6, 3, 10, 1).
    a. 1
    b. 4
    c. 3
    d. 9

14. Choose the correct median for the data set (1, 1, 3, 4, 9, 2, 6, 3, 10, 1).
    a. 1
    b. 4
    c. 3
    d. 9

15. Choose the correct range for the data set (1, 1, 3, 4, 9, 2, 6, 3, 10, 1).
    a. 1
    b. 4
    c. 3
    d. 9

16. Data (such as gender, relationship status, or blood type) that can be represented by a categorical number can be all of the following EXCEPT?
    a. Quantitative
    b. Categorical
    c. Qualitative

17. What is a graphical method to represent frequencies of the data set?
    a. Stem and leaf plot
    b. Boxplot
    c. Histogram
    d. Pie chart

18. In a stem-and-leaf plot, the leaf must _____.
    a. Be the last digit of the number
    b. Be the first digit(s) of the number
    c. Never repeat

19. This type of graphical representation is similar to a histogram.
    a. Line graph
    b. Bar graph
    c. Scatter plot
    d. Pie chart

20. An item being investigated is referred to as a(n) what?
    a. Dependent variable
    b. Variable
    c. Independent variable
    d. Normal distribution

21. A variable not subject to change is a(n) _______.
    a. Dependent variable
    b. Variable
    c. Independent variable
    d. Normal distribution

22. A variable that changes in response to another variable is a(n) ________.
    a. Dependent variable
    b. Variable
    c. Independent variable
    d. Normal distribution

23. To determine how exam scores change based on study time and student

gender, it would be best to use which of the following?

a. Crosstabulation and Chi-square
b. Regression analysis
c. Histogram
d. Correlation

24. To determine how exam scores change based on study time, it would be best to use which of the following?
    a. Cross tabulation and Chi-square
    b. Regression analysis
    c. Histogram
    d. Correlation

25. To determine how letter grades vary according to student gender, it would be best to use which of the following?
    a. Crosstabulation and Chi-square
    b. Regression analysis
    c. Histogram
    d. Correlation

26. The strength of a linear relationship between two variables is indicated by what?
    a. Chi-square coefficient
    b. Pearson correlation coefficient
    c. Mode
    d. p-value

27. Choose the value that represents a strong increasing linear relationship.
    a. -0.154
    b. 0.825
    c. -0.825
    d. 0.154

28. Choose the value that represents a strong decreasing linear relationship.
    a. -0.154
    b. 0.825
    c. -0.825
    d. 0.154

29. Choose the value that represents almost no linear relationship but slightly decreasing.
    a. -0.154
    b. 0.825
    c. -0.825
    d. 0.154

30. Choose the value that represents almost no linear relationship but slightly increasing.
    a. -0.154
    b. 0.825
    c. -0.825
    d. 0.154

31. The number of observations is represented by which symbol?
    a. N
    b. Σ
    c. X
    d. Y

32. The regression equation is:
    a. $\hat{y} = \frac{\Sigma Y - b\left(\Sigma X\right)}{N}$
    b. $\hat{y} = a + bx$
    c. $\hat{y} = \frac{N\Sigma XY - \left(\Sigma X\right)\left(\Sigma Y\right)}{N\left(\Sigma X^2\right) - \left(\Sigma X\right)^2}$
    d. $\hat{y} = \frac{a+b}{x}$

33. Categorical variables that have a natural arrangement or order are called:
    a. Nominal
    b. Scale
    c. Ordinal
    d. Discrete

34. Quantitative variables that sparse in space and occur at any interval are:
    a. Nominal
    b. Scale
    c. Ordinal
    d. Discrete

35. Probability is represented by the symbol:
    a. P
    b. Σ
    c. p
    d. A

36. Union is represented by which symbol?
    a. P
    b. Σ
    c. ∩
    d. ∪

37. Intersect is represented by which symbol?
    a. P
    b. Σ
    c. ∩
    d. ∪

38. If the probability of A is 0.55, the probability of B is ______.
    a. 1.00
    b. 1.45
    c. 0.45
    d. 0.55

39. The distribution best exemplified by a simple coin toss is called what?
    a. Bernoulli Distribution
    b. Binomial Distribution
    c. Poisson Distribution
    d. Normal Distribution

40. The distribution best exemplified by a bell curve is the referred to as what?
    a. Bernoulli Distribution
    b. Binomial Distribution
    c. Poisson Distribution
    d. Normal Distribution

41. The distribution used to describe the number of successes within a given series that are independent and have the same probability is the _________.
    a. Bernoulli Distribution
    b. Binomial Distribution
    c. Poisson Distribution
    d. Poisson Binomial Distribution

42. The distribution used to describe the number of successes within a given series that are independent and have different probability is the _______.
    a. Bernoulli Distribution
    b. Binomial Distribution
    c. Poisson Distribution
    d. Poisson Binomial Distribution

43. The distribution that measures the sum of squares of n is the ________.
    a. Bernoulli Distribution
    b. Chi-squared Distribution
    c. Poisson Distribution
    d. Poisson Binomial Distribution

44. The measurement of the ratio of two (normalized) random variables that experience chi-squared distribution is called what?
    a. Bernoulli Distribution
    b. Chi-squared Distribution
    c. F-Distribution
    d. Poisson Binomial Distribution

45. This distribution is used to describe the time between consecutive events that are random in a random, non-reoccurring process.
    a. Bernoulli Distribution
    b. Chi-squared Distribution
    c. F-Distribution
    d. Exponential Distribution

46. Choose the correct statement.
    a. A population is what the researcher wishes to study; a sample is a group selected from the population that holds the highest level of statistical power.
    b. A population is what the researcher wishes to study; a sample is a group selected from the population that the researcher chose at random.
    c. A sample is what the researcher wishes to study; a population is a group selected from the sample that holds the highest level of statistical power.
    d. A sample is what the researcher wishes to study; a population is a group selected from the sample that the researcher chose at random.

47. In this sampling technique, all subsets of the given frame have an equal opportunity of being selected.
    a. Systematic
    b. Panel
    c. Simple Random
    d. Stratified

48. In this sampling technique, sampling is grouped by geography or other defining population characteristics (such as generations).
    a. Systematic
    b. Cluster
    c. Panel
    d. Stratified

49. In this sampling technique, the sample is reduced by asking questions that would narrow down the given sample.
    a. Systematic
    b. Cluster
    c. Panel
    d. Stratified

50. The standard deviation is represented by the following symbol.
    a. σ
    b. X
    c. SD
    d. µ

51. A confidence interval is used to _______.
    a. Indicate the reliability of a statistical estimate
    b. Explain the range of values from lowest to highest
    c. Represent the standard deviation

d. Help the statistician increase accuracy

52. Calculate the 95% confidence interval for the following data set.
Mean = 122
SD = 9
N = 30
    a. 1.64
    b. 119 to 125
    c. 102 to 132
    d. 29

53. Calculate degrees of freedom for the following data set.
Mean = 122
SD = 9
N = 30
    a. 1.64
    b. 119 to 125
    c. 102 to 132
    d. 29

54. Calculate the SEM for the following data set.
Mean = 122
SD = 9
N = 30
    a. 1.64
    b. 2.045
    c. 102 to 132
    d. 29

55. When the null hypothesis is rejected but it is actually true, it is called ______.
    a. Margin of error
    b. Type I error
    c. Z-test
    d. Type II error

56. When the null hypothesis is false but it is not rejected, that is a ________.
    a. Margin of error
    b. Type I error
    c. Z-test
    d. Type II error

57. When the normality and standard deviation of a data set are known, it is best to use a(n):
    a. T-test
    b. Paired test
    c. Z-test
    d. F-test

58. When the normality and standard deviation of a data set are not known, it is best to use a(n):
    a. T-test
    b. Paired test
    c. Z-test
    d. F-test

59. ANOVA is also known as a(n):
    a. T-test
    b. Paired test
    c. Z-test
    d. F-test

60. When we have two samples and it is impossible to control all influencing variables, it is best to use a(n):
    a. T-test
    b. Paired test
    c. One-sample test

d. F-test

61. The equation (1 - β) represents:
    a. Power
    b. Degrees of freedom
    c. Significance level
    d. Critical value

62. Sampling in which every possibility (characteristic) within the given population has an equal opportunity or likelihood of being selected for the sample; produces an unbiased representation of the population.
    a. Non-probability sampling
    b. Probability sampling
    c. Stratified sampling
    d. Cluster sampling

63. Sampling in which not every possibility (characteristic) within the given population has an equal opportunity or likelihood of being selected for the sample; produces an unbiased representation of the population.
    a. Non-probability sampling
    b. Probability sampling
    c. Stratified sampling
    d. Cluster sampling

64. How many steps are involved in selecting a sample?
    a. 2
    b. 6
    c. 3
    d. 5

65. For the following data set, what is the median?
    28, 30, 31, 31, 33, 33, 33, 35, 36, 37, 38, 42
    a. There is no median
    b. 32.5
    c. 33
    d. 33.5

66. For the following data set, what is the mean?
    31, 33, 30, 31, 35, 33, 36, 28, 42, 37, 33
    a. 35
    b. 32.5
    c. 33
    d. 33.5

67. For the following data set, what is the mode?
    31, 33, 30, 31, 35, 33, 36, 28, 42, 37, 33
    a. There is no mode
    b. 32.5
    c. 33
    d. 33.5

68. For the following data set, what is the mode?
    28, 30, 31, 31, 31, 33, 33, 33, 35, 36, 37, 42
    a. There is no mode
    b. 32.5
    c. 33
    d. 33.5

69. In a scatter plot, which variable is on the y-axis?
    a. Independent

b. Dependent
c. Whichever has a greater range
d. Whichever has a greater standard deviation

70. In a scatter plot, which variable is on the x-axis?
    a. Independent
    b. Dependent
    c. Whichever has a greater range
    d. Whichever has a greater standard deviation
71. For a population variance of 144, what is the standard deviation of the population?
    a. 12
    b. 14
    c. 11
    d. 57

72. The assumption to be tested is the _______.
    a. Alternative hypothesis
    b. Null hypothesis
    c. Statistical significance
    d. Power

73. If the significance level is set to 95%, which of the following significance levels will NOT allow us to reject the null hypothesis?
    a. 0.049
    b. 0.065
    c. 0.010
    d. 0.000

74. The distribution calculated by the formula q = 1 – p is
    a. Bernoulli Distribution
    b. Chi-squared Distribution
    c. Poisson Distribution
    d. Poisson Binomial Distribution

75. Given the following data set, find the regression equation.

X	Y
2	4
4	3
6	2
8	1

    a. $-5x + 0.5$
    b. $0.5x - 5$
    c. $0.5x + 5$
    d. $-0.5x + 5$

Answer key is found in the Answer Key section.

Appendix C: Answer Keys

Chapter 1 Solutions

Chapter 1 Review

Crossword Puzzle

Across
3. Quantitative
4. Mode
5. Median

Down
1. Qualitative
2. Range
4. Mean

Practice Problems

172 pounds
174 pounds
169 pounds
98 pounds

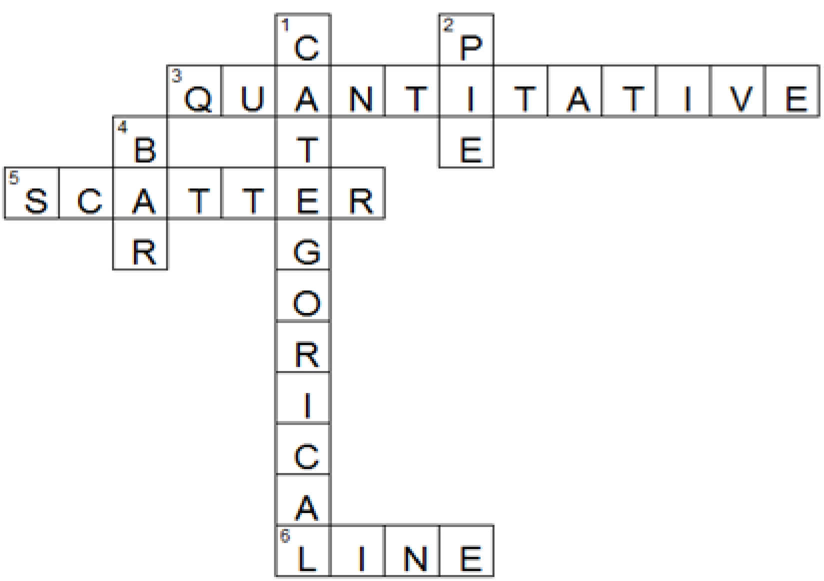
Chapter 1 Quiz

1. C
2. C
3. B
4. B

Chapter 2 Solutions

Chapter 2 Review

Crossword Puzzle



Practice Problems

1.

12	2
13	5
15	0 1 5
16	7 9
17	8 9
18	0 1 2 9
22	0

2. 158 pounds
3. No mode
4. 98 pounds

Chapter 2 Quiz

1. B
2. A
3. C
4. B
5. D

Chapter 3 Solutions

Chapter 3 Review

Fill in the Blank

1. Variable
2. Independent
3. Dependent
4. Random

1. $b = -0.5$
2. $a = 5$
3. $\hat{y} = 5 - 0.5x$
4. $\hat{y} = 2.5$ if $x = 5$

Practice Problems

Chapter 3 Quiz

1. A
2. B
3. A
4. D
5. C

Chapter 4 Solutions

Chapter 4 Review

Table

Event	Probability Equation
$P(A)$	$0 \leq P(A) \leq 1$
$P(A')$	$P(A') = 1 - P(A)$
$P(A \cup B)$ Mutually Exclusive and Non-Mutually Exclusive	Mutually Exclusive $P(A \cup B) = P(A) + P(B)$ Non-Mutually Exclusive $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(A \cap B)$ Independent and Dependent	Independent $P(A \cap B) = P(A) \times P(B)$ Dependent $P(A \cap B) = P(A) \times P(B \mid A)$
$P(B)$	$P(B) = \frac{P(A \cap B)}{P(B)}$

Practice Problems

1. 0.35
2. Yes, event A or B does not depend on the other event. There may be additional factors or events that influence turn out.
3. The carnival is dependent upon temperature which is independent.
4. $P(B) = \frac{P(A \cap B)}{P(B)}$

Chapter 4 Quiz

1. A
2. C
3. A
4. C
5. A

Chapter 5 Solutions

Chapter 5 Review

Distribution	Definition
Bernoulli	This classical probability distribution is easily calculated with the following formula: $q = 1 - p$
Binomial	This type of distribution is used to describe the number of successes within a given series of events that are independent.
Poisson Binomial	This type of distribution is used to describe the number of successes within a given series of events that are independent with a different probability of success.
Poisson	Measures a large number of individual unlikely events that happen within a certain time interval.
Normal	In this distribution, each variable can be measured as the sum of many smaller independent variables.
Chi Squared	This distribution measures the sum of the squares of *n (n = the number of observations)* of independent Gaussian random variables.
F	the measurement of the ratio of two (normalized) random variables that experience chi-squared distribution.
Exponential	This distribution is used to describe the time between consecutive events that are random in a random, non-reoccurring process.
Students T	This distribution is used in the measurement and estimation of unknown means from Gaussian (normally distributed) populations.

Chapter 5 Quiz

1. B
2. A
3. D
4. B
5. A

Chapter 6 Solutions

Chapter 6 Review

Table

Sampling	Definition
Simple	All subsets of the given frame have an equal opportunity of being selected.
Systematic	This technique arranges the population based on some pre-determined order. The sample is then selected at a determined interval from the list.
Stratified	During this sampling technique, the population experiences and is organized into distinct categories or classifications. The sampling frame is then organized into categories also called strata.
Cluster	Sampling is clustered or grouped by geography or other defining population characteristics (such as generations).
Panel	First, a given population is selected from a random list of participants. The sample is then reduced by asking questions that would narrow down the given sample.

Practice Questions

1. Six Steps
    a. Define the population or experimental group.
    b. Specify and select what is to be tested or measured.
    c. Determine the correct method of sampling
    d. Determine the population size.
    e. Test the hypothesis.
    f. Collect the data.
2. Because an entire population would be too large and not practical. Subsets or samples are taken to represent a population.
3. Probability sampling: every possibility (characteristic) within the given population has an equal opportunity or likelihood of being selected for the sample; produces an unbiased representation of the population. Non-probability sampling: not every possibility (characteristic) within the given population has an equal opportunity or likelihood of being selected for the sample; produce a biased or skewed representation of the population.

Chapter 6 Quiz

1. A   2. A   3. B   4. A   5. B

Chapter 7 Solutions

Chapter 7 Review

Fill in the Blank

1. Variance
2. Confidence interval
3. Degrees of freedom
4. T-table
5. Standard error of the mean

Practice Problems

1. 82.4
2. 50.8
3. 7.12
4. 3.19

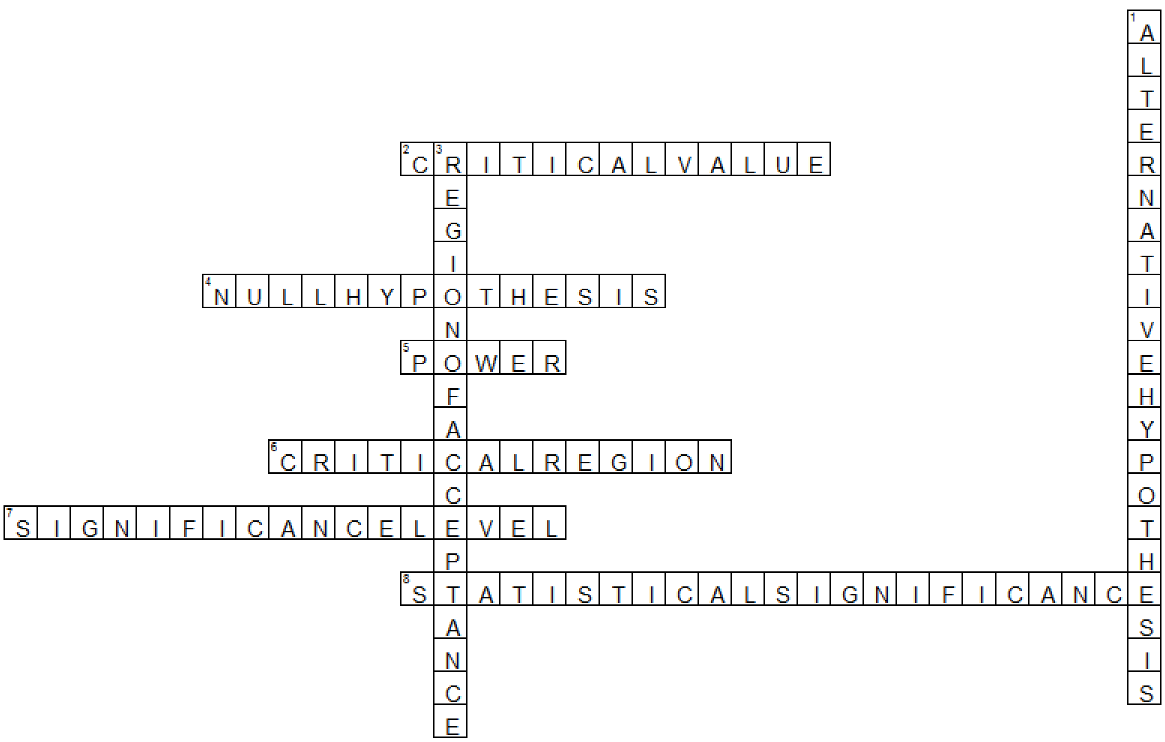
Chapter 7 Quiz

1. A 2. C 3. D 4. A 5. B

Chapter 8 Solutions

Chapter 8 Review

Crossword Puzzle



Statistics

Matching

Match the correct test with the definition:

One-sample test	Compare the sample to the population from which the hypothesis is drawn.
F-test	Also referred to the analysis of variance, (ANOVA). F-tests are used to determine whether groupings of data by category are meaningful.
Two-sample test	Compare two samples, typically experimental and control samples.
T-test	Compare the means of the data where normality and standard deviation may not be known; these are more relaxed statistical conditions.
Paired test	Compare two samples where it is impossible to control all variables that may influence the results.
Z-test	Compare the means of the data where normality and standard deviation are known; these are more tightly controlled statistical conditions.

Practice Problems

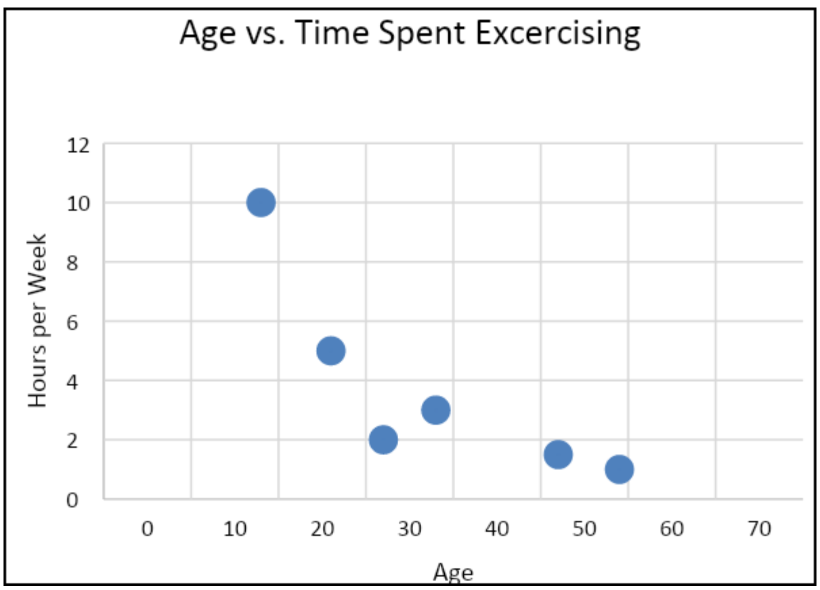
1. The suspect is not guilty
2. The suspect is guilty
3. The suspect really was guilty and has gone free
4. The suspect really was not guilty and has gone to prison

Chapter 8 Quiz

1. A
2. A
3. B
4. C
5. C

Appendix A (Homework Sets) Solutions

A.1 Homework Set Solutions

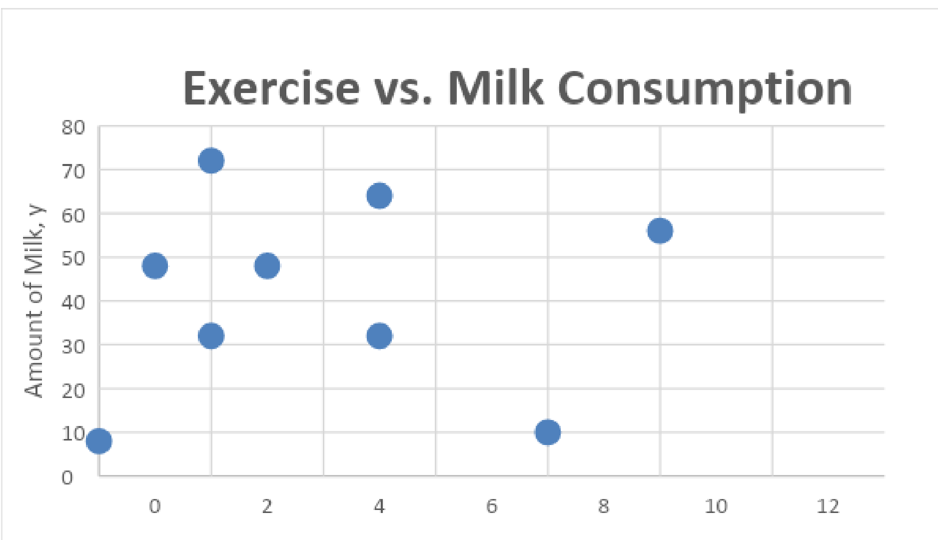
1. See below.
    a. 1
    b. 1.47
    c. -0.47
    d. 0
    e. -1
2. The grade of 43 has a better relative position.
3. See below.
    a. The percentile rank for …
        i. 5 is 8. A student whose score was 5 did better than 8% of the class.
        ii. 12 is 25. A student whose score was 12 did better than 25% of the class.
        iii. 15 is 42. A student whose score was 15 did better than 42% of the class.
        iv. 16 is 58. A student whose score was 16 did better than 58% of the class.
        v. 20 is 75. A student whose score was 20 did better than 75% of the class.
        vi. 21 is 92. A student whose score was 21 did better than 92% of the class.
    b. 12 corresponds to the 33rd percentile
4. See below.
    a. 

    b. This is negative correlation.
5. Answers will vary.
6. 6
7. 75
8. 15
9. -10
10. -6

A.2 Homework Set Solutions

1. See table below.

Subject	Hours, x	Amount, y	xy	x^2	y^2
A	3	48	144	9	2304
B	0	8	0	0	64
C	2	32	64	4	1024
D	5	64	320	25	4096
E	8	10	80	64	100
F	5	32	160	25	1024
G	10	56	560	100	3136
H	2	72	144	4	5184
I	1	48	48	1	2304
Totals	36	370	1520	232	19236

2. 

3. $r = 0.0672$
4. This is a very weak positive relationship between the variables.
5. $\hat{y} = 38.38 + 0.68x$
6. 48.58

7. 16
8. 456,976
9. 84
10. 81
11. 67,600
12. 120
13. 60
14. 42
15. 336
16. 30420
17. 792

A.3 Homework Set Solutions

1. See below.
   a. Yes
   b. No
   c. Yes
   d. No
2. See below.
   a. Continuous
   b. Continuous
   c. Discrete
   d. Continuous
   e. Continuous
3. $\mu = E(x) = 1.5$
4. See below.
   a. Mean: 2.43
   b. Variance: 1.97
   c. Standard Deviation: 1.40
5. $P(x = 3) = 0.0512 \approx 5\%$
6. See below.
   a. Mean: 2
   b. Variance: 1
   c. Standard Deviation

A.4 Homework Set Solutions

1. $P(0 \leq z < 0.56) = 0.2123$
2. $P(z > 0.23) = 0.4090$
3. $P(z \leq -1.92) = 0.9726$
4. $P(-1.43 > z) = 0.0764$
5. $1\sigma = 68\%,\ 2\sigma = 95\%$,and $3\sigma = 99.77\%$
6. See below.
   a. 0.0049
   b. 0.1624
   c. 0.7486

A.5 Homework Set Solutions

1. See below.
   a. (66.1, 82.35)
   b. It would be highly unlikely, since this is far larger than the average of 74.22 minutes, as shown by the confidence interval.
2. (2.81, 2.91)
3. (86.9, 103.1)
4. (0.096, 0.144)

A.6 Homework Set Solutions

1. See below.
   a. $H_0: \mu = 69.21$
   b. $H_A: \mu \neq 69.21$
2. $z_{\alpha/2} = \pm 1.96$
3. $z = \frac{68.43 - 69.21}{\frac{3.72}{\sqrt{30}}} \approx -7.04$
4. $z \leq z_{\frac{\alpha}{2}} \Longrightarrow -7.04 \leq -1.96$

   Reject the null hypothesis. There is enough evidence to reject the claim that the average cost of a hotel stay in Atlanta is $69.21.

Appendix B (Practice Test) Solutions

1. c
2. c
3. c
4. a
5. b
6. c
7. a
8. a
9. a
10. b
11. d
12. b
13. a
14. c
15. d
16. a
17. c
18. a
19. b
20. b
21. c
22. a
23. b
24. d
25. a
26. b
27. b
28. c
29. a
30. d
31. a
32. b
33. c
34. d
35. a
36. d
37. c
38. c
39. a
40. d
41. b
42. d
43. b
44. c
45. d
46. a
47. c
48. b
49. c
50. a
51. a
52. b
53. d
54. a
55. b
56. d
57. c
58. a
59. d
60. b
61. a
62. b
63. a
64. b
65. c
66. d
67. c
68. a
69. b
70. a
71. a
72. b
73. b
74. a
75. d

Appendix D: Distribution Tables

D.1 Standard Normal Distribution

Table of Cumulative Probability for Standard Normal Distribution

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.00	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.001	0.001
-2.90	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.80	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.002	0.0019
-2.70	0.0035	0.0034	0.0033	0.0032	0.0031	0.003	0.0029	0.0028	0.0027	0.0026
-2.60	0.0047	0.0045	0.0044	0.0043	0.0041	0.004	0.0039	0.0038	0.0037	0.0036
-2.50	0.0062	0.006	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.40	0.0082	0.008	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.30	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.20	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.011
-2.10	0.0179	0.0174	0.017	0.0166	0.0162	0.0158	0.0154	0.015	0.0146	0.0143
-2.00	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.90	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.025	0.0244	0.0239	0.0233
-1.80	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.70	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.60	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.50	0.0668	0.0655	0.0643	0.063	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.40	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.30	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.20	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.102	0.1003	0.0985
-1.10	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.123	0.121	0.119	0.117
-1.00	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.90	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.166	0.1635	0.1611
-0.80	0.2119	0.209	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.70	0.242	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.60	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.50	0.3085	0.305	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.281	0.2776
-0.40	0.3446	0.3409	0.3372	0.3336	0.33	0.3264	0.3228	0.3192	0.3156	0.3121
-0.30	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.352	0.3483
-0.20	0.4207	0.4168	0.4129	0.409	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.10	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.00	0.5	0.496	0.492	0.488	0.484	0.4801	0.4761	0.4721	0.4681	0.4641

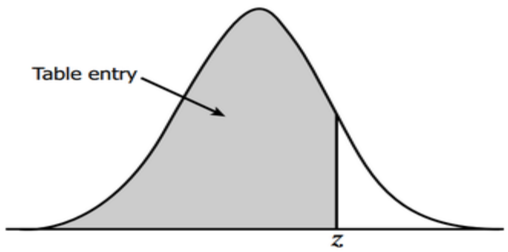


Table of Cumulative Probability for Standard Normal Distribution

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

D.2 Student's t-Distribution

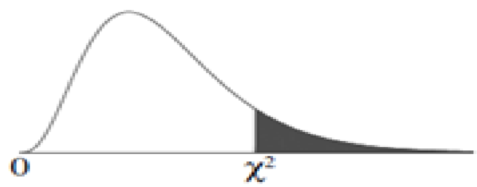
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.01
df								
1	0.0000	1.0000	1.3764	1.9626	3.0777	6.3138	12.7062	63.6567
2	0.0000	0.8165	1.0607	1.3862	1.8856	2.9200	4.3027	9.9248
3	0.0000	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	5.8409
4	0.0000	0.7407	0.9410	1.1896	1.5332	2.1318	2.7764	4.6041
5	0.0000	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	4.0321
6	0.0000	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.7074
7	0.0000	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	3.4995
8	0.0000	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	3.3554
9	0.0000	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	3.2498
10	0.0000	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	3.1693
11	0.0000	0.6974	0.8755	1.0877	1.3634	1.7959	2.2010	3.1058
12	0.0000	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	3.0545
13	0.0000	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	3.0123
14	0.0000	0.6924	0.8681	1.0763	1.3450	1.7613	2.1448	2.9768
15	0.0000	0.6912	0.8662	1.0735	1.3406	1.7531	2.1314	2.9467
16	0.0000	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.9208
17	0.0000	0.6892	0.8633	1.0690	1.3334	1.7396	2.1098	2.8982
18	0.0000	0.6884	0.8620	1.0672	1.3304	1.7341	2.1009	2.8784
19	0.0000	0.6876	0.8610	1.0655	1.3277	1.7291	2.0930	2.8609
20	0.0000	0.6870	0.8600	1.0640	1.3253	1.7247	2.0860	2.8453
21	0.0000	0.6864	0.8591	1.0627	1.3232	1.7207	2.0796	2.8314
22	0.0000	0.6858	0.8583	1.0614	1.3212	1.7171	2.0739	2.8188
23	0.0000	0.6853	0.8575	1.0603	1.3195	1.7139	2.0687	2.8073
24	0.0000	0.6848	0.8569	1.0593	1.3178	1.7109	2.0639	2.7969
25	0.0000	0.6844	0.8562	1.0584	1.3163	1.7081	2.0595	2.7874
26	0.0000	0.6840	0.8557	1.0575	1.3150	1.7056	2.0555	2.7787
27	0.0000	0.6837	0.8551	1.0567	1.3137	1.7033	2.0518	2.7707
28	0.0000	0.6834	0.8546	1.0560	1.3125	1.7011	2.0484	2.7633
29	0.0000	0.6830	0.8542	1.0553	1.3114	1.6991	2.0452	2.7564
30	0.0000	0.6828	0.8538	1.0547	1.3104	1.6973	2.0423	2.7500
50	0.0000	0.6794	0.8489	1.0473	1.2987	1.6759	2.0086	2.6778
100	0.0000	0.6770	0.8452	1.0418	1.2901	1.6602	1.9840	2.6259
500	0.0000	0.6750	0.8423	1.0375	1.2832	1.6479	1.9647	2.5857

1000	0.0000	0.6747	0.8420	1.0370	1.2824	1.6464	1.9623	2.5808

D.3 Binomial

p =		*0.05*	*0.1*	*0.15*	*0.2*	*0.25*	*0.3*	*0.35*	*0.4*	*0.45*	*0.5*
n	*k*										
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9928	0.9720	0.9393	0.8960	0.8438	0.7840	0.7183	0.6480	0.5748	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6563
	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

D.4 Chi-Square



The shaded area is equal to α for $\chi^2 = \chi^2_{\alpha}$.

	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.0002	0.0010	0.0039	0.0158	2.7055	3.8415	5.0239	6.6349
2	0.0201	0.0506	0.1026	0.2107	4.6052	5.9915	7.3778	9.2103
3	0.1148	0.2158	0.3518	0.5844	6.2514	7.8147	9.3484	11.3449
4	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877	11.1433	13.2767
5	0.5543	0.8312	1.1455	1.6103	9.2364	11.0705	12.8325	15.0863
6	0.8721	1.2373	1.6354	2.2041	10.6446	12.5916	14.4494	16.8119
7	1.2390	1.6899	2.1673	2.8331	12.0170	14.0671	16.0128	18.4753
8	1.6465	2.1797	2.7326	3.4895	13.3616	15.5073	17.5345	20.0902
9	2.0879	2.7004	3.3251	4.1682	14.6837	16.9190	19.0228	21.6660
10	2.5582	3.2470	3.9403	4.8652	15.9872	18.3070	20.4832	23.2093
11	3.0535	3.8157	4.5748	5.5778	17.2750	19.6751	21.9200	24.7250
12	3.5706	4.4038	5.2260	6.3038	18.5493	21.0261	23.3367	26.2170
13	4.1069	5.0088	5.8919	7.0415	19.8119	22.3620	24.7356	27.6882
14	4.6604	5.6287	6.5706	7.7895	21.0641	23.6848	26.1189	29.1412
15	5.2293	6.2621	7.2609	8.5468	22.3071	24.9958	27.4884	30.5779
16	5.8122	6.9077	7.9616	9.3122	23.5418	26.2962	28.8454	31.9999
17	6.4078	7.5642	8.6718	10.0852	24.7690	27.5871	30.1910	33.4087
18	7.0149	8.2307	9.3905	10.8649	25.9894	28.8693	31.5264	34.8053
19	7.6327	8.9065	10.1170	11.6509	27.2036	30.1435	32.8523	36.1909
20	8.2604	9.5908	10.8508	12.4426	28.4120	31.4104	34.1696	37.5662
21	8.8972	10.2829	11.5913	13.2396	29.6151	32.6706	35.4789	38.9322
22	9.5425	10.9823	12.3380	14.0415	30.8133	33.9244	36.7807	40.2894
23	10.1957	11.6886	13.0905	14.8480	32.0069	35.1725	38.0756	41.6384
24	10.8564	12.4012	13.8484	15.6587	33.1962	36.4150	39.3641	42.9798
25	11.5240	13.1197	14.6114	16.4734	34.3816	37.6525	40.6465	44.3141
26	12.1981	13.8439	15.3792	17.2919	35.5632	38.8851	41.9232	45.6417
27	12.8785	14.5734	16.1514	18.1139	36.7412	40.1133	43.1945	46.9629
28	13.5647	15.3079	16.9279	18.9392	37.9159	41.3371	44.4608	48.2782
29	14.2565	16.0471	17.7084	19.7677	39.0875	42.5570	45.7223	49.5879
30	14.9535	16.7908	18.4927	20.5992	40.2560	43.7730	46.9792	50.8922