Class Notes

To Accompany Elementary Statistics 6e, by Larson and Farber

Examples used in these notes are taken from the textbook and written by Larson and Farber

Instructor Tim Busken

Table of Contents

1 Introduction to Statistics

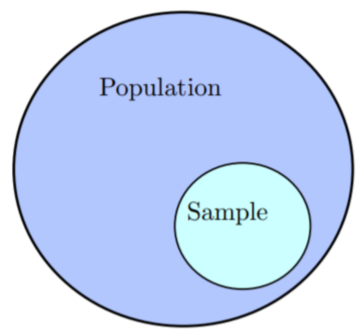
1.1 An overview of Statistics

Learning objectives:

1. The definition of statistics
2. How to distinguish between a population and a sample; and a parameter and a statistic
3. How to distinguish between descriptive statistics and inferential statistics

---

- **Data** consist of information coming from observations, counts, measurements, or responses.
- **Statistics** is the science of collecting, organizing, analyzing, and interpreting data in order to make decisions.
- There are two types of data sets we will use in this course — **populations** and **samples**.
- A **Population** is the collection of all outcomes, responses, measurements, or counts that are of interest.
- A **Sample** is a subset, or part, of the population.



A goal of statisticians is to draw reasonable conclusions about population characteristics based off of the analysis of sample data. It is extremely important to obtain sample data that are representative of the population from which the data are drawn.

Example In a recent survey from 2012, 1500 new college graduates were asked if they had taken out student loans to finance their education. 69% said yes. Identify the population and the sample. Describe the data set.

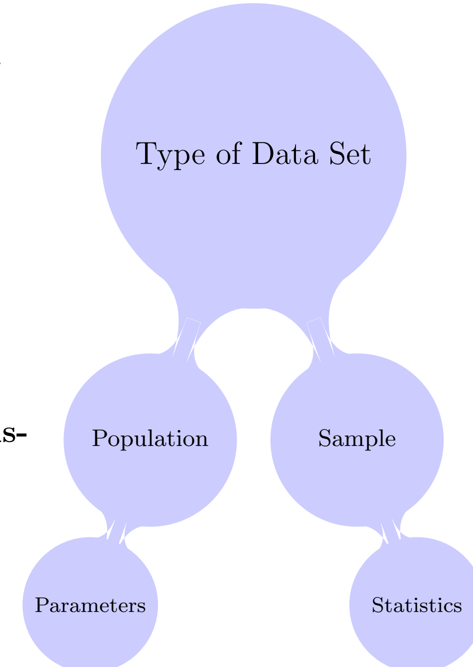
Try These! Identify the population and the sample.

1. Thirty nurses working in the San Diego area were surveyed concerning their opinions of working conditions.

2. A survey of 200 U.S. adults found that 32% drink covfefe daily.

- A **parameter** is a number that describes a population characteristic. For example, the average age of all people in the U.S. is a parameter.

- A **statistic** is a number that describes a sample characteristic. For example, the average age of people from a sample of three states.



Try These! Decide whether the numerical value describes a population parameter or a sample statistic.

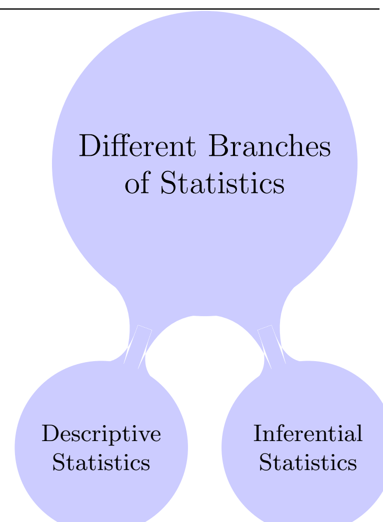
3. In 2012, Major League Baseball teams spent a total of $2,940,657,192 on players salaries.

4. In a survey of 1000 U.S. adults, 74% said they care about the next upcoming presidential election.

5. In a recent study of math majors at a university, 10 students were minoring in physics.

6. The 2182 students who accepted admission offers to Northwestern University in 2009 had an average SAT score of 1442.

- The study of statistics has two major branches: **descriptive statistics** and **inferential statistics**.
- **Descriptive Statistics** Involves organizing, summarizing, and displaying data.
- **Inferential Statistics** Involves using sample data to draw conclusions about a population.



Try These! In each of these statements, tell whether descriptive or inferential statistics have been used.

7. Expenditures for the cable industry were $5.66 billion in 1996 (Source: USA TODAY).

8. The report by the Medicare Office of the Actuary estimated that health spending will grow by an average of 5.8 percent a year through 2020, compared to 5.7 percent without the health overhaul.

9. The mean travel time to work (years 2008–2012) for San Diego County workers age 16+ is 24.2 minutes. (Source: census.gov).

10. Allergy therapy makes bees go away (Source: Prevention).

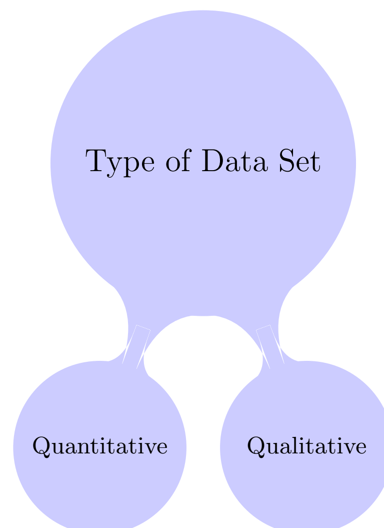
1.2 Data Classification

Learning objectives:

1. How to distinguish between qualitative and quantitative data
2. How to distinguish between a discrete variable and a continuous variable
3. How to classify data with respect to the four levels of measurement: nominal, ordinal, interval and ratio

---

- **Qualitative Data** Consists of attributes, labels, categories or nonnumerical entries, for example hair color, place of birth or major.
- **Quantitative data** Numerical measurements or counts, for example age, height or temperature



<u>Try These!</u> Classify each variable as qualitative or quantitative.

1. A Gallup poll asked Americans, "Which subject (Math, English, Art,...), if any, has been the most valuable in your life?"
2. The number of donuts sold each week at Dunkin Donuts
3. The different flavors of donuts at Dunkin Donuts
4. Heights of NBA atheletes
5. The numbers on the back of NBA atheletes' uniforms
6. Bank account numbers of each person in this class
7. Zip code of each person's residence.

- **DISCRETE VARIABLE** a variable that assumes values that can be counted. A discrete variable has gaps between values the variable can take on. Ex's of discrete variables: number of people, number of donuts, performance rating (1, 2, 3, 4 or 5), shoe size (5.5, 6, 6.5, 7, 7.5, 8, ...).
- **CONTINUOUS VARIABLE** a variable that can be equal to any number between any two specific values. A variable that can take on any number in a connected interval or set of numbers. Ex's of continuous variables: time, temps, volume.

<u>Try These!</u> Classify each variable as discrete or continuous.

8. The number of people who buy a coffee from Starbucks today.
9. The volume of water (in cubic feet) of each of the Great Lakes.
10. The number of drones Amazon.com wants to have delivering orders by 2020.
11. The amount of time it takes to answer this question.
12. The number of nuclear power plants in the world.

1.3 Data Collection and Experimental Design

Learning objectives:

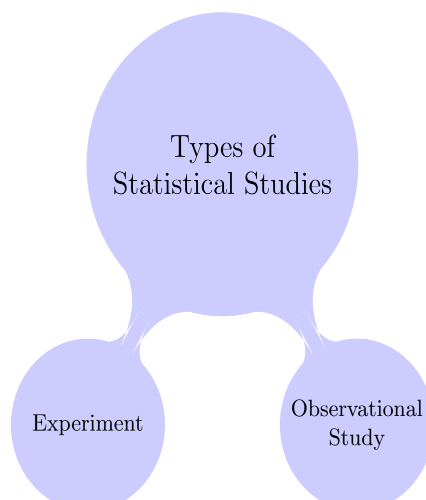
1. How to design a statistical study and how to distinguish between an observational study and an experiment
2. How to collect data by survey or simulation
3. How to design an experiment
4. How to create a sample using random sampling, simple random sampling, stratified sampling, cluster sampling, and systematic sampling

---

How to Design a Statistical Study

1. Identify the variable(s) of interest (the focus) and the population of the study.
2. Develop a detailed plan for collecting data. If you use a sample, make sure the sample is representative of the population.
3. Collect the data.
4. Describe the data using descriptive statistics techniques.
5. Interpret the data and make decisions about the population using inferential statistics.
6. Identify any possible errors.

Types of Statistical Studies



A statistical study can be categorized as observational study or an experiment.

Observational study A researcher observes and measures characteristics of interest of part of a population, but does not change existing conditions.

Experiment In performing an experiment, a treatment is applied to part of a population, called a treatment group, and responses are observed. Another part of the population may be used as a control group, in which no treatment is applied.(The subjects in the treatment and control groups are called experimental units.) In many cases, subjects in the control group are given a placebo, which is a harmless, fake treatment, that is made to look like the real treatment. The responses of the treatment group and control group can then be compared and studied to help determine if the treatment was effective, or if the treatment causes side effects. In most cases, it is a good idea to use the same number of subjects for each group.

Try These! Decide whether the study is an observational study or an experiment.

1. Researchers demonstrated in people at risk for cardiovascular disease that 2000 milligrams per day of acetyl-L-carnitine over a 24-week period lowered blood pressure and improved insulin resistance.

2. Researchers conduct a study to determine where a drug used to treat hypothyroidism works better when taken in the morning or when taken at bedtime. To perform the study, 90 patients are given one pill to take in the morning and one pill to take in the evening (one containing the drug and the other a placebo). After 3 months, patients are instructed to switch pills.

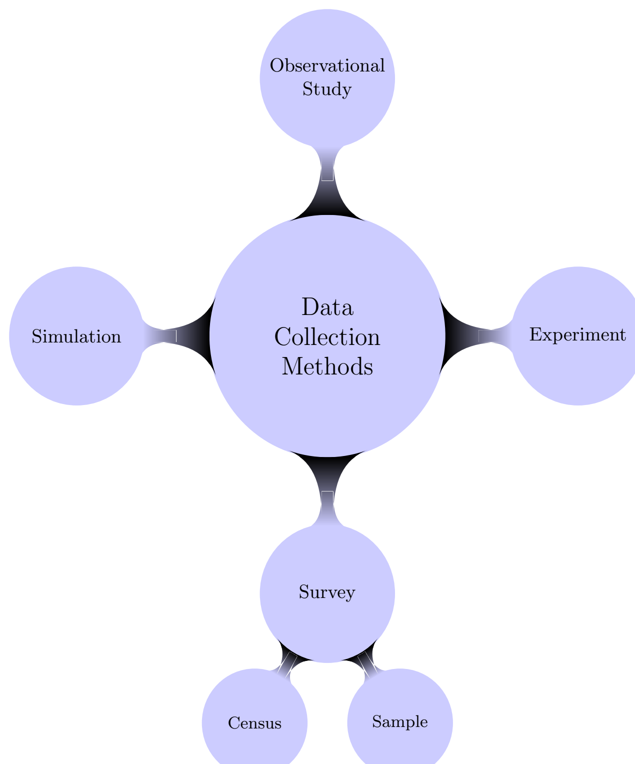
3. Researchers conduct a study to determine the number of falls women had during pregnancy. To perform the study, researchers contacted 3997 women who had recently given birth and asked them how many times they fell during their pregnancies.

Methods of Data Collection

There are several ways to collect data. Often the focus of the study dictates the best way to collect data.

The four ways we collect data are:

1. by conducting an observational study
2. by conducting an experimental study
3. by conducting a simulation
4. by conducting a survey



- ***Simulation*** uses a mathematical or physical model to reproduce the conditions of a situation or process, often by use of computers. Simulations allow you to study situations that are impractical or too dangerous to create in real life, and they often save time and money. For instance, Automobile manufacturers use simulations with dummies to study the effects of crashes on humans.
- A ***Survey*** is an investigation of one or more characteristics of a population. The most common types of surveys are done by interview, Internet, phone, or mail. An official count or survey of a population is called a ***Census***.

Consider the following statistical studies. Which method of data collection would you use to collect data for each study?

1. A study of the effect of changing flight patterns on the number of airplane accidents.

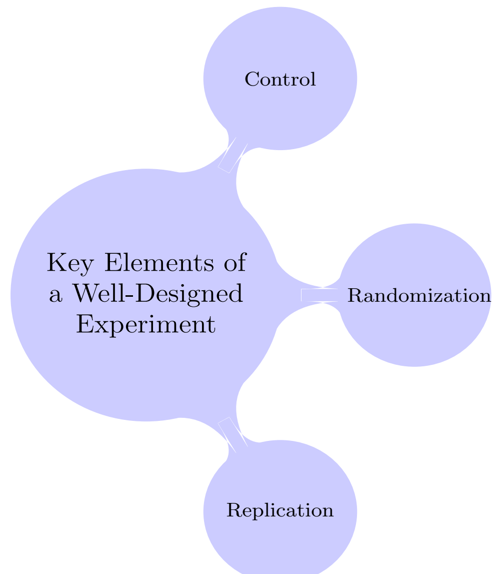
2. A study of the effect of eating carrots on lowering blood pressure.

3. A study of how sixth grade students solve a puzzle.

4. A study of U.S. residents' approval rating of the U.S. president.

Experimental Design

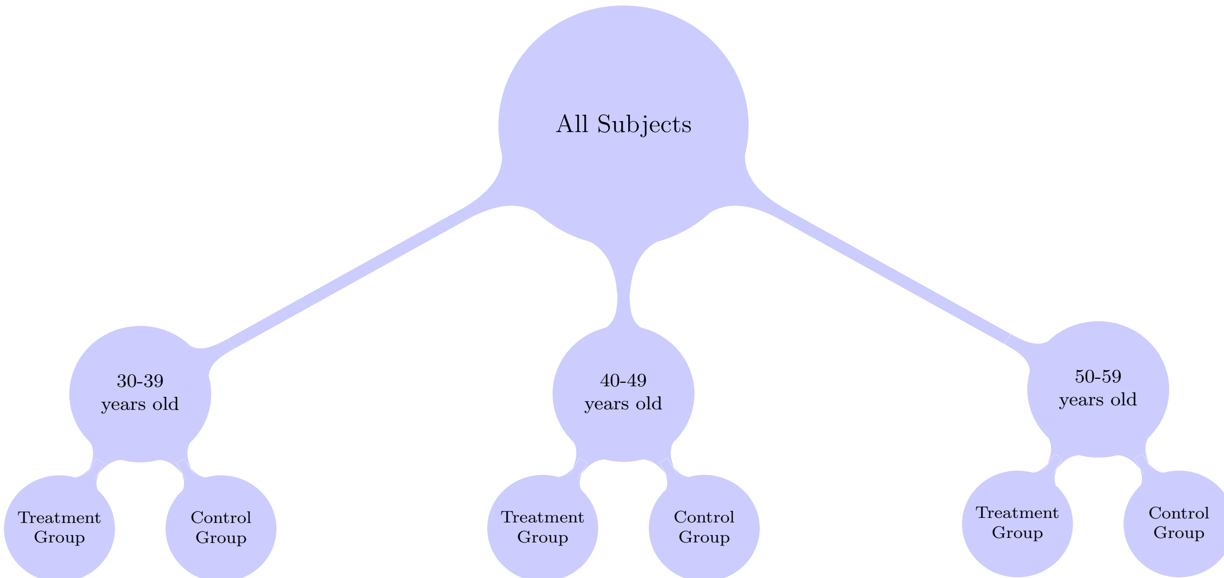
To produce meaningful unbiased results, experiments should be carefully designed and executed. It is important to know what steps should be taken to make the results of an experiment valid. ***Three key elements of a well-designed experiment are control, randomization, and replication.*** Because experimental results can be ruined by a variety of factors, being able to control these influential factors is important. One such factor is a confounding variable. A ***confounding variable*** occurs when an experimenter cannot tell the difference between the effects of different factors on the variable.



For instance, to attract more customers, a coffee shop owner experiments by remodeling her shop using bright colors. At the same time, a shopping mall nearby has its grand opening. If business at the coffee shop increases, it cannot be determined whether it is because of the new colors or the new shopping mall. The effects of the colors and the shopping mall have been confounded. Another factor that can affect experimental results is the placebo effect. The ***placebo effect*** occurs when a subject reacts favorably to a placebo when in fact the subject has been given a fake treatment. To help control or minimize the placebo effect, a technique called blinding can be used.

Blinding is a technique where the subjects do not know whether they are receiving a treatment or a placebo. In a ***double-blind experiment***, neither the experimenter nor the subjects know if the subjects are receiving a treatment or a placebo. The experimenter is informed after all the data have been collected. This type of experimental design is preferred by researchers.

Another element of a well-designed experiment is randomization. ***Randomization*** is a process of randomly assigning subjects to different treatment groups. In a completely randomized design, subjects are assigned to different treatment groups through random selection.



In some experiments, it may be necessary for the experimenter to use blocks, which are groups of subjects with similar characteristics. A commonly used experimental design is a randomized block design. To use a ***randomized block design***, the experimenter divides the subjects with similar characteristics into blocks, and then, within each block, randomly assign subjects to treatment groups. For instance, an experimenter who is testing the effects of a new weight loss drink may first divide the subjects into age categories such as 30–39 years old, 40– 49 years old, and over 50 years old, and then, within each age group, randomly assign subjects to either the treatment group or the control group (see figure at the left).

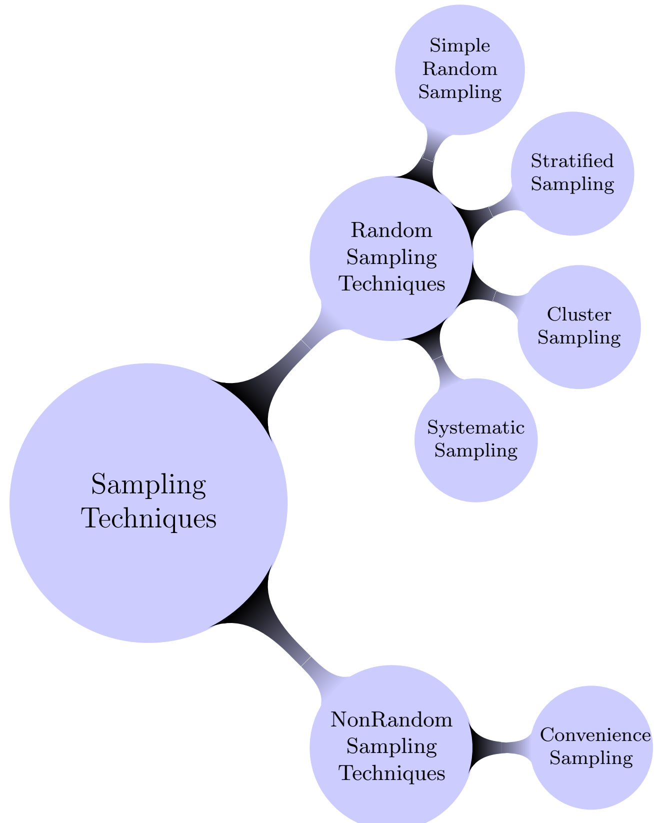
Another type of experimental design is a ***matched-pairs design***, where subjects are paired up according to a similarity. One subject in each pair is randomly selected to receive one treatment while the other subject receives a different treatment. For instance, two subjects may be paired up because of their age, geographical location, or a particular physical characteristic. Sample size, which is the number of subjects in a study, is another important part of experimental design. To improve the validity of experimental results, replication is required.

Replication is the repetition of an experiment under the same or similar conditions.

Sampling Techniques

In a Random Sample, every member of the population has an equal chance of being selected.

The five sampling techniques presented here are: Simple Random Sampling, Stratified Sampling, Cluster Sampling, Systematic Sampling and Convenience Sampling.



Simple Random Sample
Every possible sample of the same size has the same chance of being selected. One way to collect a simple random sample is to assign a number to each member of the population. Random numbers can then be generated by a random number table, a software program or a calculator. Members of the population that correspond to these numbers become members of the sample.

Example There are 37 students in Mr. Busken's Friday stats class. You wish to form a sample of five students to answer some survey questions. Select the students who will belong to the simple random sample. Use the calculator's "randInt" function.

Stratified Sample
Divide a population into groups (strata) so that subjects within the same subgroup share the same characteristics (such as gender or age bracket) and select a random sample from each group.

Example To collect a stratified sample of Mr. Busken's students, you could divide the students up into age groups, then randomly select a couple people from each age group.

Cluster Sample
Divide the population into groups (clusters) and select *all* of the members in one or more, but not all, of the clusters.

Example To collect a cluster sample of Mr. Busken's students, you could divide the students up into age groups, then select *all* the students in one or more age brackets.

Systematic Sample
Choose a starting value at random. Then choose every k^{th} member of the population.

Example To collect a systematic sample of Mr. Busken's students, you could number the students 1 through 37, randomly choose a starting number, then select every 4th person.

Convenience Sample
Choose only members of the population that are easy to get Often leads to biased studies (not recommended).

Try These! Identify the sampling technique used.

1. Using random digit dialing, researchers ask 300 U.S. adults if they care about the next election.

2. A student asks 12 friends in her dorm to participate in a psychology experiment.

3. A study of water quality in Rio de Janeiro, Brazil randomly selects and collects water samples from 33 of the 55 communities from the Rio metropolitan area.

4. Law enforcement officials at the Kaaboo concert stop and check the driver of every third vehicle for their blood alcohol content.

5. Twenty-six students are randomly selected from each grade level at a high school and surveyed about their study habits.

6. A journalist interviews 150 people waiting at an airport baggage claim and asks them how satisfied they are with the airline they flew.

2 Descriptive Statistics

2.1 Frequency Distributions and Their Graphs

Learning objectives:

1. Learn the terminology that goes along with a Frequency Distribution Table.
2. Construct a Frequency Distribution, Construct a Relative Frequency Distribution, Construct a Cumulative Frequency Distribution
3. Construct Frequency Histograms, Frequency Polygons, Relative Frequency Histograms, and Ogives

---

Defn: A ***Frequency Distribution*** is a two-column table that summarizes a list of measurements or responses. Column 1 of the frequency distribution table is a list of classes. These classes can be single numbers, intervals of numbers or categories. Column 2 of the frequency distribution is the frequency column. The numbers in this column represent the number of measurements or responses that were measurements or responses from their corresponding numeric classes or category classes.

EXAMPLE 1: Suppose that in a previous semester, a random sample of my students were asked what their age was. The frequency distribution below summarizes the list of the ages that make up the sample. The first column of the table is a list of age intervals, or age classes. The frequency number in row 2 of the table is 68. This means there were 68 students in the sample between and including ages 15 and 24. From the next row of the frequency distribution, we see that there were 72 students who said they were between and including ages 25 and 34. There were 34 students between and including ages 35 and 44, 16 students between and including ages 45 and 54, 8 students between and including ages 55 and 64, and 2 students between and including ages 65 and 74.

Classes (Age Intervals)	Frequencies, f (counts)
15—24	68
25—34	72
35—44	34
45—54	16
55—64	8
65—74	2

EXAMPLE 2: 10 randomly selected history students were asked how many absences they had in their history class last semester. The 10 responses are summarized in the table.

Classes (absences)	Frequencies, f (days absent)
0	1
1	3
2	4
3	0
4	1
5	1

You can tell 1 student had zero absences, 3 students had one absence, 4 students had two absences, and so on. Example 2 gives us an example of a frequency distribution where the classes are single numbers. Example 1 gives us an example of a frequency distribution whose classes are number intervals. It is also possible to make a ***categorical frequency distribution*** with qualitative data. The next table gives us and example of a categorical frequency distribution.

EXAMPLE 3: 100 randomly selected students were asked their blood type.

Classes (blood type)	Frequencies, f
A	40
B	11
AB	4
O	45

EXAMPLE 4: Earthquake magnitudes for the month of July were randomly selected. The data is summarized in the frequency distribution.

Classes (magnitudes)	Frequencies, f (Number of Earthquakes)
0.0—0.9	5
1.0—1.9	15
2.0—2.9	17
3.0—3.9	13
4.0—4.9	45
5.0—5.9	35
6.0—6.9	10
7.0—7.9	5

EXAMPLE 5: Gasoline prices of regular unleaded from 30 gas stations were randomly selected. The data is summarized in the frequency distribution.

Classes ($ intervals)	Frequencies, f (Number of Gas Stations)
$3.00—$3.49	15
$3.50—$3.99	8
$4.00—$4.49	5
$4.50—$5.00	2

Guideline for Collecting Quantitative Data:

—All your measurements must be numbers measured to the same precision.

Examples: In Example 1, the student ages data was a set of whole number ages. In Example 4, the earthquake data set was a list of magnitudes that were decimal numbers in the tenths, such as 7.1, 3.4, 5.6, and so on. In Example 5, the gas prices data set was a list of dollar amounts, rounded to the nearest hundredth, such as $3.09, $3.29, $3.15, etc.

VOCABULARY EXERCISES: **Use the Age Frequency Distribution from Example 1 to answer the following questions.**

Classes (Age Intervals)	frequencies, f (counts)
15—24	68
25—34	72
35—44	34
45—54	16
55—64	8
65—74	2

1) How many classes does the table have?
2) What numbers represent ***lower class limits***?
3) What numbers represent ***upper class limits***?
4) What is the ***class width***?
5) What are the ***class midpoints***?
6) What is the ***sample size***?
7) What are the ***class boundaries***?

---

1) How many classes does the table have?

2) What numbers represent ***lower class limits***?

3) What numbers represent ***upper class limits***?

Guideline for Class Limit Numbers:

—Class limit numbers must be numbers measured to the same precision as the numbers in the data set.

Examples: In Example 1, the class limits were whole numbers, so the student ages data was a list of whole number ages. In Example 2, the class limits are whole numbers, so the data (number of absences) was a list of whole numbers. In Example 4, the class limits were decimal numbers measured to the tenths decimal place value column, so the earthquake data set was a list of magnitudes that were decimal numbers measured in the tenths (such as 7.1, 3.4, 5.6, and so on). In Example 5, the class limits were dollars rounded to the nearest cent, so the gas prices data set was a list of dollar amounts, rounded to the nearest hundredth (such as \$3.09, \$3.29, \$3.15, etc). (In Example 3 there were no class limit numbers, since the data was qualitative data, not quantitative data.)

VOCABULARY EXERCISES (Continued): **Use the Age Frequency Distribution from Example 1 to answer the following questions.**

4) What is the ***class width***?

5) What are the ***class midpoints***?

6) What is the ***sample size***?

7) What are the ***class boundaries***?

Guideline for Class Boundary Numbers:

1) Class Boundary numbers are the midpoints of the gaps that separate classes
2) Class Boundary numbers are decimal numbers that are measured out to one decimal place value column more than the numbers in the data set, and they end in a 5.

An Important Property of Class Width:

—Consecutive class midpoints are separated by a distance equal to the class width.
—Consecutive lower class limits are separated by a distance equal to the class width.
—Consecutive upper class limits are separated by a distance equal to the class width.
—Consecutive lower boundaries are separated by a distance equal to the class width.
—Consecutive upper boundaries are separated by a distance equal to the class width.

How to Construct a Frequency Distribution (using interval classes)

1. Enter the data into the calculator and sort it in ascending fashion.
2. Decide on the number of classes (usually between 5 and 20).
3. Find the class width with this formula

$$\text{class width} = \frac{\text{max} - \text{min}}{\text{number of classes}}$$

   Always round the class width number up. For example, if the formula give class width as 24.1 round that up to 25.

4. Write down column 1 of your table, the class limit numbers. Write down lower class limit numbers first. Then write down upper class limit numbers.

   A First write down (use) the minimum value in the data list as the lower class limit of the 1st interval class. Use your sorted data to find the minimum of the data list.

   B Add class width to the minimum to find (and write down)the lower class limit number for the 2nd interval class. Add class width to the lower class limit of the 2nd class to find the lower class limit of the 3rd interval class. Keep on adding class width to find the next lower class limit, until you find and write down all lower class limit numbers.

   C Write down upper class limit numbers.

5. Use your sorted data to write down the numbers in the frequency column (the 2nd column) of your frequency distribution table.
6. Double check your work.

Example

The following sample data set lists the prices (in dollars) of 30 portable global positioning system (GPS) navigators. Construct a ***frequency distribution*** that has seven classes.

90	130	400	200	350	70	325	250	150	250
275	270	150	130	59	200	160	450	300	130
220	100	200	400	200	250	95	180	170	150

The sorted data

59	70	90	95	100	130	130	130	150	150
150	160	170	180	200	200	200	200	220	250
250	250	270	275	300	325	350	400	400	450

Example

Now make an Expanded Frequency Distribution (table) that includes midpoints, relative frequencies, cumulative frequencies and percentage cumulative frequencies.

---

classes	freq., f	midpoint, X_m	relative frequency	cumulative frequency	% cum. freq.
59 — 114	5				
115 — 170	8				
171 — 226	6				
227 — 282	5				
283 — 338	2				
339 — 394	1				
395 — 450	3				

The ***Midpoint of a class*** is found using the formula

$$\text{class midpoint} = \frac{\text{upper class limit} + \text{lower class limit}}{2}$$

The ***Relative Frequency of a class*** is the portion or percentage of the data that falls in a particular class.

$$\text{relative frequency} = \frac{\text{class frequency}}{\text{sample size}} = \frac{f}{n}$$

The ***Cumulative frequency of a class*** The sum of the frequency for that class and all previous classes.

The Different Types of Data Tables Commonly Used in Statistics

1. **Frequency Distribution** — graph is a histogram or frequency polygon
2. **Relative Frequency Distribution** — graph is a histogram
3. **Cumulative Frequency Distribution** — graph is an ogive
4. **The Expanded Frequency Table**
5. **Categorical Frequency Distribution** — graph is a pie chart or pareto chart

---

1. Frequency Distribution for the GPS Device Price Data

Classes	Frequencies, f
59 — 114	5
115 — 170	8
171 — 226	6
227 — 282	5
283 — 338	2
339 — 394	1
395 — 450	3

2. Relative Frequency Distribution for the GPS Device Price Data

Classes	Relative Frequencies
59 — 114	0.17
115 — 170	0.27
171 — 226	0.20
227 — 282	0.17
283 — 338	0.07
339 — 394	0.03
395 — 450	0.10

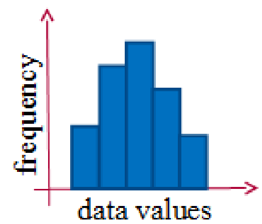
3. Cumulative Frequency Distribution for the GPS Device Price Data

Upper Class Boundary	Cumulative Frequencies
114.5	5
170.5	13
226.5	19
282.5	24
338.5	26
394.5	27
450.5	30

Graphs of Frequency Distributions

Frequency Histogram

- A bar graph that represents the frequency distribution.
- The horizontal scale is quantitative and measures the data values.
- The vertical scale measures the frequencies of the classes.
- Consecutive bars must touch.
- Bars are centered at class midpoints
- Bars are graphed between ***class boundaries***.

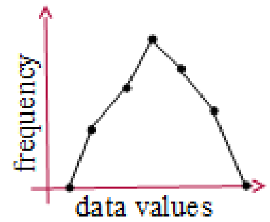


Construct a frequency histogram for the GPS navigators frequency distribution.

How to Sketch a Frequency Histogram

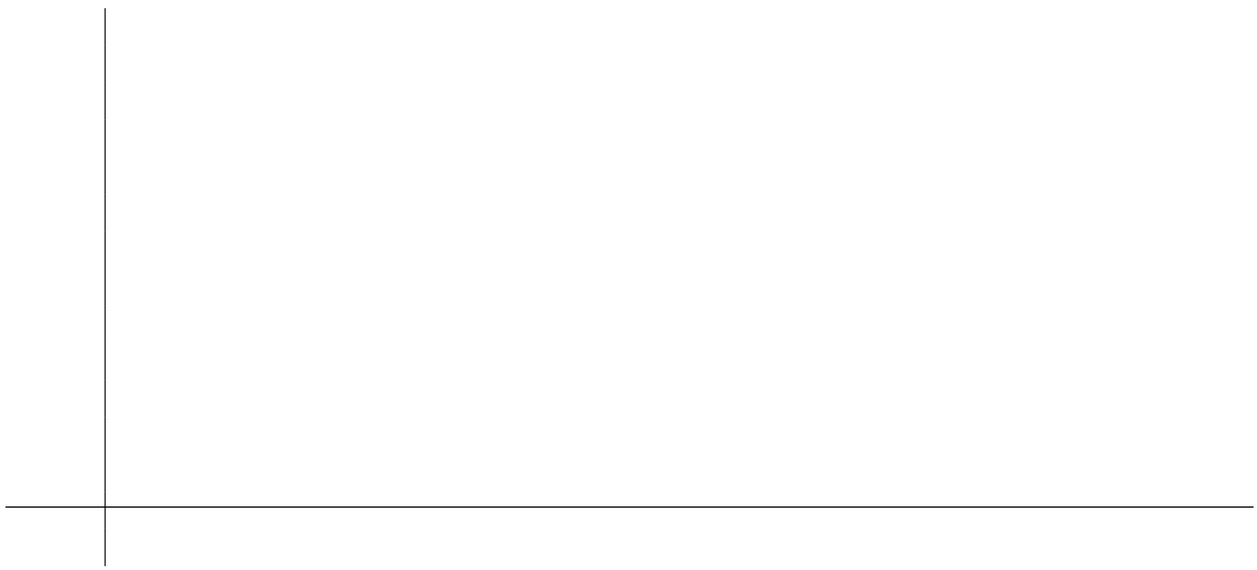
1. Locate and label class midpoints (or boundaries) along the x axis.
2. Make tick marks at the midpoint between each pair of consecutive midpoints. These are the location of your class boundaries where you histogram bars will connect.
3. Sketch a bar for each class. The height of each bar should be equal to the frequency of the class. Histogram bars should be centered around class midpoints.
4. Label your x and y axis. The y axis should be labeled 'frequency' and the x axis should be labeled as the variable you measured in the data.

Frequency Polygon



- A line graph that emphasizes the continuous change in frequencies.

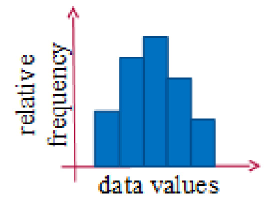
Construct a frequency polygon for the GPS navigators frequency distribution.



How to Sketch a Frequency Polygon

1. Locate and label class midpoints along the x axis.
2. Above each midpoint draw a point whose height above the x-axis is equal to the frequency of that class.
3. Locate the left-hand x-intercept. Subtract the class width from the lowest class midpoint, and locate the resulting number on the x-axis
4. Locate the right-hand x-intercept. Add the class width to the largest class midpoint, and locate the resulting number on the x-axis
5. Connect your points with line segments. The final graph should make a closed figure with the x-axis.
6. Label your x and y axis. The y axis should be labeled 'frequency' and the x-axis should be labeled as the variable you measured in the data.

Relative Frequency Histogram



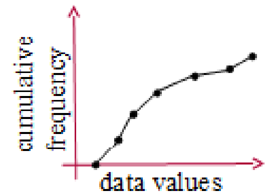
- Has the same shape and the same horizontal scale as the corresponding frequency histogram.
- The vertical scale measures the relative frequencies (percentages), not frequencies.

Construct a relative frequency histogram for the GPS navigators frequency distribution.

<u>How to Sketch a Relative Frequency Histogram</u>

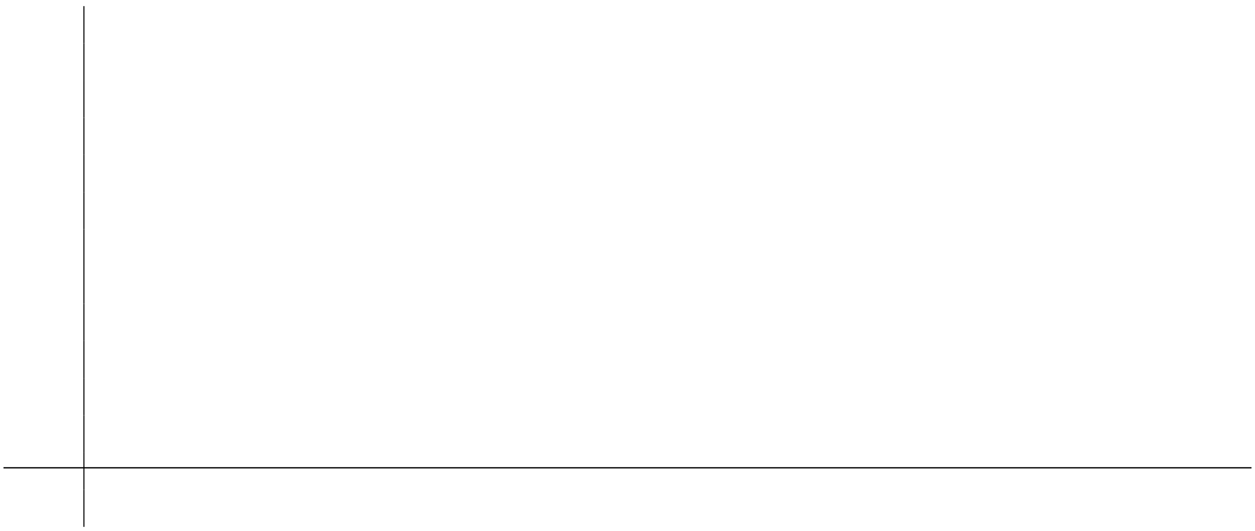
1. Locate and label class midpoints (or boundaries) along the x axis.
2. Make tick marks at the midpoint between each pair of consecutive midpoints. These are the location of your class boundaries where you histogram bars will connect.
3. Sketch a bar for each class. The height of each bar should be equal to the relative frequency (either decimals or percents) of the class. Histogram bars should be centered around class midpoints.
4. Label your x and y axis. The y axis should be labeled 'relative frequency' and the x axis should be labeled as the variable you measured in the data.

Cumulative Frequency Graph or Ogive



- A line graph that displays the cumulative frequency of each class at its upper class boundary.
- The upper boundaries are marked on the horizontal axis.
- The cumulative frequencies are marked on the vertical axis.

Construct an ogive for the GPS navigators frequency distribution.

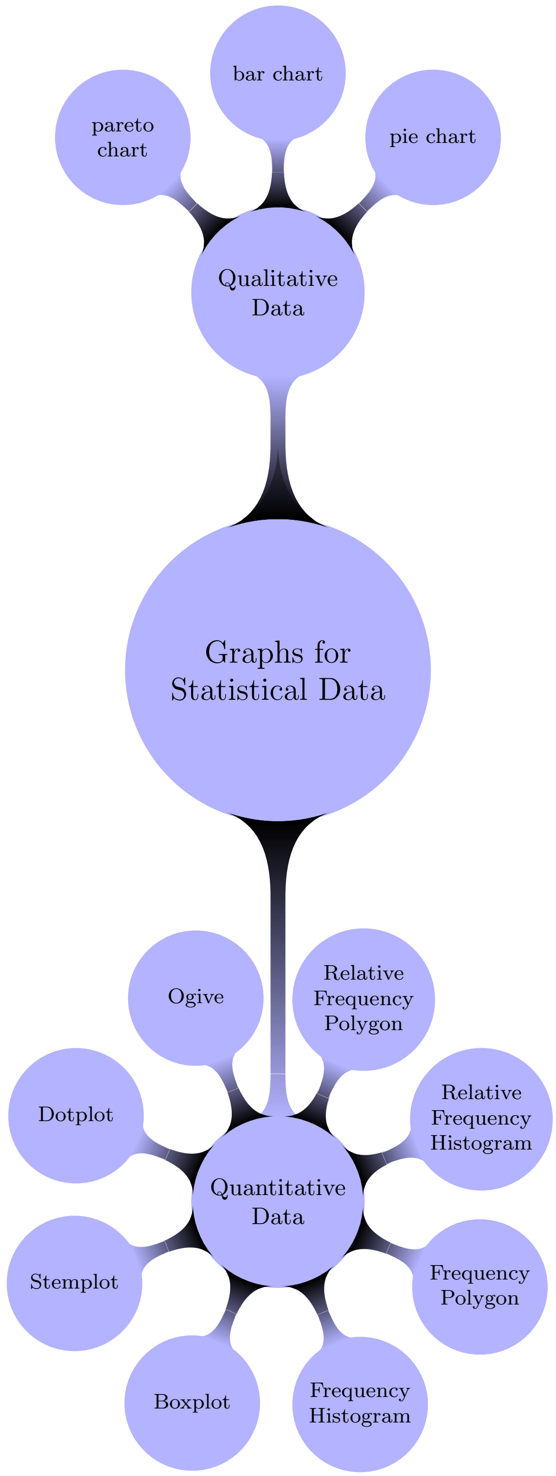


How to Construct an Ogive

1. Construct a frequency distribution table that includes cumulative frequencies as one of the columns. Find and write down values for the boundaries of your classes.
2. Locate upper class boundary numbers along the x-axis. Locate and draw an ordered pair for each class in your frequency distribution table. Each ordered pair will have the coordinates

$$(x, y) = (\text{upper class boundary}, \text{cumulative frequency})$$

   - The horizontal scale consists of the upper class boundaries.
   - The vertical scale measures cumulative frequencies.
3. Plot points that represent the upper class boundaries and their corresponding cumulative frequencies.
4. Connect the points in order from left to right.
5. The graph should start at the lower boundary of the first class (cumulative frequency is zero) and should end at the upper boundary of the last class (cumulative frequency is equal to the sample size).


bar chart
pareto chart
pie chart
Qualitative Data
Graphs for Statistical Data
Ogive
Relative Frequency Polygon
Dotplot
Relative Frequency Histogram
Quantitative Data
Stemplot
Frequency Polygon
Boxplot
Frequency Histogram

2.2 More Graphs and Displays

Learning objectives:

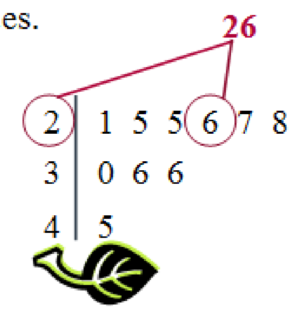
1. How to graph and interpret quantitative data using stem-and-leaf plots and dot plots
2. How to graph and interpret qualitative data using pie charts and Pareto charts
3. How to graph and interpret paired data sets using scatter plots and time series charts

---

Stem-and-leaf plot

- Each number is separated into a **stem** and a **leaf**.
- Similar to a histogram.
- Still contains original data values.

Data: 21, 25, 25, **26**, 27, 28, 30, 36, 36, 45



How to Graph a Stem-and-Leaf plot

1. List the stems to the left of a vertical line.
2. For each data entry, list a leaf to the right of its stem.

---

1. ***Example***: Use the stem-and-leaf plot to list the actual data entries.

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

Key : 3|1 means 31

2. ***Example***: Use the stem-and-leaf plot to list the actual data entries.

Key: 12|9 = 12.9

12	
12	9
13	3
13	6 7 7
14	1 1 1 1 3 4 4
14	6 9 9

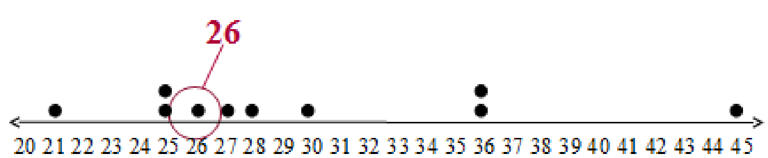
3. ***Example***: Use a stem-and-leaf plot to display the data. The data represents the number of hours 24 nurses work per week.

40	40	35	48	38	40	36	50	32	36	40	35
30	24	40	36	40	36	40	39	33	40	32	38

Dot plot

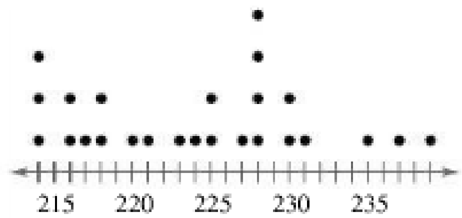
- Each data entry is plotted, using a point, above a horizontal axis

Data: 21, 25, 25, **26**, 27, 28, 30, 36, 36, 45



If an entry is repeated, plot another point above the previous point.

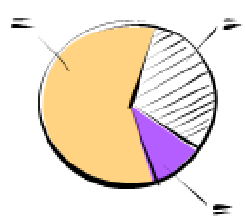
4. Construct a stem-and-leaf plot for the data represented in the dotplot below.



5. ***Construct the dotplot for the given data.*** Attendance records for a preschool class with 15 students show the number of days each student was absent during the year. The days absent for each student were as follows.

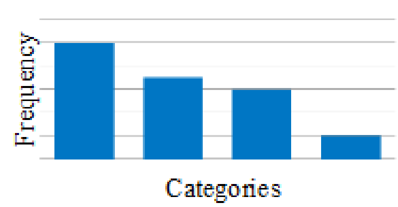
3 5 9 5 5 6 2 1 8 4 3 5 6 7 8

Pie Chart



- A circle is divided into sectors that represent categories.
- The area of each sector is proportional to the frequency of each category.

Pareto Chart



- A vertical bar graph in which the height of each bar represents frequency or relative frequency.
- The bars are positioned in order of decreasing height, with the tallest bar positioned at the left.

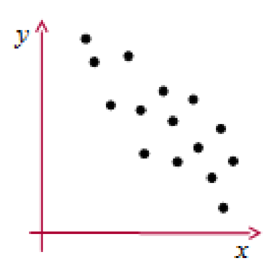
6. ***Construct a pie chart and a pareto chart for the data summarized in the following categorical frequency distribution.***

Response	At home	At friend's home	At restaurant or bar	Somewhere else	Not sure
Number	620	110	50	100	130

Graphing Paired Data Sets

Paired Data Sets

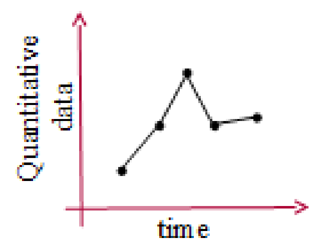
- Each entry in one data set corresponds to one entry in a second data set.
- Graph using a **scatter plot.**
  - The ordered pairs are graphed as points in a coordinate plane.
  - Used to show the relationship between two quantitative variables.



7. Use a scatter plot to display the data below. All measurements are in milligrams per cigarette.

Brand	Tar	Nicotine
Benson & Hedges	16	1.2
Lucky Strike	13	1.1
Marlboro	16	1.2
Viceroy	18	1.4
True	6	0.6

Time Series

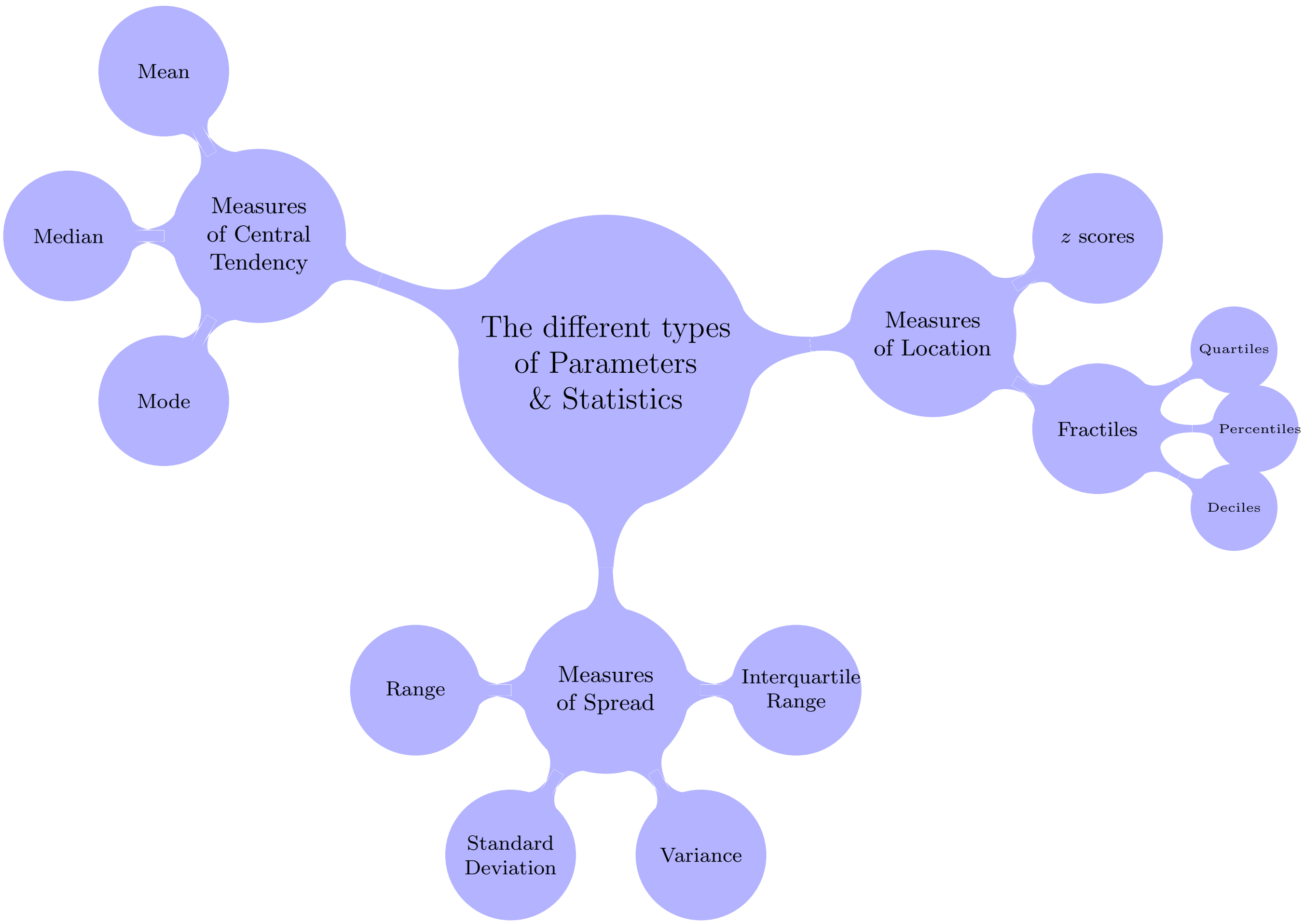


- Used when a data set is composed of quantitative entries taken at regular intervals over a period of time
- for example, The amount of precipitation measured each day for one month

8. The data below represent the smoking prevalence among U.S. adults over a 35-year period. Use a time series chart to display the data. Describe any trends shown.

Year	1965	1985	1990	1995	2000
Percent of Smokers	42	30	25	25	23

The different types of parameters and statistics for quantitative data can be grouped into three different categories: Measures of Central Tendency, Measures of Spread and Measures of Location. The next three sections in the workbook and textbook are devoted to those three topics.



Definition **Measures of Central Tendency** are numbers that represent the typical value in a data set. Measures of Central Tendency measure the location of the center of the data.

Definition **Measures of Spread** (also called Measures of Variations) are numbers that measure how spread out a data set is along the x axis. The four common measures of spread we can find for a quantitative data set are: the range, the variance, the standard deviation and the interquartile range.

Definition **Measures of Location** (also called Measures of Relative Standing) are numbers that measure the location of data value relative to the center of the data (z-scores); and relative to the other data values (fractiles).

2.3 Measures of Central Tendency

Learning objectives:

1. How to find the mean, median, and mode of a population and of a sample
2. How to find the weighted mean of a data set and the mean of a frequency distribution
3. How to describe the shape of a distribution as symmetric, uniform, or skewed and how to compare the mean and median for each

---

Measure of central tendency

1. A value that represents a typical, or central, entry of a data set.
2. Most common measures of central tendency: ***mean, median and mode***.

Mean (average)

1. The sum of all the data entries divided by the number of entries.
2. Sigma notation: $\sum x$ means "add all of the data entries (x) in the data set."
3. Population Mean: $\mu = \dfrac{\sum x}{N}$
4. Sample Mean: $\bar{x} = \dfrac{\sum x}{n}$

---

4. ***Example***: Find the mean for the census data below.

$$3 \quad 6 \quad 5 \quad 7 \quad 9$$

Median

- The value that lies in the middle of the data when the data set is **ordered**.
- Measures the center of an ordered data set by dividing it into two equal parts.
- If the data set has an
  - **odd number of entries**: median is the middle data entry.
  - **even number of entries**: median is the mean of the two middle data entries.

5. ***Example***: Find the median. The prices (in dollars) for a sample of roundtrip flights from Chicago, Illinois to Cancun, Mexico are listed below.

   872 432 397 427 388 782 397

6. ***Example***: The flight priced at $432 is no longer available. What is the median price of the remaining flights?

   872 397 427 388 782 397

Mode

- The data entry that occurs with the greatest frequency.
- If no entry is repeated the data set has no mode.
- If two entries occur with the same greatest frequency, each entry is a mode (**bimodal**).

7. At a political debate a sample of audience members was asked to name the political party to which they belong. Their responses are shown in the table. What is the mode of the responses?

Political Party	Frequency, f
Democrat	34
Republican	56
Other	21
Did not respond	9

	Advantages	Disadvantages
Mean	Is relatively reliable, means of samples drawn from the same population dont vary as much as other measures of center. Takes every data value into account	Is sensitive to every data value, one extreme value can affect it dramatically; is not a resistant measure of center.
Median	is not affected by an extreme value - is a resistant measure of the center	Doesn't always reflect the true center
Mode	is fairly easy to find	Doesn't always reflect the true center often a data set has no mode

Weighted Mean

1. The mean of a data set whose entries have varying weights.
2. $\bar{x} = \dfrac{\sum(x \cdot w)}{\sum w}$, where w is the weight of each entry x.

---

8. You are taking a class in which your grade is determined from five sources: 50% from your test mean, 15% from your midterm, 20% from your final exam, 10% from your computer lab work, and 5% from your homework. Your scores are 86 (test mean), 96 (midterm), 82 (final exam), 98 (computer lab), and 100 (homework). What is the weighted mean of your scores? If the minimum average for an A is 90, did you get an A?

9. For the month of May, a checking account has a balance of \$759 for 15 days, \$1985 for 5 days, \$1410 for 5 days and \$348 for 6 days. What is the account's mean daily balance for May?

Mean of a Frequency Distribution

1. $\bar{x} = \dfrac{\sum(x_m \cdot f)}{n}$, with $n = \sum f$, where x_m and f are the midpoints and frequencies of a class, respectively
2. Can be used to find the mean of a data set that is summarized in a table.

How to find the Mean of a Frequency Distribution

1. Find the midpoint of each class with $x_m = \dfrac{lowerclasslimit + upperclasslimit)}{2}$.
2. Next find $\sum(x_m \cdot f)$, the sum of the products of the midpoints and the frequencies.
3. Find $n = \sum f$, the sum of the frequencies (the sample size).
4. Divide using $\dfrac{\sum(x_m \cdot f)}{n}$

---

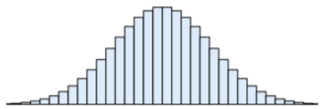
10. Use the frequency distribution to approximate the mean number of minutes that a sample of Internet subscribers spent online during their most recent session.

Class	Midpoint	Frequency, f
7 – 18	12.5	6
19 – 30	24.5	10
31 – 42	36.5	13
43 – 54	48.5	8
55 – 66	60.5	5
67 – 78	72.5	6
79 – 90	84.5	2

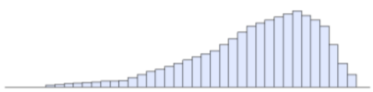
Skewness

Definition

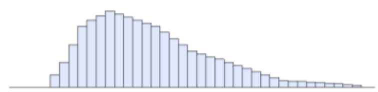
A distribution of data is **skewed** if it is not symmetric and extends more to one side than to the other. (A distribution of data is symmetric if the left half of its histogram is roughly a mirror image of its right half.) [**?**]



(b) Symmetric Distribution



(a) Skewed to the Left

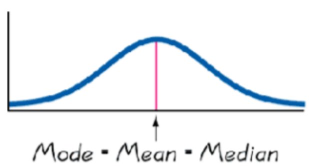


(c) Skewed to the Right

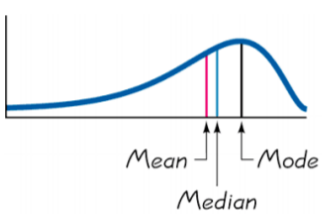
The distribution in (a) is called "skewed left" because most of the data falls to the left of the mode (the value along the *x*-axis associated with the largest bar in the histogram). The distribution in (c) is called "skewed right" because most of the data falls to the right of the mode.

Definition

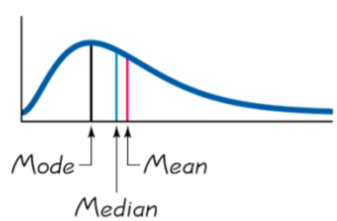
A distribution of data is **skewed** if it is not symmetric and extends more to one side than to the other. (A distribution of data is symmetric if the left half of its histogram is roughly a mirror image of its right half.) [**?**]



(b) Symmetric



(a) Skewed to the Left (Negatively)



(c) Skewed to the Right (Positively)

2.4 Measures of Variation

Learning objectives:

1. How to find the range of a data set
2. How to find the variance and standard deviation of a population and of a sample
3. How to use the Empirical Rule and Chebychev's Theorem to interpret standard deviation
4. How to approximate the sample standard deviation for grouped data
5. How to use the coefficient of variation to compare variation in different data sets

---

Measures of Variations (also called Measures of Spread) are numbers that measure how spread out a data set is along the x-axis. The four common measures of spread we can find for a quantitative data set are the range, the variance, the standard deviation and the interquartile range. (The interquartile range is given in Section 2.5, the next section.)

The ***Range*** of a quantitative data set is the difference between the maximum and minimum data entries in the set.

$$\text{Range } = (\text{Max. data entry}) - (\text{Min. data entry})$$

Example 1 A corporation hired 5 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries.

Starting salaries (in 1000s of dollars)	41	38	39	45	47

The Deviation Value for a particular number in a data set is the distance (difference) between the data entry, x, and the mean of the data set.

Formula used when you have a Population: Deviation of $x = x - \mu$

Formula used when you have a Sample: Deviation of $x = x - \bar{x}$

Example 2

Find the deviation of the starting salaries from Ex. 1.

To obtain another measure of spread it was proposed that we use the average of the deviations for a data set, as this would give us a measure of how much the typical data entry deviated from the mean (center) of the data. The problem with that is that for any data set, the mean of the deviations will equal 0! It was then proposed that we take the absolute value of each deviation before we find the average deviation. The problem with getting an average deviation this way was that it almost always overestimated the value of the average deviation. The decision was then made to approximate average deviation by first squaring the deviation lengths before finding the average (called "finding the **variance**"), and then finding the square root of the average (of the squared deviations). This became the standard way of finding average deviation, hence the name ***Standard Deviation***.

***Mathematical Notation Note:* Variables used for Variance and Standard Deviation**

- We use σ (the Greek lowercase s) as the symbol representing the *population* standard deviation.
- We use σ^2 as the symbol representing the *population* variance.
- We use s (the Roman lowercase s) as the symbol representing the *sample* standard deviation.
- We use s^2 as the symbol representing the *sample* variance.
- σ and σ^2 are *parameters* while s and s^2 are *statistics.*

Formulas for Population Variance and Standard Deviation

Population Variance $$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

This gives us a formula for the "average of the squared deviations" for population data (census data). However, the units of variance is equal to the units of the data raised to the second power. For example, if x is in dollars, then the units on variance will be dollars squared.

Population Standard Deviation $$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

If we take the square root of the variance however, we get σ, the population standard deviation — a parameter that has the same units as numbers in the data set.

How to find the Population Variance & Standard Deviation

1. Find the mean of the population data set, $\mu = \dfrac{\sum x}{N}$
2. Find deviation of each entry, $x - \mu$
3. Square each deviation, $(x - \mu)^2$.
4. Add to get the sum of squares, $\sum (x - \mu)^2$
5. Divide by N to get the population variance, $\dfrac{\sum (x - \mu)^2}{N}$
6. Find the square root to get the population standard deviation, $\sqrt{\dfrac{\sum (x - \mu)^2}{N}}$

Example 3 A corporation hired 5 graduates. The starting salaries for each graduate are shown. Find the population variance and standard deviation of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47

Properties of Standard Deviation

- The standard deviation is the typical amount that a data entry deviates from the mean (center of the data).
- The standard deviation measures the variation of the data set about (around) the mean (middle of the data), and has the same units as the data.
- The standard deviation is always greater than or equal to 0. When $\sigma = 0$, the data has no variation and all the entries have the same value.
- As the data entries get farther from the mean (that is, more spread out), the value of σ increases.

Formulas for Sample Variance and Standard Deviation

Sample Variance $$s^2 = \frac{\sum(x - \bar{x})^2}{n-1}$$

Sample Standard Deviation $$s = \sqrt{s^2} = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}}$$

How to find the Sample Variance & Standard Deviation

1. Find the mean of the sample data set, $\bar{x} = \dfrac{\sum x}{n}$
2. Find deviation of each entry, $x - \bar{x}$
3. Square each deviation, $(x - \bar{x})^2$.
4. Add to get the sum of squares, $\sum(x - \bar{x})^2$
5. Divide by $(n-1)$ to get the sample variance, $\dfrac{\sum(x - \bar{x})^2}{n-1}$
6. Find the square root to get the sample standard deviation, $\sqrt{\dfrac{\sum(x - \bar{x})^2}{n-1}}$

When calculating standard deviation (and variance) for a sample, we adopt the convention (the same behavior) of dividing the sum of the squared deviations by $(n-1)$ ***because statisticians noticed this would, on average, give a better estimate for the value of population variance.***

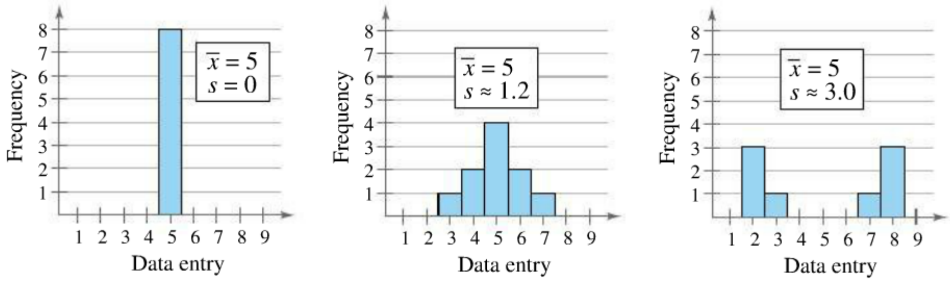
Example 4 A corporation hired 5 graduates. The starting salaries for each graduate are shown. Find the ***sample*** variance and standard deviation of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47

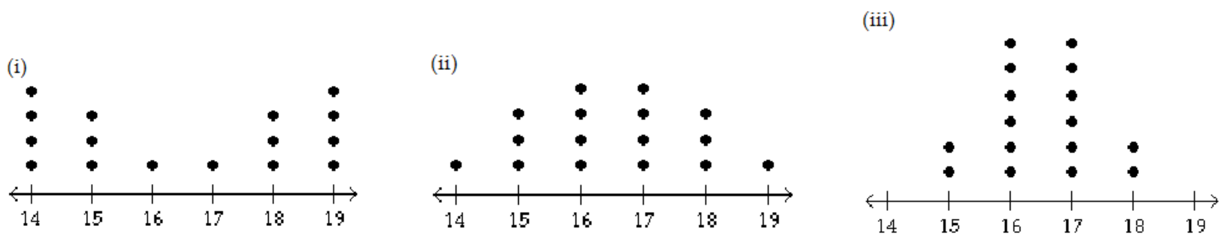
INTERPRETING STANDARD DEVIATION

When interpreting the standard deviation, remember that it is a measure of the typical amount an entry deviates from the mean. The more the entries are spread out, the greater the standard deviation.



Try This! For each graph above, list the sample data being graphed.

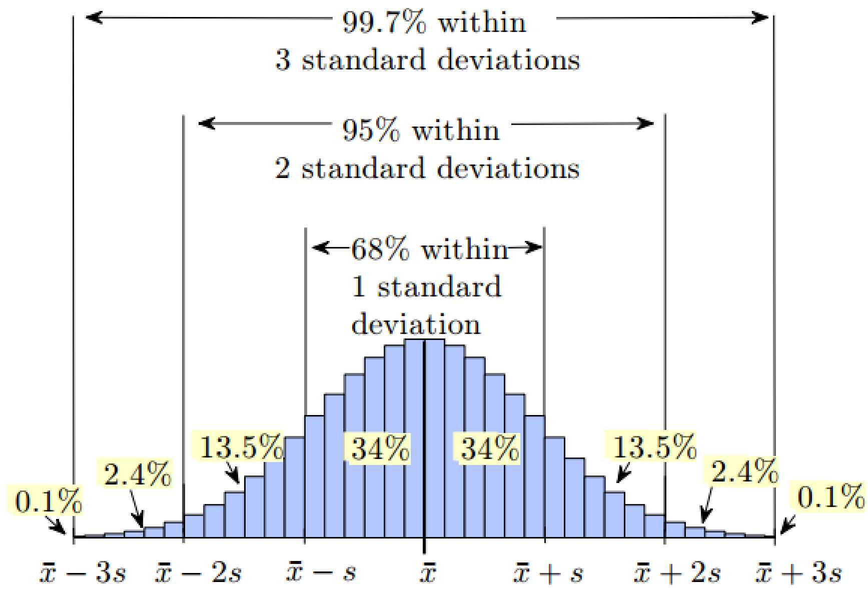
Try This! You are asked to compare three data sets (dotplots below). Without calculating, determine which data set has the greatest sample standard deviation and which has the least sample standard deviation.



GUIDELINE: HOW TO IDENTIFY USUAL AND UNUSUAL VALUES

Data entries that lie more than two standard deviations from the mean are considered unusual, while those that lie more than three standard deviations from the mean are very unusual. Unusual and very unusual entries have a greater influence on the standard deviation than entries closer to the mean. This happens because their deviation lengths are squared.

Try This! A sample of 500 monthly utility bills for households in a city was collected. The mean of the sample was \$70 and the sample standard deviation was \$8. Here is a short list of a few of the 500 measurements from the sample: \$74, \$52, \$62, \$98 Are any of the data entries unusual or very unusual? Explain your reasoning.



The Empirical Rule

For data with a (symmetric) bell-shaped distribution, the standard deviation has the following characteristics:

- About 68% of the data lie within one standard deviation of the mean.
- About 95% of the data lie within two standard deviations of the mean.
- About 99.7% of the data lie within three standard deviations of the mean.

Example 5: The mean IQ score of students in a particular calculus class is 110, with a standard deviation of 5. (Assume the data set has a bell-shaped distribution.) Hint: use the Empirical Rule

a) Use the Empirical Rule to find the percentage of students with an IQ below 110.

b) Use the Empirical Rule to find the percentage of students with an IQ above 110.

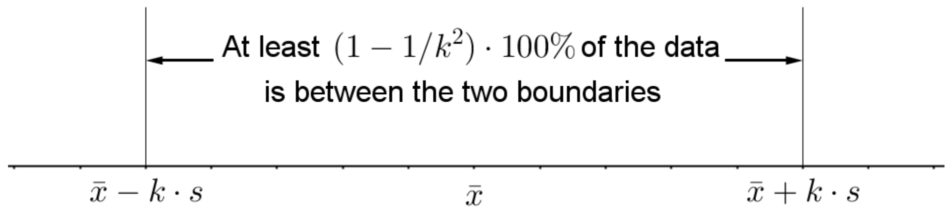
c) Use the Empirical Rule to find the percentage of students with an IQ above 120.

d) Use the Empirical Rule to find the percentage of students with an IQ between 100 and 110

e) Use the Empirical Rule to find the percentage of students with an IQ between 105 and 120

1. ***Try This!***: In a survey conducted by the National Center for Health Statistics, the sample mean height of women in the United States (ages 20-29) was 64.3 inches, with a sample standard deviation of 2.62 inches. Use the Empirical Rule to estimate the percent of the women whose heights are between 59.06 inches and 64.3 inches.

Chebychev's Theorem



1. The percentage of any data set lying within k standard deviations (where k is any number greater than one) of the mean is at least: $\left(1 - \dfrac{1}{k^2}\right) \cdot 100\%$
2. For example, when $k = 2$: In any data set, at least $1 - \dfrac{1}{2^2} = \dfrac{3}{4}$ or 75% of the data lie within 2 standard deviations of the mean.
3. When $k = 3$: In any data set, at least $1 - \dfrac{1}{3^2} = \dfrac{8}{9}$ or 88.9% of the data lie within 3 standard deviations of the mean.
4. k can be any number greater than one

Try This! The mean time in the finals for the women's 800-meter freestyle at the 2012 Summer Olympics was 502.84 seconds, with a standard deviation of 4.68 seconds. Apply Chebychev's Theorem to the data using $k = 1.5$. Interpret the results.

Try This! Heights of adult women have a mean of 63.6 in. and a standard deviation of 2.5 in. Does Chebyshev's Theorem say about the percentage of women with heights between 58.6 in. and 68.6 in.? At least how many women in a sample of 50 would have heights between 58.6 in. and 68.6 in.?

2.5 Measures of Position

Learning objectives:

1. How to find the first, second, and third quartiles of a data set, how to find the interquartile range of a data set, and how to represent a data set graphically using a box-and whisker plot
2. How to interpret other fractiles such as percentiles and how to find percentiles for a specific data entry
3. Determine and interpret the standard score (z-score)

Fractiles

- In this section, you will learn how to use fractiles to specify the location of a data entry within a data set.
- Fractiles are numbers that partition (divide) an ordered data set into equal parts.

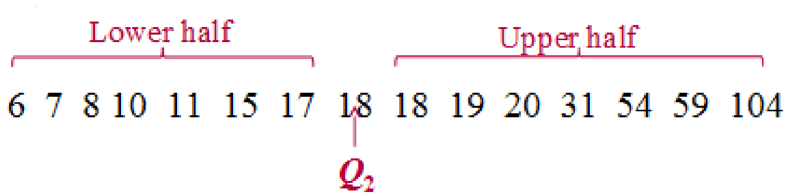
Quartiles

- First quartile, Q_1: About one quarter of the data fall on or below Q_1.
- Second quartile, Q_2: About one half of the data fall on or below Q_2 (median).
- Third quartile, Q_3: About three quarters of the data fall on or below Q_3.

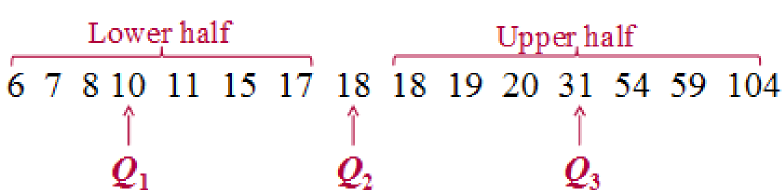
Example: The number of nuclear power plants in the top 15 nuclear power-producing countries in the world are listed. Find the first, second, and third quartiles of the data set.

7 18 11 6 59 17 18 54 104 20 31 8 10 15 19

Solution: Q2 divides the data set into two halves.



The first and third quartiles are the medians of the lower and upper halves of the data set.



Interquartile Range (IQR)

- The difference between the third and first quartiles.
- IQR $= Q_3 - Q_1$.

Recall that for the nuclear power plant data, $Q_1 = 10$, $Q_2 = 18$ and $Q_3 = 31$. The interquartile range is

$$\text{IQR} = Q_3 - Q_1 = 31 - 10 = 21$$

The number of power plants in the middle portion of the data set vary by at most 21.

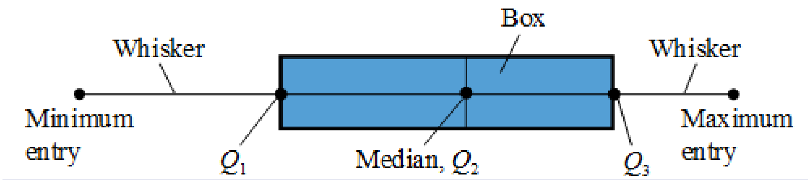
Box-and-whisker plot

- Exploratory data analysis tool.
- Highlights important features of a data set.
- Requires (five-number summary): Minimum entry, First quartile Q_1, Median Q_2, Third quartile Q_3, Maximum entry

---

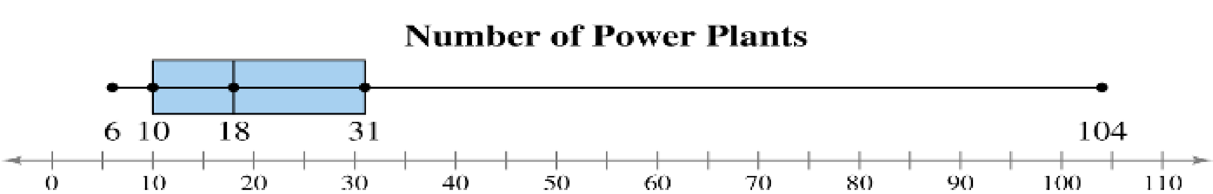
How to draw a Box-and-Whisker Plot

1. Find the five-number summary of the data set.
2. Construct a horizontal scale that spans the range of the data.
3. Plot the five numbers above the horizontal scale.
4. Draw a box above the horizontal scale from Q_1 to Q_3 and draw a vertical line in the box at Q_2.
5. Draw whiskers from the box to the minimum and maximum entries.



Example The five number summary for the power plant data is:

$$\text{Min} = 6, \quad Q_1 = 10, \quad Q_2 = 18, \quad Q_3 = 31, \quad \text{Max} = 104$$



About half the scores are between 10 and 31. By looking at the length of the right whisker, you can conclude 104 is a possible outlier. Another way to identify outliers is to use the interquartile range.

Using the Interquartile Range to Identify Outliers

1. Find Q_1 and Q_3
2. Find the interquartile range, IQR $= Q_3 - Q_1$
3. Multiply the IQR by 1.5
4. Subtract (1.5)×(IQR) from Q_1 to find the lower fence.

$$lower fence = Q_1 - 1.5 \cdot (IQR)$$

   Any data entry less than the lower fence is an outlier.

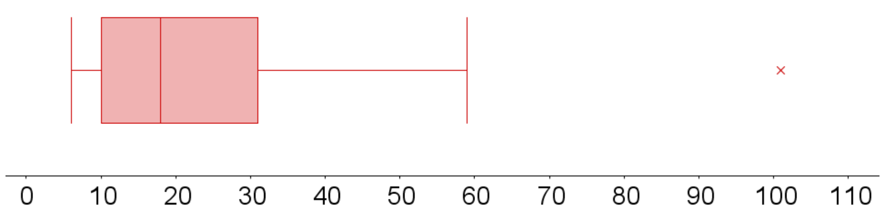
5. Add 1.5×IQR to Q_3 to find the upper fence.

$$upper fence = Q_3 + 1.5 \cdot (IQR)$$

   Any data entry greater than the upper fence is an outlier.

Example When carrying out this list of steps for the power plant data we find:

1. $Q_1 = 10$ and $Q_3 = 31$
2. Find the interquartile range, IQR $= Q_3 - Q_1 = 21$
3. Multiply the IQR by 1.5: $1.5 \times 21 = 31.5$
4. lower fence $= 10 - 31.5 = \boxed{-21.5}$. There is no data entry less than -21.5 value.
5. upper fence $= 31 + 31.5 = \boxed{62.5}$. So, 104 is an outlier since it is greater than 62.5.



Modified Boxplot The boxplot is redrawn with the whisker on *the right extended out to the largest value in the data set that is not larger than the upper fence*, namely data value 59. Had there been an outlier on the lower end, then our graph would have *the left whisker extended to the lowest value in the data set that was not an outlier.* Also, notice that outliers are marked on the graph with a cross marker.

Your Turn! Use the sample data below answer the questions on the next page. The sample represents the number of paid vacation days used by 20 employees in a recent year.

16 25 1 33 15 5 18 8 20 14 17 19 16 10 21 28 14 37 18 15

1. List the data in ascending fashion.

2. Find the minimum — 2. ____________

3. Find Q_1 — 3. ____________

4. Find Q_2 — 4. ____________

5. Find Q_3 — 5. ____________

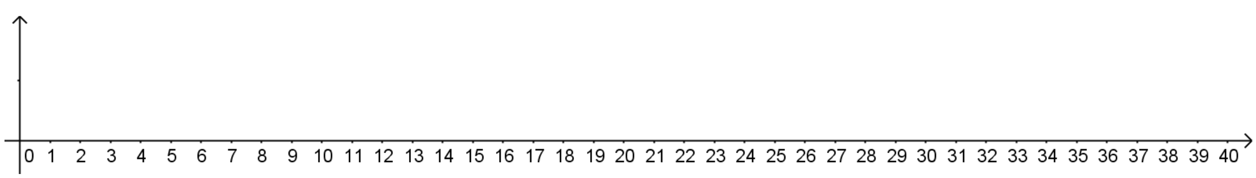
6. Find the maximum — 6. ____________

7. Find the lower fence. — 7. ____________

8. Find the upper fence — 8. ____________

9. What numbers are outliers? — 9. ____________

10. Sketch the graph of a modified boxplot. Label the x axis.



11. About 75% of the employees in the sample took at least how many days off?

    11. ____________

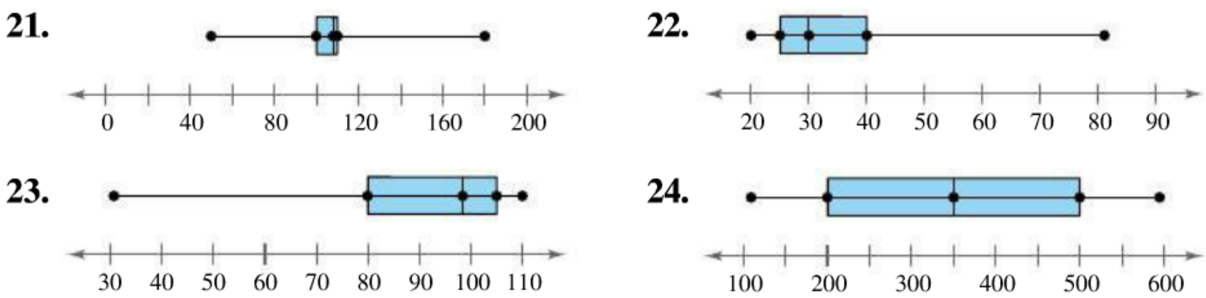
12. What percentage of employees took more than 16 and a half days off?

    12. ____________

13. You randomly select one employee from the sample. What is the likelihood that the person took less than 20.5 days off?

    13. ____________

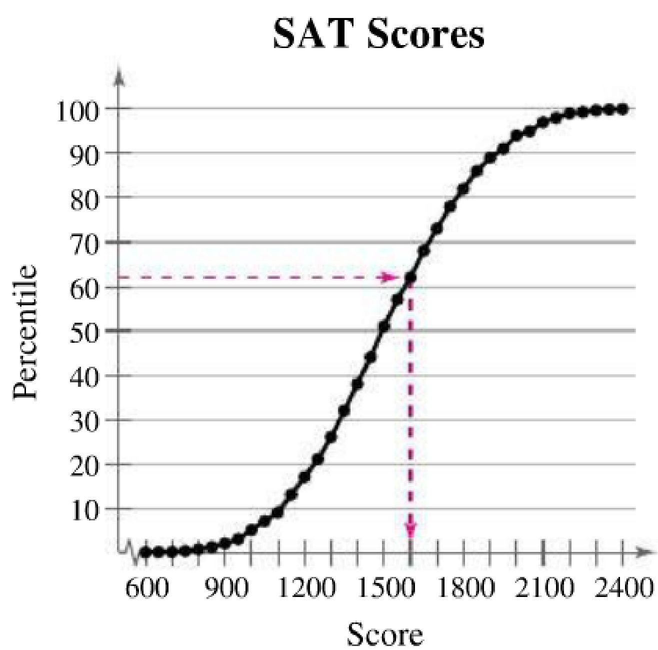
Graphical Analysis *In Exercises 21–24, use the box-and-whisker plot to determine whether the shape of the distribution represented is symmetric, skewed left, skewed right, or none of these. Justify your answer.*



Fractiles	Summary	Symbols
Quartiles	Divides data into 4 equal parts	Q_1, Q_2, Q_3
Deciles	Divides data into 10 equal parts	$D_1, D_2, D_3, \ldots, D_9$
Percentiles	Divides data into 100 equal parts	$P_1, P_2, P_3, \ldots, P_{99}$

The ogive at the right represents the cumulative frequency distribution for SAT scores of college-bound students in a recent year. What SAT score represents the 62nd percentile?

Answer/Interpretation: An SAT score of 1600. This means that approximately 62% of the students had an SAT score of 1600 or less.



14. What percentage of students scored at most 1500?

15. Approximately what SAT score represents the 90th percentile? How should you interpret this?

Definition To find the **percentile that corresponds to a specific data entry** x, use the formula

$$\text{Percentile of } x = \frac{\text{number of data entries less than } x}{\text{total number of data entries}} \cdot 100$$

and then round to the nearest whole number.

Example For the vacation days data, find the percentile that corresponds to 8 vacation days.

Answer There are two data entries less than 8 and the total number of data entries is 20.

$$\text{Percentile of } 8 = \frac{\text{number of data entries less than } x}{\text{total number of data entries}} \cdot 100 = \frac{2}{20} \cdot 100 = 10$$

Interpretation 8 vacation days off corresponds to the 10th percentile. 10% of the data values are less than or equal to 8 vacation days. (8 vacation days is greater than 10% of the entries in the sample).

Try These!: The exam scores of a particular geography class are listed.

26	54	59	64	67	73	73	73	74	76
77	78	80	80	81	83	83	85	86	87
87	88	88	89	90	91	92	93	94	97

16. Find the percentile that corresponds to an exam score of 67. How should you interpret this?

17. Find the percentile that corresponds to an exam score of 89. How should you interpret

this?

Standard Score (z-score)

- Every number x in a data set has a z-score
- A z-score represents the number of standard deviations a given value x is away from the mean μ.
- z-score formula $\qquad z = \dfrac{x - \mu}{\sigma}$
- Always round z-scores to the hundredths.
- A z-score can be negative, positive or zero.
- When z is negative the corresponding x value is less than the mean
- When z is positive the corresponding x value is greater than the mean
- When $z = 0$ the corresponding x value is equal to the mean

Life Spans of Fruit Flies The life spans of a species of fruit fly have a bell-shaped distribution, with a mean of 33 days and a standard deviation of 4 days.

(a) The life spans of three randomly selected fruit flies are 34 days, 30 days, and 42 days. Find the z-score that corresponds to each life span. Determine whether any of these life spans are unusual.

(b) The life spans of three randomly selected fruit flies are 29 days, 41 days, and 25 days. Using the Empirical Rule, find the percentile that corresponds to each life span.

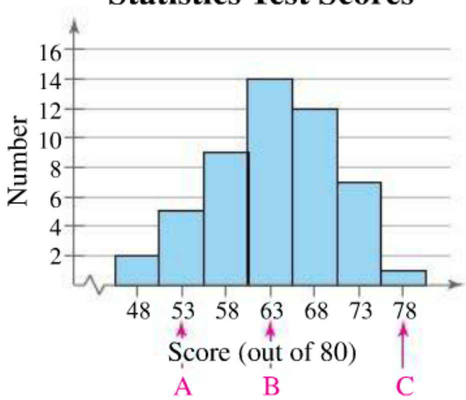
Graphical Analysis *In Exercises 39 and 40, the midpoints A, B, and C are marked on the histogram. Match them with the indicated z-scores. Which z-scores, if any, would be considered unusual?*

39. $z = 0$

$z = 2.14$

$z = -1.43$

Statistics Test Scores

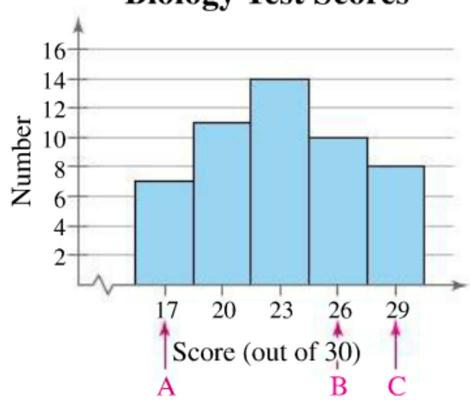


40. $z = 0.77$

$z = 1.54$

$z = -1.54$

Biology Test Scores



VIDEO LINK: **https://www.youtube.com/watch?v=_Np8mJGeQ3g**

How to use the calculator to graph a Boxplot

The TI-83, TI-83 Plus, and TI-84 Plus calculators will take a list of data and automatically draw a box-and-whisker plot for that data. Since there are a number of different types of plots available on the calculator, it is important to make sure that all other plots are turned off before you begin or your graph will be cluttered with several unrelated plots being graphed at the same time. Even worse, it is possible for previous plots to become invalid or the data sets that were used before are changed or deleted, causing an error whenever you try to graph anything new. Before you begin any plotting:

- Press the **Y=** key at the top left of the keyboard, and delete or deselect any equations being plotted there.

- Press STAT PLOT (above the **Y=** key) and choose 4 to turn off any other statistical plots.

There are three stages to creating a box-and-whisker plot.

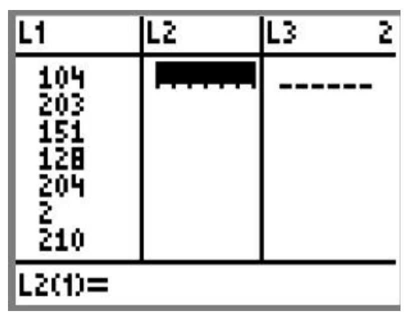
1. Enter the data into a list as before.
2. Tell the calculator what kind of plot you want.
3. Tell the calculator what size to draw the window for the plot.

Example: Doctor Incomes

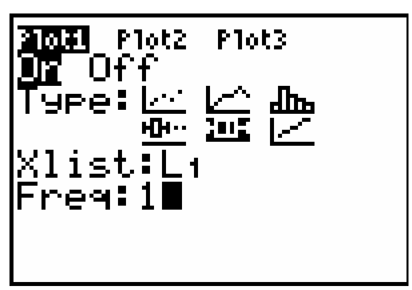
The following numbers represent 2016 incomes in thousands of dollars for seventeen doctors:

104 203 151 128 204 2 210 162 185 169 135 178 350 213 122 135 5

We'll use this data to construct a box-and-whisker plot.
First store the data in the list L1.



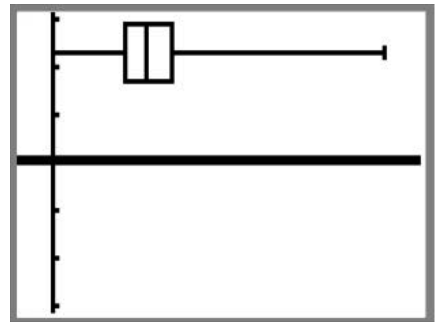
Press the **2nd** key and then press the **Y=** key to get to STATPLOT.
Press the number **1** key.
Move the cursor to On and press the **ENTER** key.
Use the arrow keys and highlight the box-and-whisker plot picture.
Type L1 in for Xlist:



Set the Freq: to 1.

Press the GRAPH key.
Press the **ZOOM** key.
Press the number **9** key.

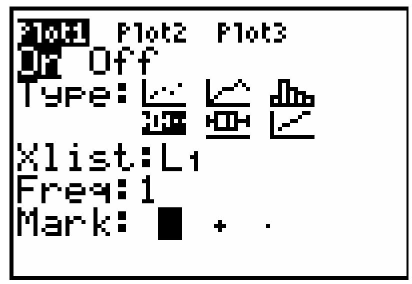
(to select "zoomstat")



We have two different types of box-and-whisker plots to select from. One will separate outliers from the maximum or minimum value (the modified boxplot) and the other will include the outliers in to whiskers (the boxplot).

If you select the picture that shows two dots after the maximum on the plot, the outliers will be shown outside of the whiskers.

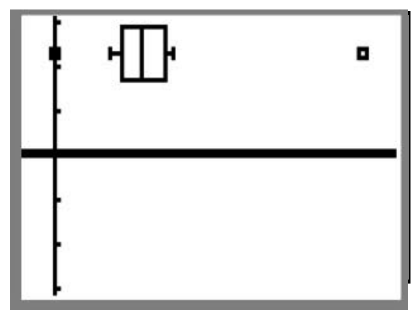
This is how you graph the "modified boxplot."



Press the "trace" button, and use your arrow keys to find the five-number summary and the outliers.

min = 2   Q1 = 125   Q2 = 162   Q3 = 203.5   max = 350

outliers = 2,   5,   350



2.6 WEBLINK — Making Percentile Graphs with StatCrunch (Extra Credit Lab)

http://timbusken.com/lab-OGIVE.html

3 Probability

3.1 Basic Concepts of Probability

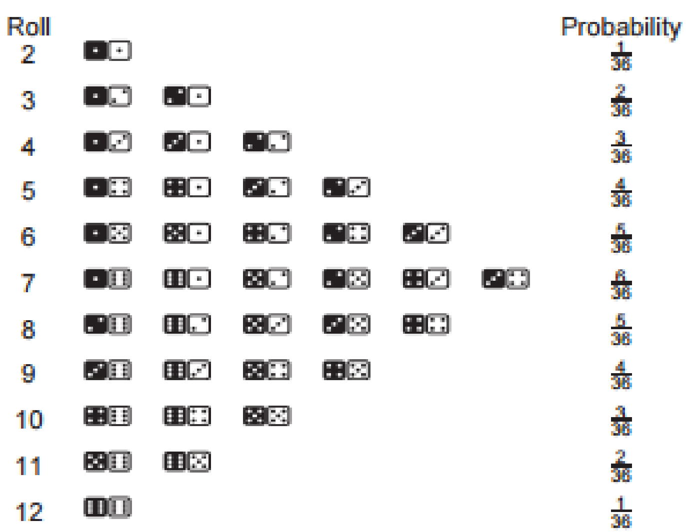
Learning objectives:

1. How to identify the sample space of a probability experiment and how to identify simple events
2. How to use the Fundamental Counting Principle to find the number of ways two or more events can occur
3. How to distinguish among classical probability, empirical probability, and subjective probability
4. How to find the probability of the complement of an event
5. How to use a tree diagram and the Fundamental Counting Principle to find probabilities

---

- **Probability** as a general concept can be defined as the chance of an event occurring.
- Processes such as flipping a coin, rolling a die, or drawing a card from a deck are called *probability experiments*.
- A **probability experiment** is an action, or chance process, through which specific results (counts, measurements, or responses) are obtained.
- A **trial** means flipping a coin once, rolling one die once, or the like. A trial is the action part of the experiment.
- An **outcome** is the result of a single trial in a probability experiment.
- When a coin is tossed, there are two possible *outcomes*: head or tail. In the roll of a single die, there are six possible *outcomes*: 1, 2, 3, 4, 5, or 6.
- A **sample space** is the set of all possible outcomes of a probability experiment.

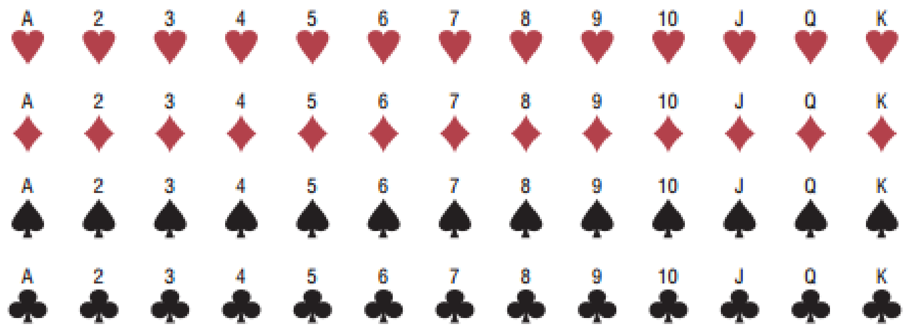
Example: Probability Experiment: Roll two dice and record the sum of the two numbers on the faces of the dice. That sum will equal a whole number between and including 2 and 12. There are 36 outcomes in the sample space. Each outcome is illustrated in the picture below.



Some more sample spaces for various probability experiments are shown here

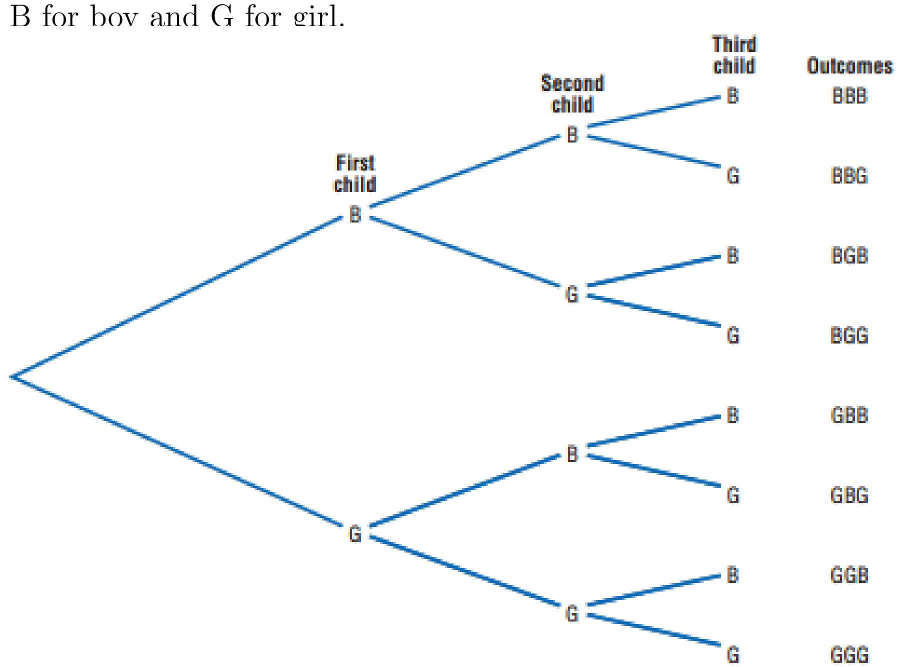
Experiment	Sample space
Toss one coin	Head, tail
Roll a die	1, 2, 3, 4, 5, 6
Answer a true/false question	True, false

Example Find the sample space for drawing one card from an ordinary deck of cards.

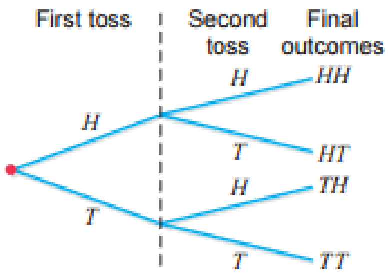


Definition A ***tree diagram*** is a device consisting of line segments emanating from a starting point and also from the outcome point. It is used to determine all possible outcomes of a probability experiment.

Example Use a tree diagram to find the sample space for the gender if a family has three children. Use B for boy and G for girl.



Example Use a tree diagram to find the sample space for the experiment tossing a coin twice.



Your Turn! A coin is flipped and a die is rolled. Use a tree diagram to find the sample space for the sequence of events. How many outcomes are in the sample space? What is the probability that the coin lands on tails and the die shows a 4?

Your Turn! Three coins are tossed. Make a tree diagram and determine the number of outcomes in the sample space. List the sample space. Also find the probability that

- no heads are tossed
- exactly one head is tossed
- three heads are tossed

More Definitions!

- An outcome was defined previously as the result of a single trial of a probability experiment. In many problems, one must find the probability of two or more outcomes. For this reason, it is necessary to distinguish between an outcome and an event.
- An ***event*** consists of a set of outcomes of a probability experiment.
- In the rest of this chapter, you will learn how to calculate the probability of an event. Events are often represented by uppercase letters, such as A, B or C.
- An event can be one outcome or more than one outcome. For example, if a die is rolled and a 6 shows, this result is called an *outcome*, since it is a result of a single trial.
- An event with one outcome is called a ***simple event***.
- The event of getting an odd number when a die is rolled is called a ***compound event***, since it consists of three outcomes or three simple events. In general, a compound event consists of two or more outcomes or simple events

---

Identifying Simple Events

A card is randomly selected from a standard deck that includes two joker cards. Some events have been defined below. Determine the number of outcomes in each event. Then decide whether the event is a simple event or not. Explain your reasoning.

6. Event A: select a heart card
7. Event B: select a 4
8. Event C: select a joker card
9. Event D: select a 4 of hearts

Identifying Simple Events

You roll a six-sided die. Some events have been defined below. Determine the number of outcomes in each event. Then decide whether the event is a simple event or not. Explain your reasoning.

10. Event A: roll at least a 3
11. Event B: roll less than 4
12. Event C: roll an odd number
13. Event D: roll a 4

In some cases, an event can occur in so many different ways that drawing a tree diagram becomes too cumbersome, and it is not practical to write out all the outcomes. When this happens, use the Fundamental Counting Principle.

Fundamental Counting Principle

- If one event can occur in m ways and a second event can occur in n ways, the number of ways the two events can occur in sequence is m times n.
- In words, the number of ways that events can occur in sequence is found by multiplying the number of ways one event can occur by the number of ways the other event(s) can occur.
- The Fundamental Counting Principle can be extended for any number of events occurring in sequence.

Using the Fundamental Counting Principle

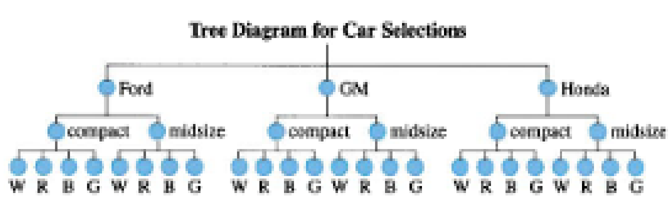
Example: Suppose you are purchasing a new car and the possible manufacturers, car sizes, and colors you have selected for your car are listed below.

Manufacturer: Ford, GM, Honda
Car size: compact, midsize
Color: white (W), red (R), black (B), green (G)

Question: How many different ways can you select one manufacturer, one car size, and one color? Use a tree diagram to check your result.

Answer: There are three choices of manufacturers, two car sizes, and four colors. Using the Fundamental Counting Principle:

$$(3)(2)(4) = 24 \; ways$$



ATM passwords have four digits. Each digit can be any number from 0 and 9.

14. How many ATM passwords are possible when each digit can be used only once and not repeated?

15. How many ATM passwords are possible when each digit can be repeated?

16. How many ATM passwords are possible when each digit can be repeated but the first digit cannot be 0 or 1?

17. How many license plates can you make when a license plate consists of six (out of 26) alphabetical letters, each of which can be repeated?

18. How many license plates can you make when a license plate consists of six (out of 26) alphabetical letters, each of which cannot be repeated?

Probabilities can be expressed as fractions, decimals percentages. If you ask, "What is the probability of getting a head when a coin is tossed?" typical responses can be any of the following three.

"One-half." "Point-five" "fifty percent"

These answers are all equivalent.

Rounding Rule for Probabilities

- Probabilities should be expressed as reduced fractions or rounded to two or three decimal places. Percentages are also acceptable.
- When the probability of an event is an extremely small decimal, it is permissible to round the decimal to the first nonzero digit after the point. For example, 0.0000587 would be 0.00006.

There are three different ways to obtain a probability:
1. Classical probability
2. Empirical probability (or relative frequency probability)
3. Subjective probability

Classical (or theoretical) probability uses sample spaces to determine the numerical probability that an event will happen. Classical probability assumes that all outcomes in the sample space are equally likely to occur

$$P(E) = \frac{\text{Number of outcomes in E}}{\text{Total number of outcomes in the sample space}}$$

Exercises: use the Classical technique to find the probability of each event below.

19. Rolling a Die If a die is rolled one time, find these probabilities.
    (a) Of getting a 4
    (b) Of getting an even number
    (c) Of getting a number greater than 4
    (d) Of getting a number less than 7
    (e) Of getting a number greater than 0
    (f) Of getting a number greater than 3 or an odd number
    (g) Of getting a number greater than 3 and an odd number

20. If one card is drawn from a deck, find the probability of getting these results.
    (a) An ace
    (b) A diamond
    (c) An ace of diamonds
    (d) A 4 or a 6
    (e) A 4 or a club

Empirical (Relative Frequency) Probability Based on observations obtained from probability experiments. Relative frequency of an event.

$$P(E) = \frac{\text{frequency of event E}}{\text{Total frequencies in the distribution}} = \frac{f}{n}$$

The difference between classical and empirical probability is that classical probability assumes that certain outcomes are equally likely (such as the outcomes when a die is rolled), while empirical probability relies on actual experience to determine the likelihood of outcomes. Empirical probability is based on observations obtained from probability experiments.

Relative Frequency Probability Examples:

- The probability that the next car that comes out of an auto factory is a "lemon"
- The probability that a randomly selected family in San Diego owns a home
- The probability that an 80-year-old person will live for at least 1 more year
- The probability that a randomly selected California driver owns a Toyota Prius.

These probabilities cannot be computed using the classical probability rule because the various outcomes for the corresponding experiments are not equally likely.

Exercises: use the Empirical (Relative Frequency) technique of finding probability.

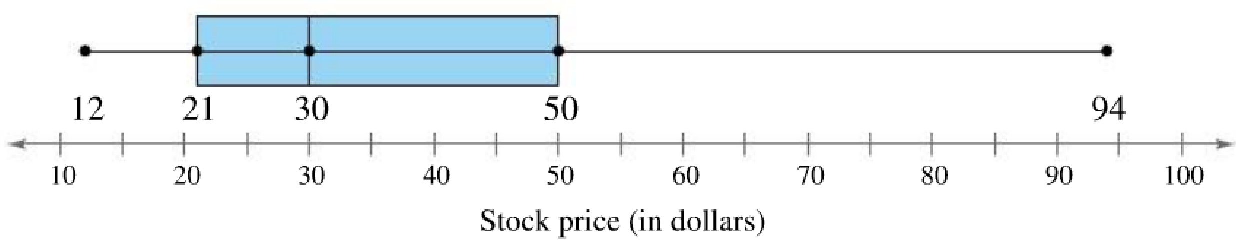
21. Hospital records indicated that maternity patients stayed in the hospital for the number of days shown in the distribution. Find these probabilities.
    (a) A patient stayed exactly 5 days.
    (b) A patient stayed less than 6 days.
    (c) A patient stayed at most 4 days.
    (d) A patient stayed at least 5 days. number

Number of days stayed	Frequency
3	15
4	32
5	56
6	19
7	5
	127

22. In a sample of 50 people, 21 had type O blood, 22 had type A blood, 5 had type B blood, and 2 had type AB blood. Set up a frequency distribution and find the following probabilities.
    (a) A person has type O blood.
    (b) A person has type A or type B blood.
    (c) A person does not have type AB blood number

23. On a construction site, there are 10 carpenters and 18 laborers; 3 carpenters and 5 laborers are females. A person working on the construction site is randomly selected. Set up a frequency distribution and find the following probabilities.
    (a) The person selected is a female
    (b) The person selected is a laborer
    (c) The person selected is a female carpenter

24. ***Try this!*** An individual stock is selected at random from the portfolio represented by the box-and-whisker plot shown. Find the probability that the stock price is (a) less than \$21, (b) between \$21 and \$50 and (c) \$30 or more.



Subjective probability uses a probability value based on an educated guess or estimate, employing opinions and inexact information. In subjective probability, a person or group makes an educated guess at the chance that an event will occur. This guess is based on the person's experience and evaluation of a solution.

- The probability that Carol, who is taking a statistics course, will earn an A in the course
- The probability that the Dow Jones Industrial Average will be higher at the end of the next trading day
- A doctor may feel a patient has a 90% chance of a full recovery.

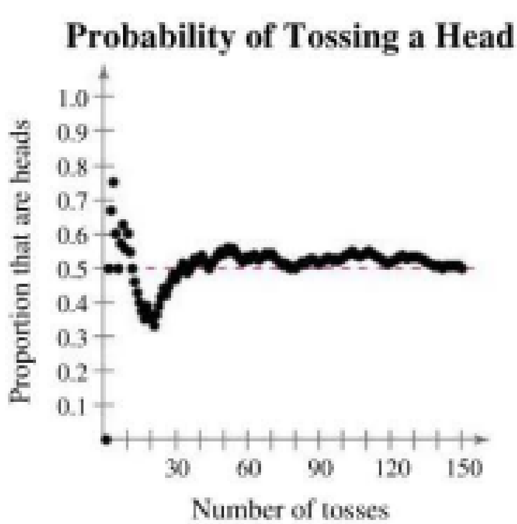
All three types of probability (classical, empirical, and subjective) are used to solve a variety of problems in business, engineering, and other fields.

---

25. ***<u>Exercises</u>*: Classify each statement as an example of classical probability, empirical probability, or subjective probability**
    (a) The probability that a person will watch the 6 o'clock evening news is 0.15.
    (b) The probability of winning the final round of wheel of fortune
    (c) The probability that a city bus will be in an accident on a specific run is about 6%.
    (d) The probability of getting a royal flush when five cards are selected at random is 1/649,740
    (e) An analyst feels that a certain stock's probability of decreasing in price over the next week is 0.75.

Law of Large Numbers

As an experiment is repeated over and over, the empirical probability of an event approaches the theoretical (actual) probability of the event.



Properties of Probability

1. The probability of any event E is a number (either a fraction or decimal) between and including 0 and 1. This rule states that probabilities cannot be negative or greater than 1.
2. If an event E cannot occur (i.e., the event contains no members in the sample space), its probability is 0
3. If an event E is certain, then the probability of E is 1 (or 100%).
4. The sum of the probabilities of all the outcomes in the sample space is 1.
5. An event is considered unusual if occurs with a probability of 5% or less.

The Complement of An Event

- Recall that an event is a set of outcomes.
- The **complement** of an event E is the set of all outcomes in the sample space that are not included in event E.
- The complement of E is denoted either as E' E^c or $\bar{E}$. For example, if we define an event A then the event's complement is denoted as A' A^c or $\bar{A}$.
- Using the definition of the complement of an event and the fact that the sum of the probabilities of all outcomes in the sample space is one, we can determine that

$$P(E) + P(E^c) = 1$$

- Write the complement rule here:

26. In 2013, 33% of LeastWorst Airlines customers who purchased a ticket spent an additional \$20 to be in the first boarding group. Choose one LeastWorst customer at random. What is the probability that the customer didn't spend the additional \$20 to be in the first boarding group?

27. A card is randomly selected from a standard deck of 52 playing cards. Some events have been defined below. Find the complement of each event.
    (a) Event A: the card is a four
    (b) Event B: the card is a heart card
    (c) Event C: the card is a four of hearts

3.2 Conditional Probability and the Multiplication Rule

1. How to find the probability of an event given that another event has occurred
2. How to distinguish between independent and dependent events
3. How to use the Multiplication Rule to find the probability of two or more events occurring in sequence and to find conditional probabilities

---

Definition A **conditional probability** of an event is a probability obtained with the additional information that some other event has already occurred. $P(B|A)$ denotes the conditional probability of event B occurring, given that (it is known that) event A has already occurred.

	Nonsmoker	Light Smoker	Heavy Smoker	Total
Men	306	74	66	446
Women	345	68	81	494
Total	651	142	147	940

Consider the following events:

Event N:	The person selected is a nonsmoker
Event L:	The person selected is a light smoker
Event H:	The person selected is a heavy smoker
Event M:	The person selected is a male
Event F:	The person selected is a female

Try These! 1. Suppose one of the 940 subjects is chosen at random. Determine the following probabilities:

a. $P(N|F)$

b. $P(F|N)$

c. $P(H|M)$

d. $P(\text{the person is a smoker})$

2. The human resources division at the Krusty-O cereal factory reports a breakdown of employees by job type and sex, summarized in the table below.

	Sex		
Job Type	Male	Female	total
Management	7	6	13
Supervision	8	12	20
Production	45	72	117
total	60	90	150

One of these workers is randomly selected.

(a) Find the probability that the worker is a female.

(a) ____________

(b) Find the probability that the worker is a female given she is a supervisor.

(b) ____________

(c) Find the probability that the worker is male with the Supervision job type.

(c) ____________

(d) Find the probability that the worker is female, given that the person works in production.

(d) ____________

(e) Find the probability that the worker works in production given that she is a female.

(e) ____________

(f) Find the probability that the worker is a male given he is in management.

(f) ____________

Notation $P(B|A)$ represents the probability of event B occurring after it is assumed that event A has already occurred (read $B|A$ as "B given A").

Definition Two events A and B are **independent** if the occurrence of one does not affect the probability of the occurrence of the other. If A and B are not independent, they are said to be **dependent** .

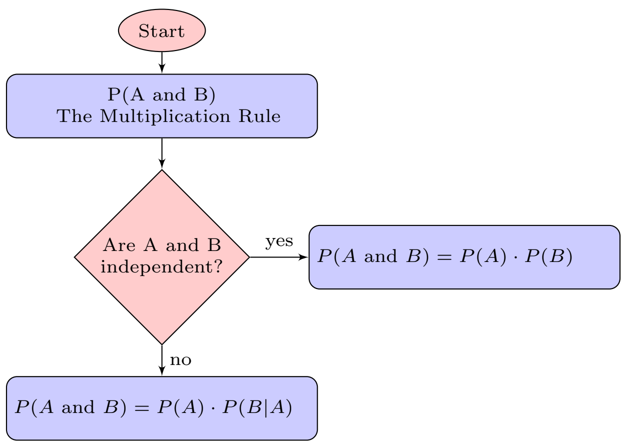
Definition Two events A and B are said to be **independent** if and only if either

$$P(B|A) = P(B) \qquad \text{or} \qquad P(A|B) = P(A)$$

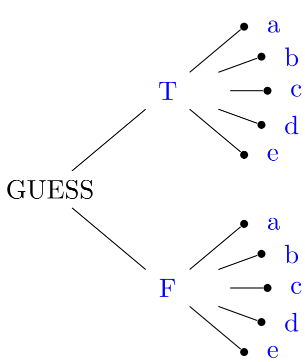
Theorem The Multiplication Rule

$$P(A \text{ and } B) = P(A) \cdot P(B) \qquad \text{(if A and B are independent)}$$

$$P(A \text{ and } B) = P(A) \cdot P(B|A) \qquad \text{(if A and B are dependent)}$$



Example: Suppose you are given a two-question quiz, where the first question is a true/false question and the second question is a multiple choice question with 5 possible answers. Suppose you guess on both questions. What is the probability that you correctly answered both questions?



Notice that the notation $P(\text{both correct})$ is equivalent to $P(\text{the first answer is correct AND the second answer is correct})$.

The sample space,

$$S = \{Ta, Tb, Tc, Td, Te, Fa, Fb, Fc, Fd, Fe\},$$

has 10 simple events.

Only one of these is a correct outcome, so

$$P(\text{both correct}) = \frac{1}{10} = 0.1$$

Suppose the correct answers are T and c. We can also obtain the correct probability by multiplying the individual probabilities:

$$P(\text{both correct}) = P(T \text{ and } c)$$
$$= P(T) \cdot P(c) = \frac{1}{2} \cdot \frac{1}{5} = \frac{1}{10} = 0.1$$

3. If 28% of U.S. medical degrees are conferred to women, find the probability that 2 randomly selected medical school graduates are women. Would you consider this event to be unusual?

4. Find the probability that 3 randomly selected medical school graduates are women. Would you consider this event to be unusual?

5. Find the probability that 3 randomly selected medical school graduates are men. Would you consider this event to be unusual?

6. A candy dish contains four red candies, seven yellow candies and fourteen blue candies. You close your eyes, choose two candies one at a time (without replacement) from the dish, and record their colors.
   (a) Find the probability that both candies are red.
   (b) Find the probability that the first candy is red and the second candy is blue.

7. Two cards are randomly selected ***without replacement***. Find the probability the first card is an ace and the second card is an ten.

8. Two cards are randomly selected ***with replacement***. Find the probability the first card is an ace and the second card is an ten.

9. Two cards are randomly selected ***without replacement***. Find the probability that the draw includes and ace and a ten.

10. Use the data in the following table, which summarizes blood type and Rh types for 100 subjects.

Blood Type

		O	A	B	AB
Rh Type	Rh^+	39	35	8	4
	Rh^-	5	6	2	1

If 2 out of the 100 subjects are randomly selected, find the probability that they are both blood group O and Rh type Rh^+.

(a) Assume that the selections are made with replacement.

(b) Assume that the selections are made without replacement.

	Nonsmoker	Light Smoker	Heavy Smoker	Total
Men	306	74	66	446
Women	345	68	81	494
Total	651	142	147	940

Consider the following events:

Event N: The person selected is a nonsmoker
Event L: The person selected is a light smoker
Event H: The person selected is a heavy smoker
Event M: The person selected is a male
Event F: The person selected is a female

11. Now suppose that two people are selected from the group, ***without replacement***. Let A be the event "the first person selected is a nonsmoker," and let B be the event "the second person is a light smoker." What is $P(A \cap B)$?

12. Two people are selected from the group. What is the probability that both people are smokers? Assume that the selections are made without replacement.

13. Three people are selected from the group. What is the probability that both people are smokers? Assume that the selections are made without replacement.

14. It is reported that 16% of households regularly eat Krusty-O cereal. Choose 4 households at random. Find the probability that

 (a) none regularly eat Krusty-O cereal

 (b) all of them regularly eat Krusty-O cereal

 (c) at least one regularly eats Krusty-O cereal

Let A be the event "a randomly selected household regularly eats Krusty-O cereal." Then $P(A) = 0.16$ and the complement of A (the event "a randomly selected household does not regularly eat Krusty-O cereal"), $P(\overline{A}) = 1 - P(A) = 1 - 0.16 = 0.84$.

 (a) P(none regularly eat Krusty-O cereal)

 = P(1st does not AND 2nd does not AND 3rd does not AND 4th does not)

 $= (0.84) \cdot (0.84) \cdot (0.84) \cdot (0.84) = (0.84)^4 = \boxed{0.4979}$

 (b) P(all 4 of them regularly eat Krusty-O cereal)

 = P(1st does AND 2nd does AND 3rd does AND 4th does)

 $= (0.16) \cdot (0.16) \cdot (0.16) \cdot (0.16) = (0.16)^4 = \boxed{0.000655}$

 (c) P(at least one regularly eats Krusty-O cereal)

 = 1 − P(none regularly eat Krusty-O cereal) $= 1 - 0.4979 = \boxed{0.5021}$.

15. 24% of teens go online "almost constantly," facilitated by the widespread availability of smartphones . (source: pew research 2013). Choose 3 teens at random. Find the probability that

 (a) none go online "almost constantly,"

 (b) all of them go online "almost constantly,"

 (c) at least one goes online "almost constantly,"

3.3 The Addition Rule

1. How to determine whether two events are mutually exclusive
2. How to use the Addition Rule to find the probability of two or more events

---

Definition Two events are mutually exclusive events if they cannot occur at the same time (*i.e.*, they have no outcomes in common).

1. Determine whether these events are mutually exclusive.
   (a) Roll a die: Get an even number, and get a number less than 3.
   (b) Roll a die: Get a prime number (2, 3, 5), and get an odd number.
   (c) Roll a die: Get a number greater than 3, and get a number less than 3.
   (d) Select a student in your class: The student has blond hair, and the student has blue eyes.
   (e) Select a student in your college: The student is a sophomore, and the student is a business major.
   (f) Select any course: It is a calculus course, and it is an English course.
   (g) Select a registered voter: The voter is a Republican, and the voter is a Democrat

---

The probability of two or more events can be determined by the addition rules. The first addition rule is used when the events are mutually exclusive.

Addition Rule 1

When two events A and B are mutually exclusive, the probability that A or B will occur is

$$P(A \text{ or } B) = P(A) + P(B)$$

---

2. At a political rally, there are 20 Republicans, 13 Democrats, and 6 Independents. If a person is selected at random, find the probability that he or she is either a Democrat or an Independent.

3. A single card is drawn at random from an ordinary deck of cards. Find the probability that it is either an ace or a king.

	Nonsmoker	Light Smoker	Heavy Smoker	Total
Men	306	74	66	446
Women	345	68	81	494
Total	651	142	147	940

Consider the following events:

Event N: The person selected is a nonsmoker
Event L: The person selected is a light smoker
Event H: The person selected is a heavy smoker
Event M: The person selected is a male
Event F: The person selected is a female

4. A person is randomly selected from this sample. What is the probability that person is a smoker?

5. A person is randomly selected from this sample. What is the probability that person is either a nonsmoker or a light smoker?

6. The human resources division at the Krusty-O cereal factory reports a breakdown of employees by job type and sex, summarized in the table below.

	Sex		
Job Type	Male	Female	total
Management	7	6	13
Supervision	8	12	20
Production	45	72	117
total	60	90	150

One of these workers is randomly selected.

(a) Find the probability that the worker is a supervisor or in management.

(b) Find the probability that the worker is a supervisor or a worker in production.

(c) Find the probability that the worker is not a supervisor.

(c) ____________

Addition Rule 2

When two events A and B are not mutually exclusive, the probability that A or B will occur is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Note: This rule can also be used when the events are mutually exclusive, since P(A and B) will always equal 0. However, it is important to make a distinction between the two situations.

---

7. A single card is drawn at random from an ordinary deck of cards. Find the probability that it is either an ace or a heart card.

	Nonsmoker	Light Smoker	Heavy Smoker	Total
Men	306	74	66	446
Women	345	68	81	494
Total	651	142	147	940

Consider the following events:

Event N: The person selected is a nonsmoker
Event L: The person selected is a light smoker
Event H: The person selected is a heavy smoker
Event M: The person selected is a male
Event F: The person selected is a female

8. A person is randomly selected from this sample. What is the probability that person is a nonsmoker or a male?

9. A person is randomly selected from this sample. What is the probability that person is either a female or a light smoker?

10. The human resources division at the Krusty-O cereal factory reports a breakdown of employees by job type and sex, summarized in the table below.

	Sex		
Job Type	Male	Female	total
Management	7	6	13
Supervision	8	12	20
Production	45	72	117
total	60	90	150

One of these workers is randomly selected.

(a) Find the probability that the worker is a supervisor or a female.

(b) Find the probability that the worker is in management or a male.

11. In a hospital unit there are 8 nurses and 5 physicians; 7 nurses and 3 physicians are females. If a staff person is selected, find the probability that the subject is a nurse or a male.

3.4 Counting Techniques

Counting the number of simple events in a sample space is one of the hardest problems to deal with when finding probabilities.

The Multiplication Rule
For a sequence of two events in which the first event can occur m ways and the second event can occur n ways, the events together can occur a total of $m \cdot n$ ways.

---

1. Suppose you are given a two-question quiz, where the first question is a true/false question and the second question is a multiple choice question with 5 possible answers. Suppose you guess on both questions. How many outcomes are in the sample space?

   1. ____________

2. Suppose you have 4 pairs of jeans, 12 clean T-shirts, and 4 wearable pairs of sneakers. How many outfits (jeans, T-shirt, and sneakers) can you create?

   2. ____________

3. Suppose you roll a pair of dice and record the sum of the two numbers that land on the upper faces of the die. How many outcomes are in the sample space?

   3. ____________

4. Suppose a couple plans to have three children. How many simple events are in the sample space? Find the probability of getting two boys and a girl?

   4. ____________

5. Suppose you have 3 different cars and a 3-car garage. How many different ways can you arrange (order) the way you park the cars in your garage?

   5. ____________

6. Suppose you have 4 math books you refuse to sell or throw away. How many different ways can you arrange (order) the books on a bookshelf?

   6. ____________

The Factorial Symbol Definition

The factorial symbol ! denotes the product of decreasing positive whole numbers. For example,

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

By special definition, $0! = 1$.

7. What number is 6! equal to?

7. ____________

8. What number is 8! equal to?

8. ____________

9. What number is 12! equal to?

9. ____________

Definition (Factorial Rule)

A collection of n different items can be arranged in order $n!$ different ways. (This factorial rule reflects the fact that the first item may be selected in n different ways, the second item may be selected in $n - 1$ ways, the third item may be selected $n - 2$ ways, and so on.)

10. A debate has five different speakers. How many different ways can the speakers be arranged from left to right on the debate stage?

10. ____________

Sometimes we have n different items to arrange, but we need to select some of them instead of all of them.

Example: Suppose a television producer has four prizes to give away to a studio audience of 50 people. The first prize is a car, the second prize is a \$6000 TV, third prize is a \$2500 gift certificate to the mall, and fourth prize is \$500 cash. How many different ways can the producer select the four prize winners?

$$50 \cdot 49 \cdot 48 \cdot 47 = 5,527,000 \qquad \text{using the Multiplication Rule}$$

Another way to obtain the same result is to evaluate $\dfrac{50!}{46!}$, since

$$\frac{50!}{46!} = \frac{50 \cdot 49 \cdot 48 \cdot 47 \cdot \cancel{46!}}{\cancel{46!}} = 50 \cdot 49 \cdot 48 \cdot 47 = 5,527,000$$

This result is generalized by the *permutations rule*: if we have n different items available and we want to select r of them, then the number of different orderings is $n!/(n - r)!$

Permutations Rule
Requirements:

1. There are n different items available, with none of the items identical to any other item under consideration.
2. We select r of the n items (without replacement).
3. The ordering of the selections matter.

The number of permutations (or sequences) of r items selected from n available items (without replacement), denoted ${}_nP_r$, is

$$ {}_nP_r = \frac{n!}{(n-r)!} $$

11. There are 13 members on a board of directors. How many different ways can the group select a president, vice-president and treasurer?

11. ____________

12. There are 10 horses in a particular race. How many ways can the horses finish first, second and third?

12. ____________

Distinguishable Permutations
The number of distinguishable permutations on n objects, where n_1 are of one type, n_2 are of another type, and so on, is

$$ \frac{n!}{n_1! \cdot n_2! \cdot n_3! \cdots n_k!} $$

where $n_1 + n_2 + n_3 + \cdots + n_k = n$

13. You may want to order a group of n object in which some objects are the same. For instance, consider the letters AAAABBC. This group has four A's, two B's, and one C. How many ways can you order such a group?

13. ____________

14. A student advisory board consists of 20 members. Two members serve as the board's chair and secretary. Each member is equally likely to serve in either of the positions. What is the probability of selecting at random the two members who currently hold the two positions?

14. ____________

Combinations Rule
Requirements:

1. There are n different items available.
2. We select r of the n items (without replacement).
3. The ordering of the selections does not matter.

The number of combinations of r items selected from n available items (without replacement), denoted ${}_nC_r$, is

$$ {}_nC_r = \frac{n!}{(n-r)!r!} $$

---

15. There are 13 members on a board of directors. How many different ways can the group form a subcommittee with 3 members?

16. A study is conducted in a hospital to determine the attitudes of nurses towards various administrative procedures. If a sample of 10 nurses is to be selected from a total of 90, how many different samples can be selected?

17. Five cards are selected from a 52 card deck for a poker hand. How many possible poker hands can be dealt?

4 Discrete Probability Distributions

4.1 Probability Distributions

Learning objectives:

1. How to distinguish between discrete random variables and continuous random variables
2. How to construct a discrete probability distribution and its graph and how to determine if a distribution is a probability distribution
3. How to find the mean, variance, and standard deviation of a discrete probability distribution
4. How to find the expected value of a discrete probability distribution

---

Definition: A ***probability distribution*** is a two-column (or two row) table in which the first column lists all the outcomes from a probability experiment, and column two is a list of probabilities associated with each outcome.

Example: Consider again the probability experiment: roll a single die. The probability distribution is then

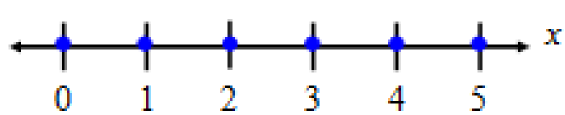
outcomes, x	1	2	3	4	5	6
probabilities, $p(x)$	1/6	1/6	1/6	1/6	1/6	1/6

Definition: A ***Random Variable*** Represents a numerical value associated with each outcome of a probability experiment.

- We use x to represent a ***Random Variable***
- Random variables can be Discrete or Continuous
- When a Random variable is Discrete (Chapter 4) the probability experiment has a finite (countable) number of outcomes.
- When a Random variable is Continuous (Chapter 5) the probability experiment has a infinite (uncountable) number of outcomes.

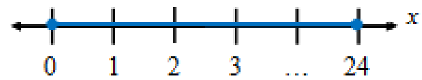
Definition: A ***Discrete Random Variable*** has a finite or countable number of possible outcomes that can be listed.

$x =$ Number of sales calls a salesperson makes in one day.



Definition: A ***Continuous Random Variable*** has an uncountable number of possible outcomes, represented by an interval on the number line.

$x =$ Number of hours a salesperson spends on the phone in one day



Exercises: Determine if the random variable is discrete or continuous.

1. x represents the number of people out of 20 with type O^+ blood
2. x represents the number of households that regularly eat Krusty-O cereal.
3. x represents the number of wins by the Boston Celtics
4. x represents the height of an ocean's tide at your favorite beach
5. x represents the length of a king salmon
6. x represents the number of gallons of milk produced by a single cow
7. x represents the number of aircraft near-collisions in a year

---

Definition: A ***Discrete probability distribution*** lists each possible value the random variable can assume, together with its probability.

THEOREM Every ***Discrete probability distribution*** has the following two properties

1) $0 \leq P(x) \leq 1$ The probability of each value of the random variable is between (and including) 0 and 1.

2) $\sum P(x) = 1$ The sum of all the probabilities in the distribution table is 1 (or 100%).

Example Construct a probability distribution using the census data (below) from the golf ball factory, and graph the probability distribution using a relative frequency histogram. The discrete random variable, x, of your probability distribution should represent the number of overtime hours worked in one week per employee at the golf ball factory.

overtime hours	0	1	2	3	4	5	6
employees	6	12	29	57	42	30	16

Example Sketch a hand-drawn graph of your probability distribution using a relative frequency histogram.

Exercises: Use your probability distribution to answer the probabilities questions below.

1. Find the probability of randomly selecting an employee whose overtime every week is :
   (a) exactly 1 hour (a) ____________

   (b) exactly two or three hours (b) ____________

   (c) between two and five hours, inclusive (c) ____________

2. Would it be unusual for an employee to work two hours of overtime? Use the 5% rule that was introduced in Chapter 3.

3. Find $P(X > 4)$

Measures of Center and Variation for discrete random variables, x, and their probability distributions

$\mu = \sum[x \cdot P(x)]$	Mean
$\sigma^2 = \sum[(x-\mu)^2 \cdot P(x)]$	Variance
$\sigma^2 = \sum[x^2 \cdot P(x)] - \mu^2$	Variance (the shortcut formula for variance)
$\sigma = \sqrt{\sigma^2} = \sqrt{\sum[x^2 \cdot P(x)] - \mu^2}$	Standard Deviation

Here is a link to my Chapter 4 pdf slides (pdf web link below) that will show you an example of how to find the mean, variance and standard deviation of a probability distribution using these formulas.
http://timbusken.com/assets/statistics/chapter-4-larson.pdf

VIDEO LINK: How to use the calculator to find mean, variance and standard deviation

https://www.youtube.com/watch?v=cI8lCZoJr1c&t=0s

APPLICATION

4. Find the Mean, Variance and Standard Deviation of the probability distribution table (below), and interpret/explain the meaning of each numerical measure as it pertains to the discrete random variable, x. The discrete random variable, x, represents the number of school related extracurricular activities per student.

Activities, x	0	1	2	3	4	5	6	7
Probability, $p(x)$	0.059	0.122	0.163	0.178	0.213	0.128	0.084	0.053

5. What is the expected number of extracurricular activities per student at this school?

Expected Value Word Problems The expected value of a discrete random variable of a probability distribution is the theoretical average of the variable. Write the formula for expected value here.

The symbol $E(x)$ is used for expected value.

APPLICATIONS

6. **Find the expected value.** Suppose you play a "Pick 4 Lotto where you pay 50 cents to select a sequence of four digits, such as 2118. If you select the same sequence of digits that are drawn, you win and collect \$2000.
   (a) How many different selections are possible?
   (b) What is the probability of winning?
   (c) If you win, what is your net profit?
   (d) Find the expected net gain.

7. **Find the expected value.** A Lottery offers one \$1000 prize, one \$500 prize, and five \$100 prizes. One thousand tickets are sold at \$3 each. Find the expected net gain if a person buys one ticket.

4.2 The Binomial Probability Distribution

Many probability experiments have only two outcomes. For example, when you guess at a multiple choice question, your answer is either right or wrong. A medical treatment can be considered effective or ineffective. Many survey questions can have only two possible answers: yes or no. When a coin is tossed it can land either heads or tails. Situations like these are called binomial experiments.

Binomial experiments are defined by the following properties:

1. The procedure has a fixed number of trials.
2. The trials must be independent. (The outcome of any individual trial doesn't affect the probabilities in the other trials.)
3. Each trial must have only two possible outcomes (commonly referred to as success and failure).
4. The probability of a success remains the same in all trials.

The outcomes of a binomial experiment and the corresponding probabilities of these outcomes are called a *binomial probability distribution*.

Notation for Binomial Probability Distributions:

n	denotes the fixed number of trials.
x	denotes a specific number of successes in n trials, so x can be any whole number between 0 and n, inclusive.
p	denotes the probability of success for a trial.
q	denotes the probability of failure for a trial, where $q = 1 - p$
$P(x)$ or $P(X = x)$	denotes the probability of getting exactly x successes among the n trials.

Here is an example of a binomial experiment: Pick a card from a standard deck and note whether or not the card is a club card. Then put the card back into the deck. Repeat the experiment five times, so $n = 5$. The outcomes of each trial can be classified in two categories: S = selecting a club and F = selecting another suit. The probability of success and failure are

$$p = \frac{1}{4} \quad \text{and} \quad q = 1 - \frac{1}{4} = \frac{3}{4}$$

The random variable x represents the number of clubs selected in five trials. So the possible values of the random variable are

$$0, 1, 2, 3, 4 \; and \; 5.$$

For instance, if $x = 2$, then exactly two of the five cards are clubs and the other three are not clubs.

Identifying Binomial Experiments

Determine whether the experiment is a binomial experiment. If it is, specify the values of n, p and q, and list the possible values of the random variable, x. If it is not, explain why.

1. A certain surgical procedure has an 85% chance of success. A doctor performs the procedure on eight patients. The random variable x represents the number of successful surgeries.

2. A jar contains five red marbles, nine blue marbles and six green marbles. You randomly select three marbles, *without replacement*. The random variable x represents the number of red marbles.

There are several ways to find the probability of x successes in n trials of a binomial experiment. One way is to use a tree diagram and the Multiplication Rule. Another way is to use the **binomial probability formula**.

The Binomial Probability Formula

$$P(x) = \left({}_nC_x \right) \cdot p^x \cdot q^{n-x}$$

for $x = 0, 1, 2, ..., n$, and recall that ${}_nC_x = \dfrac{n!}{(n-x)! \cdot x!}$.

Example: Rotator cuff surgery has a 90% chance of success. The surgery is performed on three patients. Find the probability of the surgery being successful on exactly two patients.

1. Use a tree diagram and the Multiplication Rule to find the probability.

2. Use the binomial probability formula to find the probability.

By listing the possible values of x with the corresponding probabilities, you can construct a **binomial probability distribution** (table).

Example: Construct a binomial probability distribution for the rotator cuff surgery example, using the binomial distribution formula.

Example: Verify that the numbers in your table are correct by comparing them to the ones given in the Binomial Tables (Link Below)
http://timbusken.com/assets/statistics/binomPDF.pdf

Example: One easy way to get the computer to give you the graph of your binomial distribution table from the rotator cuff surgery problem is to use the probability calculator on my website. (Link Below)
http://timbusken.com/prob-calc.html

1. ***Example*** Let X be a binomial random variable with $n = 12$ and $p = 0.3$.
   Use the Binomial Tables (Link Below) to obtain the correct probability distribution needed to answer probability questions.
   http://timbusken.com/assets/statistics/binomPDF.pdf

2. ***Find the probabilities given below.***
   (a) $P(X = 5)$ (a) ____________

   (b) $P(X = 8)$ (b) ____________

   (c) $P(X \leq 3)$ (c) ____________

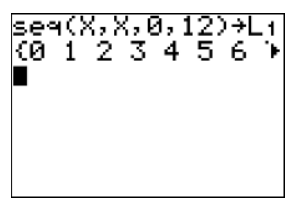
   (d) $P(X < 4)$ (d) ____________

   (e) $P(X < 2)$ (e) ____________

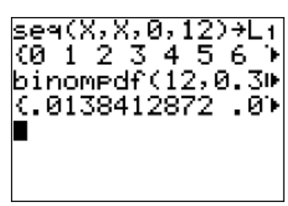
How to put a Binomial Distribution Table in your calculator

VIDEO LINK https://www.youtube.com/watch?v=_IJRsPay54g&t=0s

Step 1: Put the numbers 0 through 12 in L1.
Use $\text{seq}(X, X, 0, 12) \rightarrow \text{L1}$



Step 2: Use $\text{binomPdf}(n, p) \rightarrow \text{L2}$
to put the associated probabilities in L2.



Step 3: Open your list environment (stat→enter) and notice the distribution table is in L1 and L2.

Example Let X be a binomial random variable with $n = 12$ and $p = 0.3$. Find the following:

(a) $P(X > 9)$ (a) ____________

(b) $P(X \geq 4)$ (b) ____________

(c) $P(3 \leq X \leq 8)$ (c) ____________

(d) $P(6 < X < 10)$ (d) ____________

Measures of Center and Variation for the Binomial Prob. Distn.

$\mu = n \cdot p$	Mean
$\sigma^2 = n \cdot p \cdot q$	Variance
$\sigma = \sqrt{\sigma^2} = \sqrt{npq}$	Standard Deviation

---

3. ***Example*** Find the mean, variance and standard deviation of the random variable X from the previous problem using
   (a) the formulas above, and
   (b) using the calculator's ***1-var-stats*** command.

APPLICATIONS

1. A survey found that 53% of men believe there is a link between playing violent video games and teens exhibiting violent behavior. You randomly select five U.S. men and ask them whether they believe there is a link between playing violent video games and teens exhibiting violent behavior.
   (a) Find the probability that exactly two of the five men selected respond yes,
   (b) Find the probability that at least two of the five men respond yes, and
   (c) Find the probability that fewer than two of the five men respond yes.

2. Assume that 13% of people are left-handed. If we select 42 students at random, find the probability of each outcome described below. Use a binomial probability distribution.
   (a) There is at least one lefty in the group
   (b) There are exactly 3 lefties in the group
   (c) There are not more than 3 lefties in the group
   (d) What is the expected number of left-handers in the group?
   (e) What is the standard deviation of the number of left-handers in the group of 42 students?

3. People with type O-negative blood are said to be "universal donors." About 7% of the U.S. population has this blood type. Suppose that 335 people show up at a blood drive. Use a binomial probability distribution.
   (a) What is the expected number of universal donors in the group?
   (b) What is the probability that exactly 3 universal donors are in the group?
   (c) What is the probability that more than 3 universal donors are in the group?
   (d) What is the variance of the number of universal donors in the group of 335?
   (e) What is the standard deviation of the number of universal donors in the group of 335?

OPTIONAL — Link to the binomial cumulative distribution tables.
http://timbusken.com/assets/statistics/binomCDF.pdf

OPTIONAL — More Calculator Facts — Binomial Model

Your calculator has two binomial functions in its list of distributions (probability models). These are the binomPdf and binomCdf functions.

- pdf stands for probability distribution function and gives the probability $P(X = x)$
- cdf stands for cumulative distribution function and gives the probability $P(X \leq x)$

Both functions take can take 3 input arguments: n, p, and x, each separated by a comma.

	TI–83/84	Let $x = 3$	TI–83/84	
$P(X = x)$	binomPdf(n, p, x)	$P(X = 3)$	binomPdf$(6, .4, 3)$	0.27648
$P(X \leq x)$	binomCdf(n, p, x)	$P(X \leq 3)$	binomCdf$(6, .4, 3)$	0.8208
$P(X < x)$	binomCdf$(n, p, x - 1)$	$P(X < 3)$	binomCdf$(6, .4, 2)$	0.54432
$P(X > x)$	1 – binomCdf(n, p, x)	$P(X > 3)$	1 – binomCdf$(6, .4, 3)$	0.1792
$P(X \geq x)$	1 – binomCdf$(n, p, x - 1)$	$P(X \geq 3)$	1 – binomCdf$(6, .4, 2)$	0.45568

5 Normal Probability Distributions

5.1 Normal Distributions and the Standard Normal Distribution

Section Learning objectives:

1. How to interpret graphs of normal probability distributions
2. How to find areas under the any normal curve

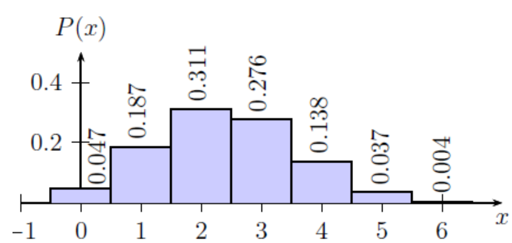
<u>**Definition**</u>: A ***Continuous Random Variable*** is a list of the infinite number of outcomes from a probability experiment.

<u>**Definition**</u>: The probability distribution associated with a Continuous Random Variable is called a ***Continuous Probability Distribution***.

In Chapter 4, you studied discrete random variables, and you learned that Discrete Probability Distributions can be described by (1) a formula or function definition, (2) a table or (3) a probability histogram.

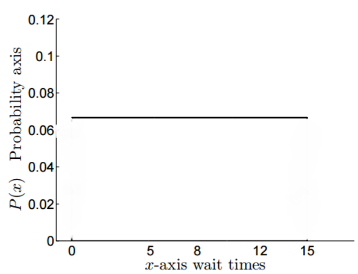
$$P(X = x)$$
$$= \left({}_nC_x\right) \cdot p^x \cdot q^{n-x}$$
$$= \left({}_6C_x\right) \cdot (0.4)^x \cdot (0.6)^{6-x}$$

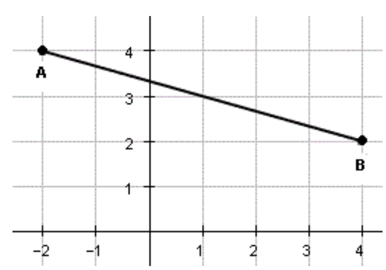
x	$P(X = x)$
0	0.047
1	0.187
2	0.311
3	0.276
4	0.138
5	0.037
6	0.004

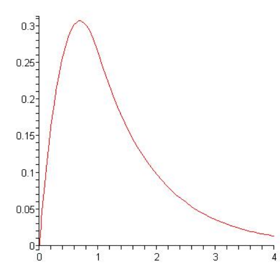


For example, the formula, table and graph above give a description of a discrete probability distribution for a binomial random variable with $n = 6$ and $p = 0.4$.

Continuous probability distributions can also be described by a table, graph, or function formula, called a **probability density function (PDF)**. However, since a continuous random variable can take on an infinite number of values it is impossible to list every row in the table. In addition, the graph of a continuous probability distribution looks like a function curve and not a histogram.







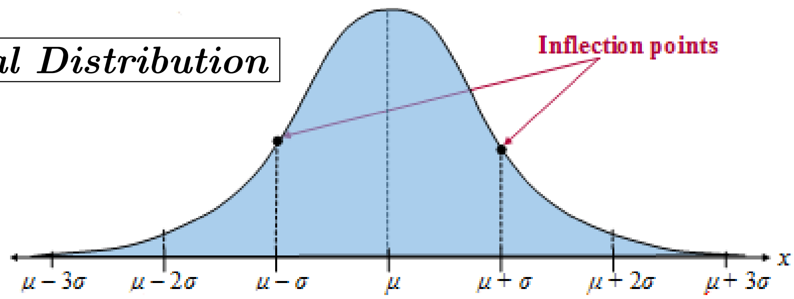
<u>**Definition**</u>: A **probability density function (PDF)** is the graph of a continuous probability distribution. It must satisfy the following properties:

1. The total area under the curve must equal 1.

2. Every point on the curve must have a vertical height that is 0 or greater. (That is, the curve cannot fall below the x-axis.)

In this chapter, we study the most important continuous probability distribution in statistics — ***the normal distribution***. Normal distributions are bell-shaped curves that model many sets of measurements in nature, industry and business. For instance, height, weights and blood pressures of humans, lengths of pregnancies, the lifetimes of washers and dryers and even housing costs are all normally distributed.

The Normal Distribution

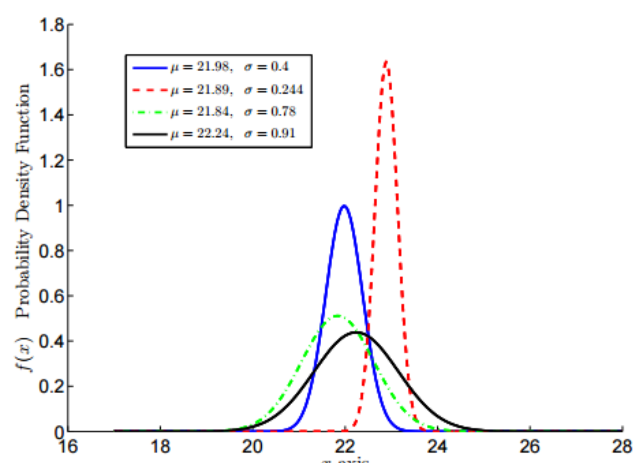


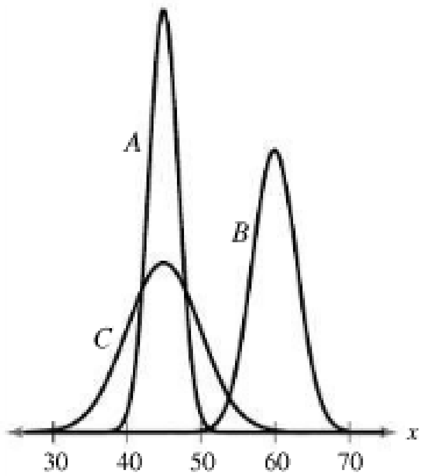
Properties of a Normal Distn.

1. The graph of a normal distribution is called the ***normal curve***.
2. A normal curve is bell-shaped and can be graphed using the function formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty \leq x \leq \infty$$

3. The mean, $x = \mu$, locates the center of the distribution. The vertical line, $x = \mu$ is an axis of symmetry for the PDF.
4. The total area under the curve is equal to one.
5. The normal curve approaches, but never touches the x-axis as it extends farther and farther away from the mean.
6. Between $\mu - \sigma$ and $\mu + \sigma$ (in the center of the curve), the graph curves downward. The graph curves upward to the left of $\sigma - \mu$ and to the right of $\sigma + \mu$. The points at which the curve changes from curving upward to curving downward are called the ***inflection points***.
7. The population standard deviation, σ, describes how spread out over the x-axis the PDF curve is. Large values of σ decrease the height of the peak and increase the spread of the distribution (along the x axis; small values of σ raise the height of the peak and decrease the spread.





Use the normal curves in the figure (left) to answer the following questions.

1. Which normal curve has the greatest mean?

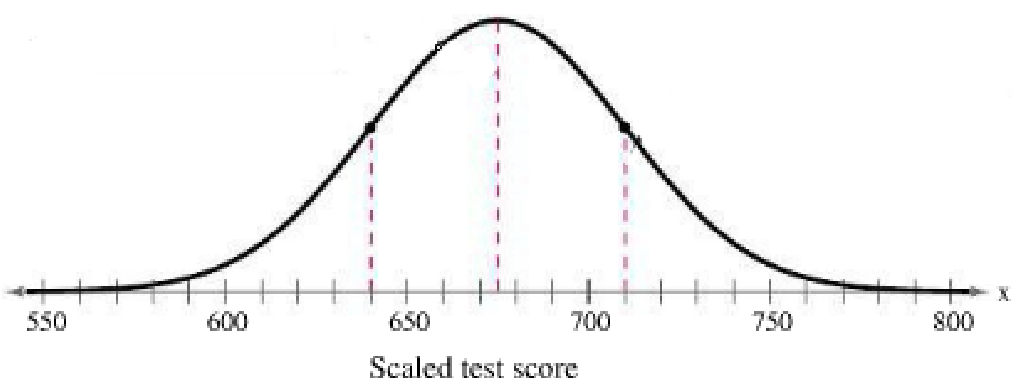
2. Which normal curve has the greatest standard deviation?

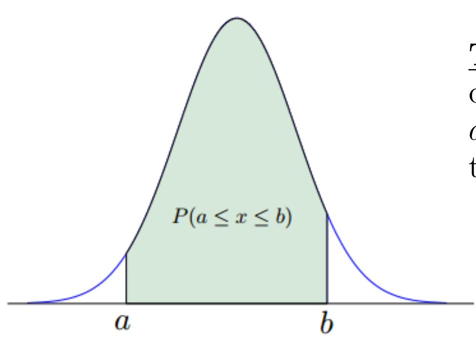
3. Which normal curve has the least standard deviation?

4. What equation describes the line (axis) of symmetry for curve C?

---

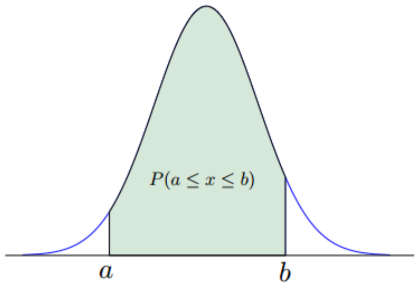
5. The scaled test scores for a New York State Grade 8 Math Test are normally distributed. The normal curve shown below represents this distribution. What is the mean test score? Estimate a value for the standard deviation of this normal distribution.





<u>**Theorem**</u>: The probability that a continuous random variable x assumes a value in the interval from a to b is the area under the probability density function between vertical lines $x = a$ and $x = b$.

6. Use your calculator's "normalCDF" command to find the probability that a randomly selected New York State Grade 8 Math Test score is between and including 600 and 650.



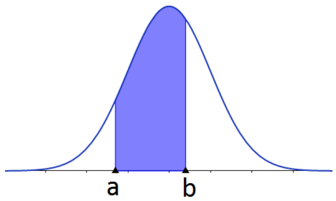
Theorem: The probability that a continuous random variable x assumes a value in the interval from a to b is the area under the probability density function between vertical lines $x = a$ and $x = b$.

Guidelines: How to find area under a normal curve that has mean, μ, and standard deviation, σ. Use the calculator's "normalCdf" command.

To access the normalCdf command press [2nd] [VARS] [2]

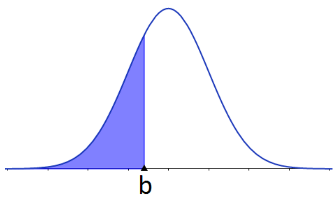
1) For $P(a \le x \le b)$ or $P(a < x < b)$

use $normalCdf(a, b, \mu, \sigma)$



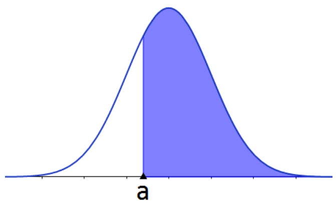
2) For $P(x \le b)$ or $P(x < b)$

use $normalCdf(-10^9, b, \mu, \sigma)$



3) For $P(x \ge a)$ or $P(x > a)$

use $normalCdf(a, 10^9, \mu, \sigma)$



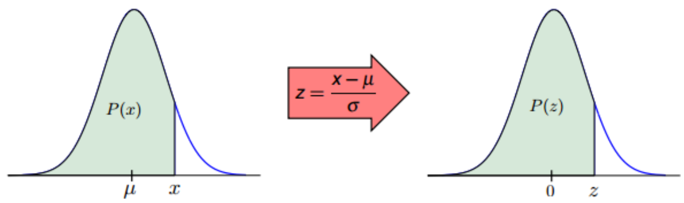
Example

7. A Honda Y2J diesel engine is used to run each packaging assembly line at the Lay's Potato Chip Factory. Let x represent the lifespan of a Honda Y2J diesel engine. Accepted values for the mean value and standard deviation of x are 25 years and 2 years, respectively. Suppose that the probability distribution of x is approximately normal. What is the probability that the lifespan of a randomly selected Honda Y2J diesel engine will be
   (a) not more than 23 years
   (b) at least 23 years
   (c) between 23 and 25 years
   (d) answer part a using the z-table

Definition: The phrase **"cumulative area"** is used to indicate an area under a curve that is LEFT of a specific number located on the horizontal axis.

Before we had calculators, statisticians had to find areas (probabilities) using calculus routines; or they used a "standardized table" (also called a "z-table") that lists several ***cumulative areas*** (probabilities) for one normal curve, called the ***Standard Normal Distribution***.

Definition: The ***Standard Normal Distribution*** is the normal distribution that has a mean, μ equal to 0, and a standard deviation, σ equal to 1.



Properties of the Standard Normal Distn.

- The Standard Normal Distribution is a collection of z scores.
- The Standard Normal Curve is centered at the origin.
- The mean, $\mu = 0$ and standard deviation, $\sigma = 1$.
- The cumulative area is close to 0 for z-scores close to $z = -3.49$.
- The cumulative area increases as the z-scores increase.
- The cumulative area for z = 0 is 0.5000 or 50%
- The cumulative area is close to 1 (or 100%) for z-scores close to z = 3.49.

How did statisticians use the Standard Normal Distribution and a "z-table" to find areas (probabilities) under *non-standard* normal curves?

The answer is that each data value of the normally distributed random variable x was transformed into a z score (using the z-score formula). This standardization process converts any non-standard normal curve whose mean is not 0 or whose standard deviation is not 1 to the **Standard Normal Distribution**.

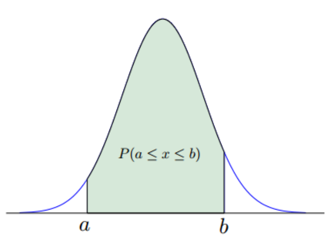


Figure 1: A Non-Standard Normal Distribution

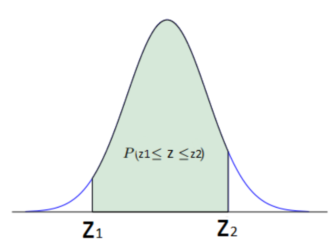


Figure 2: The Standard Normal Distribution

After this transformation takes place, **the area under the nonstandard normal curve that falls in the interval from 'a' to 'b' is the same as that under the standard normal curve within the corresponding z-scores of 'a' and 'b'**. The area between the two z-boundaries is then found in the z-table.

Link to z table

http://timbusken.com/assets/statistics/chapter-6/z-table.pdf

Link to "z-table explained"

http://timbusken.com/assets/statistics/chapter-5-larson.pdf

Link to the geogebra applet used in class

http://timbusken.com/normalcdf.html

Classroom Exercises

Use the Standard Normal Distribution to answer these questions. Use the z-table or the calculator.

1. Find the cumulative area that corresponds to a z score of -1.51.

2. Find the cumulative area that corresponds to a z score of 2.37.

3. Find the z score associated with a probability value of 0.8461

4. Find the z score associated with P_{90}, the 90th percentile.

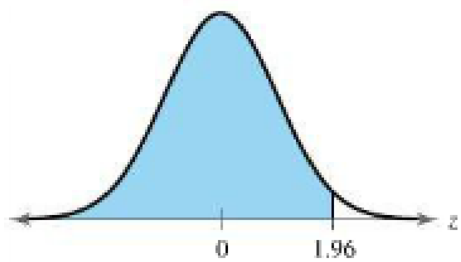
5. Find the probability $P(z < 1.35)$.

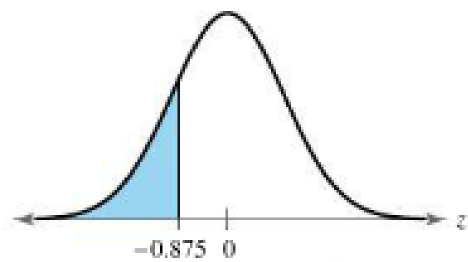
6. Find the probability $P(z < -0.18)$.

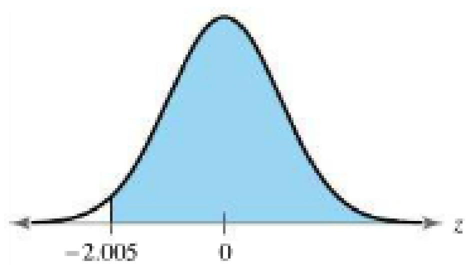
7. Find the probability $P(z > 2.358)$.

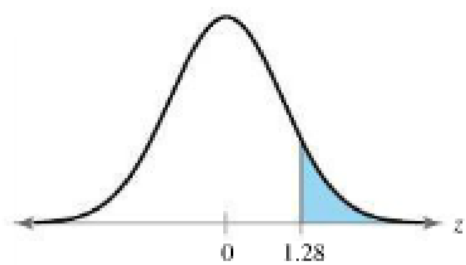
8. Find the probability $P(-1.2 < z < 0.18)$.

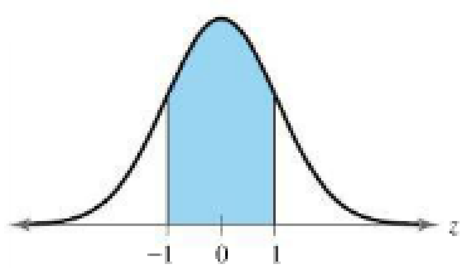
9. Find the probability of z occurring in the shaded region of each standard normal distribution.

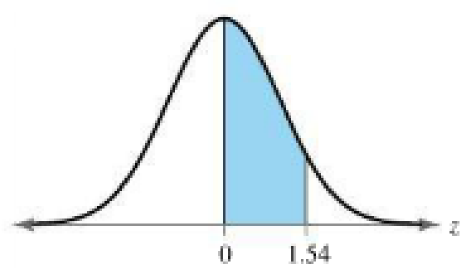












5.2 Normal Distributions: Finding Probabilities

Section Learning objective: How to find probabilities for normally distributed variables using a table and using technology

Classroom Exercises

In exercises 1 - 3, the random variable x is normally distributed with mean $\mu = 174$ and standard deviation $\sigma = 20$. Find each probability.

1. Find $P(172 < X < 192)$

2. Find $P(X < 200)$

3. Find $P(X > 155)$

4. **Heights of Men** In a survey of U.S. men, the heights in the 20 –29 age group were normally distributed, with a mean of 69.4 inches and a standard deviation of 2.9 inches. Find the probability that a randomly selected study participant has a height that is (a) less than 66 inches, (b) between 66 and 72 inches, and (c) more than 72 inches, and (d) identify any unusual events. Explain your reasoning.

5. **Utility Bills** The monthly utility bills in a city are normally distributed, with a mean of \$100 and a standard deviation of \$12. Find the probability that a randomly selected utility bill is (a) less than \$70, (b) between \$90 and \$120, and (c) more than \$140.

5.3 Normal Distributions: Finding Values

Section Learning objectives:

- How to find a z-score given the area under the normal curve
- How to transform a z-score to an x-value
- How to find a specific data value of a normal distribution given the probability

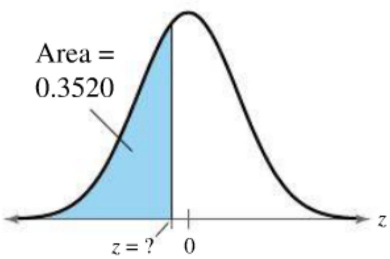
---

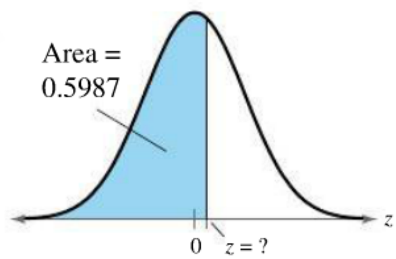
In exercises 1 - 4, use the Standard Normal Table or your calculator's invnorm command to find the z-score that corresponds to the cumulative area or percentile. Round each value of z to the hundredths.

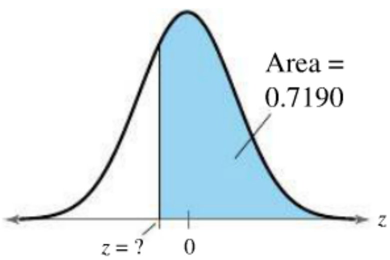
1. 0.4321
2. 0.88
3. P_{35}
4. P_{40}

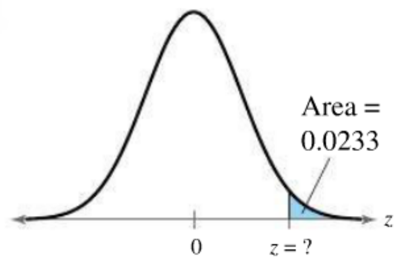
---

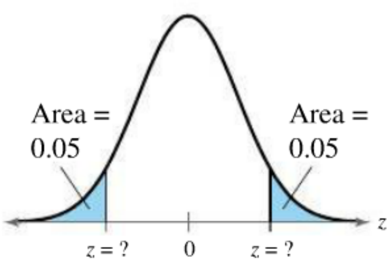
5. **Find the indicated z-score(s) shown in each graph (below). Round each value of z to the hundredths.**

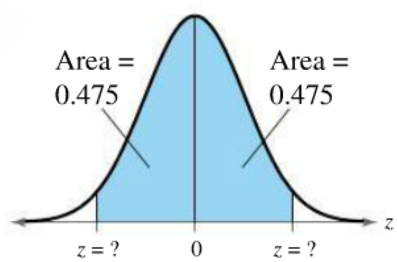




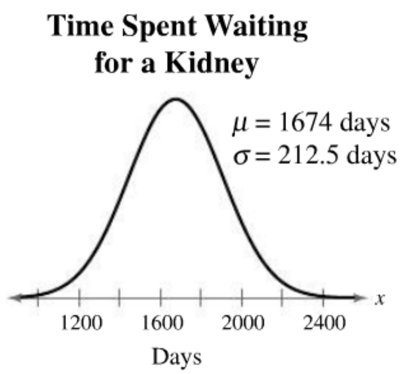








6. **Heights of Women** In a survey of women in the United States (ages 20 –29), the mean height was 64.2 inches with a standard deviation of 2.9 inches. Assume that heights are normally distributed. (Adapted from National Center for Health Statistics)
   (a) What height represents the 95th percentile?
   (b) What height represents the first quartile?



7. **Kidney Transplant Waiting Times** The time spent (in days) waiting for a kidney transplant for people ages 35– 49 can be approximated by a normal distribution, as shown in the figure. (Adapted from Organ Procurement and Transplantation Network)
   (a) What waiting time represents the 80th percentile?
   (b) What waiting time represents the first quartile?

8. **Bananas** The annual per capita consumption of fresh bananas (in pounds) in the United States can be approximated by a normal distribution, with a mean of 10.4 pounds and a standard deviation of 3 pounds. (Adapted from U.S. Department of Agriculture)
(a) What is the smallest annual per capita consumption of bananas that can be in the top 10% of consumptions?
(b) What is the largest annual per capita consumption of bananas that can be in the bottom 5% of consumptions?

5.4 Sampling Distributions and The Central Limit Theorem

Classroom Exercises

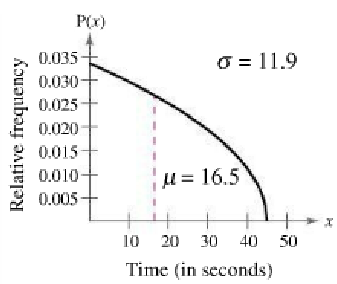
In exercises 1 and 2, a population has a mean of $\mu = 150$ and a standard deviation of $\sigma = 25$. Find the mean and standard deviation of the sampling distribution of sample means with sample size n.

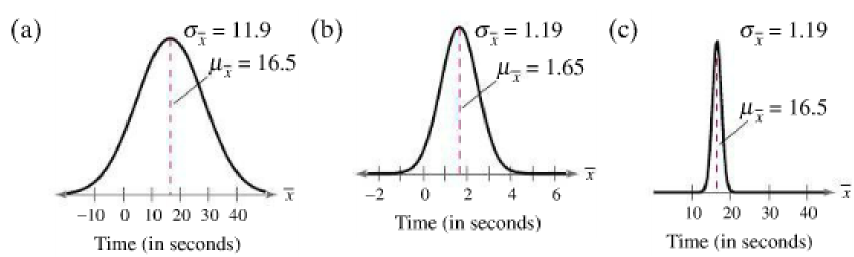
1. $n = 50$

2. $n = 100$

In the next exercise the graph of a population distribution is shown with its mean and standard deviation. A sample of size 100 is drawn from the population. Determine which of the figures labeled (a)-(c) would most closely resemble the sampling distribution of sample means. Explain your reasoning.

3. The waiting time (in seconds) at a traffic signal during a red light





4. For a sample of $n = 100$, find the probability of a sample mean being greater than 24.3 when $\mu = 24$ and $\sigma = 1.25$.

5. **Salaries** The mean annual salary for environmental compliance specialists is about \$66,000. A random sample of 35 specialists is selected from this population. What is the probability that the mean salary of the sample is less than \$60,000? Assume $\sigma = \$12,000$.

6. **Gas Prices: California** During a certain week, the mean price of gasoline in California was \$4.117 per gallon. A random sample of 38 gas stations is selected from this population. What is the probability that the mean price for the sample was between \$4.128 and \$4.143 that week? Assume $\sigma = \$0.049$.

7. **Heights of Men** The mean height of men in the United States (ages 20–29) is 69.4 inches. A random sample of 60 men in this age group is selected. What is the probability that the mean height for the sample is greater than 70 inches? Assume $\sigma = 2.9$ inches.

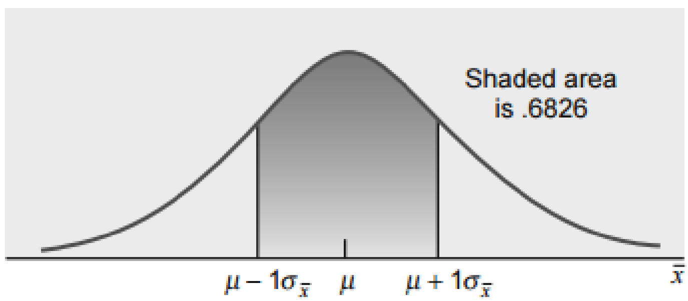
The Empirical Rule and the Central Limit Theorem

From the central limit theorem, for large samples, the sampling distribution of $\bar{x}$ is approximately normal with mean μ and standard deviation $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ Based on this result, we can make the following statements about $\bar{x}$ for large samples ($n \geq 30$). The areas under the curve of $\bar{x}$ mentioned in these statements come from the Empirical Rule, and can be found from the z-table or using the normalCdf calculator command.

1. If we take all possible samples of the same (large) size from a population and calculate the mean for each of these samples, then about 68.26% of the sample means will be within one standard deviation of the population mean. Alternatively, we can state that if we take one sample (of $n \geq 30$) from a population and calculate the mean for this sample, the probability that this sample mean will be within one standard deviation of the population mean is .6826. That is,

$$P(\mu - 1\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 1\sigma_{\bar{x}}) = .6826$$

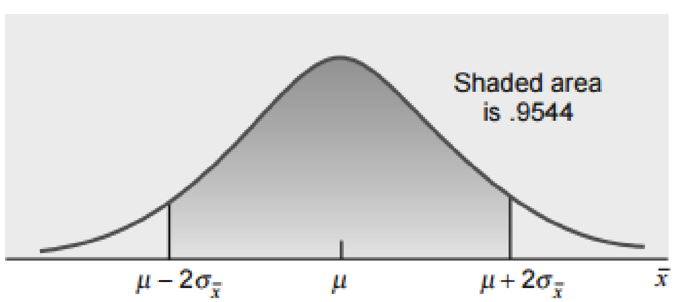
This probability is shown in the figure below.



2. If we take all possible samples of the same (large) size from a population and calculate the mean for each of these samples, then about 95.44% of the sample means will be within two standard deviations of the population mean. Alternatively, we can state that if we take one sample (of $n \geq 30$) from a population and calculate the mean for this sample, the probability that this sample mean will be within two standard deviations of the population mean is .9544. That is,

$$P(\mu - 2\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 2\sigma_{\bar{x}}) = .9544$$

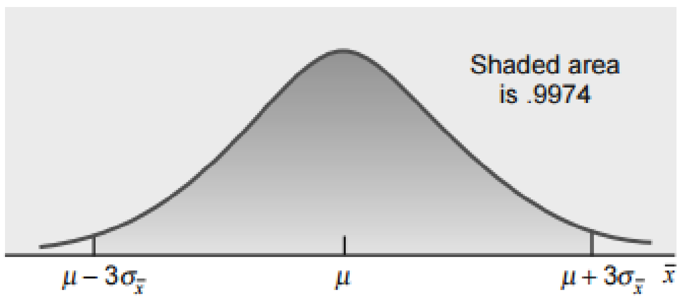
This probability is shown in the figure below.



3. If we take all possible samples of the same (large) size from a population and calculate the mean for each of these samples, then about 99.74% of the sample means will be within three standard deviations of the population mean. Alternatively, we can state that if we take one sample (of $n \geq 30$) from a population and calculate the mean for this sample, the probability that this sample mean will be within three standard deviations of the population mean is .9974. That is,

$$P(\mu - 3\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 3\sigma_{\bar{x}}) = .9974$$

This probability is shown in the figure below.



When conducting a survey, we usually select one sample and compute the value of $\bar{x}$ based on that sample. We never select all possible samples of the same size and then prepare the sampling distribution of $\bar{x}$. Rather, we are more interested in finding the probability that the value of $\bar{x}$ computed from one sample falls within a given interval.

6 Confidence Intervals and Sample Size

Chapter 6 shows us how we can use the Central Limit Theorem (CLT) to

1. estimate a population parameter (such as the mean or proportion) using a sample, and
2. determine how large should the sample be in order to make an accurate estimation.

Here is a web link to a web page I made with more information about Confidence Intervals!

6.1 Estimating a Population Mean μ with σ known

Section 6.1 shows us how to

1. estimate a population mean, μ, using a large sample, and how to
2. determine how large of a sample we should take in order to make an accurate estimation for μ.

Suppose a college student wants to estimate the average number of units per semester a student at her college takes. The student could take a random sample of 100 students and find the average number of units they are taking this semester. Suppose the mean of her sample, $\bar{x}$, is 7.3 units. The student could then use the sample mean to infer that the average number of units of all the students, μ, is 7.3 units. This type of estimate is called a point estimate.

Definition: A ***point estimate*** is a specific numerical value estimate of a parameter. The best point estimate of the population mean μ is the sample mean $\overline{X}$.

Usually different samples selected from the same population will give different values for sample mean, $\bar{x}$, because the samples contain different sets of numbers. Additionally, the mean obtained from any one sample will generally be not exactly equal to the population mean, μ. **This difference, given by the formula $\bar{x} - \mu$ is called *sampling error.***

Because *sampling error* exists, one might ask a the question: How good is a point estimate? The answer is that there is no way of knowing how close a particular point estimate is to the population mean. This answer places some doubt on the accuracy of point estimates. For this reason, statisticians prefer another type of estimate, called an interval estimate.

Definition: An ***interval estimate*** of a parameter is an interval or a range of values used to estimate the parameter. This estimate may or may not contain the value of the parameter being estimated.

In an interval estimate, the parameter is specified as being between two values. For example, an interval estimate for the average number of units all students take each semester might be $7.0 < \mu < 7.6$, or 7.3 $\pm$0.3 units. A degree of confidence (usually a percent) can be assigned before an interval estimate is made. For instance, you may want to be 95% confident that the interval contains the true population mean.

Definition: The ***confidence level,*** c**,** of an interval estimate of a parameter is the probability that the interval estimate will contain the parameter, assuming that a large number of samples are selected and that the estimation process on the same parameter is repeated.

Definition: A ***confidence interval*** is a specific interval estimate of a parameter determined by using data obtained from a sample and by using the specific confidence level of the estimate. Three common confidence intervals are used: the 90, the 95, and the 99% confidence intervals.

Definition: The ***margin of error,*** E (also called the maximum error in the estimate of a parameter), is the maximum likely difference between the point estimate, $\bar{x}$, and the population parameter μ.

$$\Big|\bar{x} - \mu\Big| \leq \text{E}$$

Formula for a c-percent Confidence Interval Estimate of a Population Mean μ

$$\overline{X} - E \ < \mu < \ \overline{X} + E$$

where margin of error, E, is

$$E = z_c \cdot \frac{\sigma}{\sqrt{n}}$$

and

$$z_c = invnorm\left(\frac{1}{2}(1 + c)\right)$$

where z_c is the z value corresponding to an area of $\frac{1}{2}(1 - c)$ in the right tail of a standard normal z distribution, and

$$\begin{aligned} n &= \text{sample size} \\ \sigma &= \text{standard deviation of the sampled population} \\ c &= \text{the confidence level} \end{aligned}$$

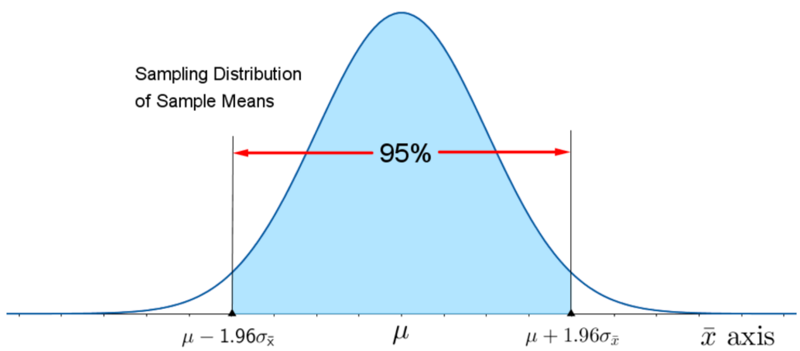
Assumptions:

1. The sample must be a random sample.
2. The value of σ is known or given along with the statement of the problem.
3. Either $n \geq 30$ or the population is normally distributed if $n < 30$, so that the Central Limit Theorem guarantees that the sampling distribution will be a normal curve centered at $\mu = \mu_{\bar{x}}$ with standard deviation, $\sigma_{\bar{x}}$, where $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$
4. If no confidence level, c, is given along with the statement of the problem, then use $c = 0.95$.
5. If the value of σ is unknown or not given, then we can use the sample standard deviation, s in place of σ in the margin of error formula, so long as $n \geq 30$.

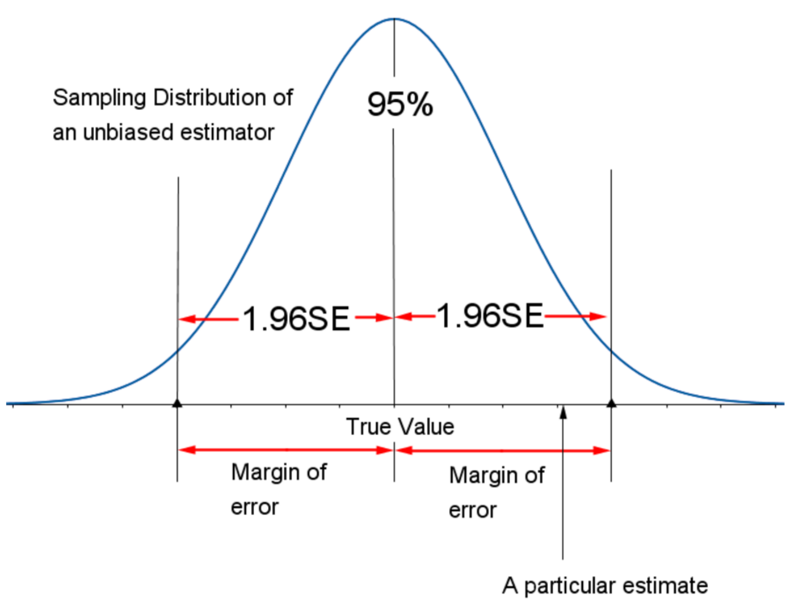
Example: A marketing analyst wants to estimate the average amount spent by a dating site customer per year. A random sample of $n = 50$ dating site customers were polled about the amount they spend each year on dating websites. The results of the poll produced a mean amount of \$240 with a standard deviation of \$20. ***Construct a 95% confidence interval estimate for the population mean amount spent by a dating website customer per year.***

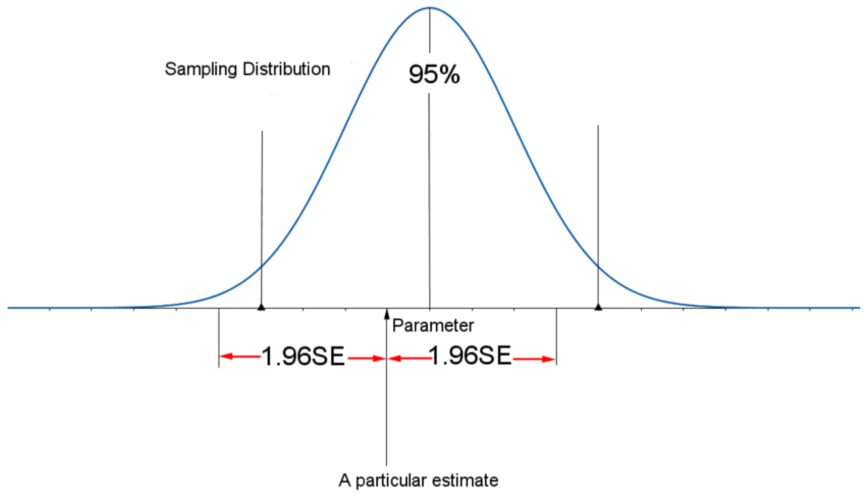
Definition: Recall that $\sigma_{\bar{x}}$ is the symbol we use for the standard deviation of the sampling distribution of $\bar{x}$. Recall that $\sigma_{\bar{x}}$ is also called the *standard error in estimating the mean*, or simply the *standard error* (SE).

For the work we do, you may assume that the sample sizes are always large and, therefore, that the estimators you will study have sampling distributions that can be approximated by a normal distribution (because of the Central Limit Theorem). Then, for any point estimator with a normal distribution, the Empirical Rule states that approximately 95% of all the point estimates will lie within two (or more exactly, 1.96) standard deviations of the mean of that distribution.



This implies that the difference between the point estimator and the true value of the parameter $(\bar{x} - \mu)$ will be less than 1.96 standard deviations or 1.96 standard errors (SE) in length for 95% of sample means, and this quantity $1.96 \cdot SE$, called the margin of error (E), provides a practical upper bound for the error of estimation (see the figures below). It is possible that the error of estimation $(\bar{x} - \mu)$ will exceed this margin of error (E), but that is very unlikely since for the 95% confidence interval that only happens for 5% of sample means drawn from the population.





How Can I Estimate a 95% Confidence Interval for the Population Mean?

To estimate the population mean μ for a quantitative population, the point estimator $\bar{x}$ has standard error given as

$$\text{SE} = \frac{\sigma}{\sqrt{n}}$$

The margin of error is then calculated as

$$\text{E} = \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

This margin of error is then added and subtracted to the point estimator, $\bar{x}$, to construct the confidence interval. If no value for σ is given or known, then we use the sample standard deviation, s, in place of σ in the margin of error formula.

Example: A marketing analyst wants to estimate the average amount spent by a dating site customer per year. A random sample of $n = 50$ dating site customers were polled about the amount they spend each year on dating websites. The results of the poll produced a mean amount of \$240 with a standard deviation of \$20. Construct a 95% confidence interval estimate for the population mean amount spent by a dating website customer per year.

Solution: The random variable is the amount spent by a dating site customer per year. The point estimate of μ is $\bar{x} = \$240$. The margin of error is

$$1.96 \cdot SE = 1.96 \cdot \sigma_{\bar{x}} = 1.96 \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{\sigma}{\sqrt{50}}$$

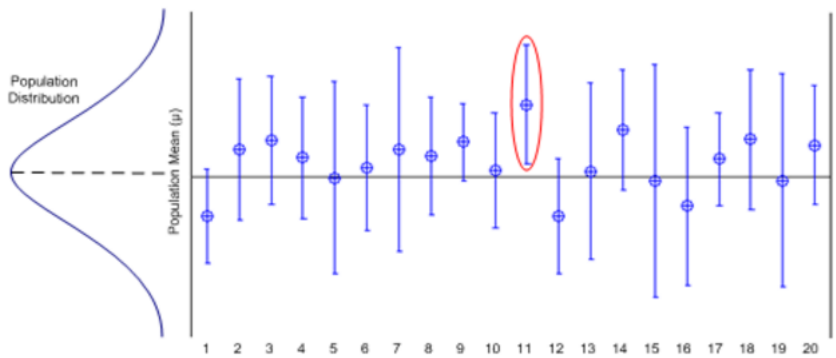
Since the sample size is large (greater than or equal to 30) the analyst can approximate the value of σ with s. Therefore, the margin of error is approximately

$$1.96 \cdot \frac{s}{\sqrt{n}} = 1.96 \cdot \frac{20}{\sqrt{50}} \doteq \$5.54$$

Adding and subtracting E from $\bar{x}$ gives the two numbers: \$240-\$5.54=\$334.46 and \$240+\$5.54=\$245.54. The 95% confidence interval estimate for the population mean is the interval (\$234.46, \$245.54).

Interpreting Confidence Intervals

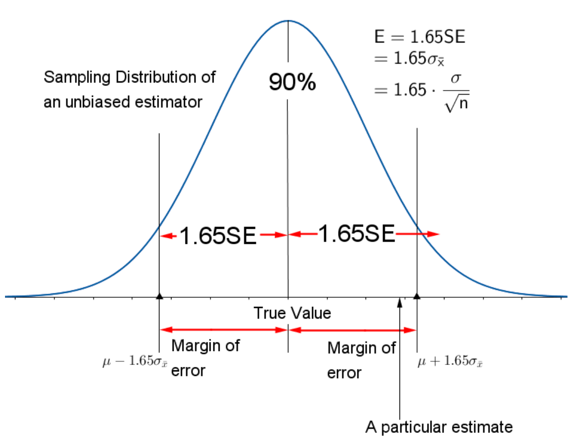
What does mean it mean to say you are "95% confident" that the true value of the population mean μ is within a given interval? If you were to construct 20 such intervals, each using different sample information, your intervals might look like those shown in the figure below. Of the 20 intervals, you might expect that 95% of them, 19 out of 20, will perform as planned and contain μ within their upper and lower bounds. Remember that you cannot be absolutely sure that any one particular interval contains the mean μ. You will never know whether your particular interval is the one out of the 19 that "worked," or whether it is the one interval that "missed." Your confidence in the estimated interval follows from the fact that when repeated intervals are calculated, 95% of these intervals will contain μ.

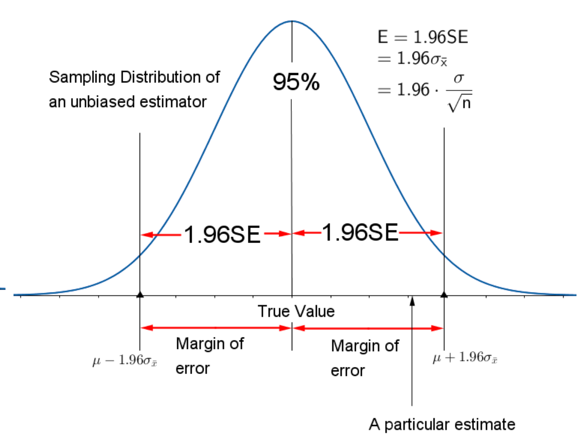


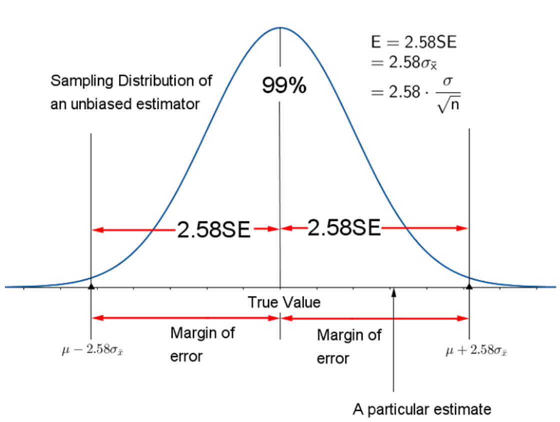
A good confidence interval has two desirable characteristics:

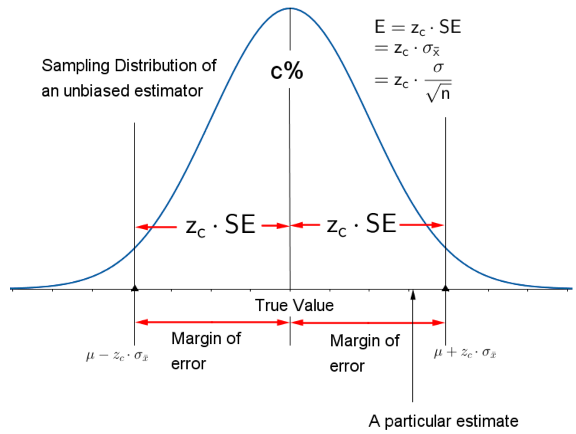
- It is as narrow as possible. The narrower the interval, the more exactly you have located the estimated parameter.
- It has a large level of confidence, near 100%. The larger the confidence level, the more likely it is that the interval will contain the estimated parameter.

The sampling distribution for the three confidence levels: 90%, 95% and 99% are shown below, along with their corresponding margin of errors. Each time we change the confidence level, c, to a different number, we get a different length for the margin of error.





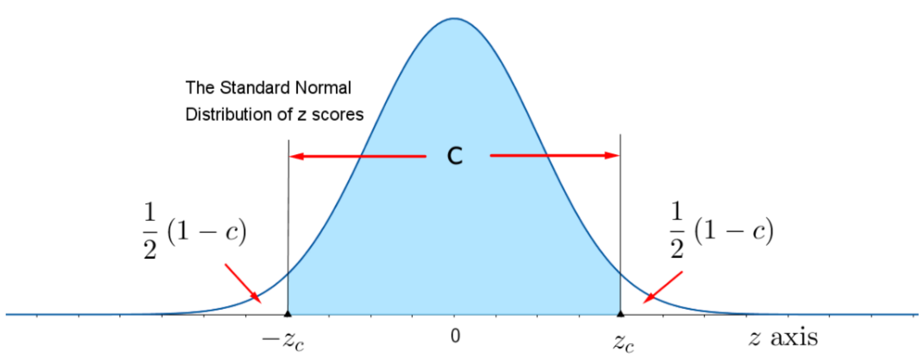




For the 90% confidence interval, $E = 1.65SE$, for the 95% confidence interval, $E = 1.96SE$, for the 99% confidence interval, $E = 2.58SE$, and, in general, for the c percent confidence interval, $E = z_c \cdot SE$.

Definition **The critical value** z_c is the z-score associated with the boundary value, $\mu + z_c \cdot \sigma_{\bar{x}}$, located along the $\bar{x}$ axis of the sampling distribution graph. We need to remember that $z_c = invnorm\left(\frac{1}{2}(1+c)\right)$.

You may want to change the level of confidence from $c = 0.95$ to an other confidence level, c. When you change c to something other than 95%, a value different from $z = 1.96$ will need to be used to find the Margin of Error. You will need to change the value of $z = 1.96$ — which locates an area 0.95 in the center of the standard normal curve — to a value of z that locates the area c in the center of the curve, as shown in the figure below. Since the total area under the curve is 1, the remaining area in the two tails is $1 - c$, and each tail contains area $\frac{1}{2}(1 - c)$. Then c is the percent of the area under the normal curve between $-z_c$ and z_c.



Notice in the figure above that the area under the standard normal distribution, left of a vertical line at z_c is

$$\begin{aligned}\frac{1}{2}(1-c)+c &= \frac{1}{2}-\frac{1}{2}\cdot c+1\cdot c\\ &= \frac{1}{2}+\left(-\frac{1}{2}c\right)+1c\\ &= \frac{1}{2}+\left(-\frac{1}{2}+1\right)\cdot c\\ &= \frac{1}{2}+\frac{1}{2}\cdot c\\ &= \frac{1}{2}(1+c)\end{aligned}$$

Use the "invnorm" command on the calculator with this value to find the critical value.

$$z_c = invnorm\left(\frac{1}{2}(1+c)\right)$$

Some of the TI calculators prompt you to enter values for μ and σ when using the invnorm command. If that is the case, make sure you are using the values of μ and σ that correspond to the standard normal distribution of z scores. That is, use $\mu = 0$ and $\sigma = 1$.

The value of z that has "tail area" $\frac{1}{2}(1-c)$ to its right is called z_c and the area between $-z_c$ and z_c is the level of confidence c. Values of z_c that are typically used by experimenters will become familiar to you as you begin to construct confidence intervals for different practical situations. Some of these values are given in the table below.

Confidence Level c	$1-c$	z_c
0.90	0.10	1.645
0.95	0.05	1.96
0.98	0.02	2.33
0.99	0.01	2.58

The width of a confidence interval depends on the size of the margin of error, $z_c \cdot \sigma_{\bar{x}}$, which depends on the values of z, σ, and n because $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. However, the value of σ is not under the control of the investigator. Hence, the width of a confidence interval can be controlled using

1. The value of z, which depends on the confidence level
2. The sample size n

The confidence level determines the value of z, which in turn determines the size of the margin of error. The value of z increases as the confidence level increases, and it decreases as the confidence level decreases. For example, the value of z is approximately 1.65 for a 90% confidence level, 1.96 for a 95% confidence level, and approximately 2.58 for a 99% confidence level. Hence, the higher the confidence level, the larger the width of the confidence interval, other things remaining the same.

For the same value of σ, an increase in the sample size decreases the value of $\sigma_{\bar{x}}$, which in turn decreases the size of the margin of error when the confidence level remains unchanged. Therefore, an increase in the sample size decreases the width of the confidence interval. Thus, if we want to decrease the width of a confidence interval, we have two choices:

1. Lower the confidence level
2. Increase the sample size

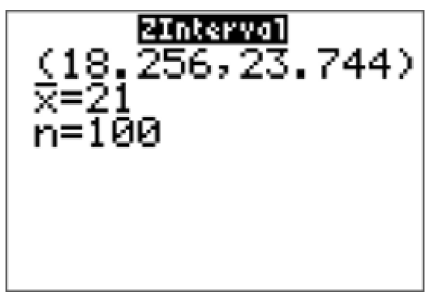
Lowering the confidence level is not a good choice, however, because a lower confidence level may give less reliable results. Therefore, we should always prefer to increase the sample size if we want to decrease the width of a confidence interval.

Exercises

1. Find the critical value z_c that must be used in the margin of error formula for a 92% confidence interval estimate for the population mean.

2. Find a 92% confidence interval estimate for μ assuming sample size n was 200. Suppose σ was known to be 33.3 from past history and that the sample mean was found to be 250.5.

3. The calculator screen below displays the results from counting the number of different types of donuts in a sample of 100. Use the given confidence interval to find the point estimate $\bar{x}$ and the margin of error E.



4. A researcher wants to estimate the average amount of time, in hours per day, that a U.S. teenager spends consuming media — watching TV, listening to music, surfing the Web, social networking, and playing video games. A random sample of $n = 1000$ U.S. teenagers were polled about the amount of time they spend daily consuming media. The results of the poll produced a mean amount of 7.6 hours. Assume the standard deviation of the population is 2.2 hours. Use this information to estimate the population mean. Use a 90% level of confidence.

   A. Press STAT and move the cursor to TESTS.
   B. Press 7 for ZInterval.
   C. Move the cursor to Stats and press ENTER.
   D. Type in the appropriate values.
   E. Move the cursor to Calculate and press ENTER.

5. Find the 95% confidence interval for the mean using this sample.
   45 52 35 22 62 34 42 46 53 58 36 40 43 16 23
   54 27 32 24 53 62 67 84 36 44 49 57 35 25 30

   A. Enter the data into L1. (Press STAT → ENTER to access L1)
   B. Press STAT and move the cursor to TESTS.
   C. Press 7 for ZInterval.
   D. Move the cursor to Data and press ENTER.
   E. Type in the appropriate values.
   F. Move the cursor to Calculate and press ENTER

Sometimes, researchers first decide how large they want the margin of error to be (and in turn how wide the confidence interval will be), then they determine how large of a sample size they need to take to build their confidence interval for μ with the desired margin of error length.

Formula to Find a Minimum Sample Size
Given a c-confidence level and a margin of error E, the minimum sample size n needed to estimate the population mean μ is

$$n = \left(\frac{z_c \cdot \sigma}{E} \right)^2$$

Always round the value up to the next whole number. When σ is unknown, you can estimate n using using sample standard deviation, s, provided you have a preliminary sample with at least 30 members.

6. A pizza shop owner wishes to find the 95% confidence interval estimate for the true mean cost of a large pepperoni pizza. How large should the sample be if she wishes to be accurate within \$0.15? A previous study showed that the standard deviation of the price was \$0.26.

7. A beverage company uses a machine to fill one-liter bottles with water. Assume the population of volumes is normally distributed. The company wants to estimate the mean volume of water the machine is putting in the bottles within one milliliter (ml). Determine the minimum sample size required to construct a 96% confidence interval estimate for μ. Assume the population standard deviation is 3 milliliters.

6.2 Confidence Intervals for the Mean (σ unknown)

Section Learning objectives:

1. How to interpret the t-distribution and use a t-distribution table
2. How to construct and interpret confidence intervals for a population mean when σ is not known

---

When the population standard deviation is unknown, the sample size is less than 30, and the random variable x is approximately normally distributed, then the z-scores of sample means taken from the sampling distribution of $\bar{x}$ follow a t-distribution, not the standard normal distribution.

This fact was discovered by an Englishman named W. S. Gosset in 1908. He derived a complicated formula for the density function of t-values.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

for random samples of size n from a normal population, and he published his results under the pen name "Student." Ever since, the statistic has been known as Student's t. A t-distribution curve has the following characteristics:

- The mean, median and mode are all equal to zero.
- The total area under the t-distribution curve is 1.
- It is mound-shaped and symmetric about t=0, just like the Standard Normal distribution of z.
- It is more variable than a z distribution with "heavier tails"; that is, the t curve does not approach the horizontal axis as quickly as z does. This is because the t statistic involves two random quantities, $\bar{x}$ and s, whereas the z statistic involves only the sample mean, $\bar{x}$.
- The shape of the t distribution depends on the sample size n. As n increases, the variability of t decreases because the estimates of s of σ is based on more and more information. Eventually, when n is infinitely large, the t and z distributions are identical!

The divisor $(n-1)$ in the formula for the sample variance s^2 is called the number of degrees of freedom (df) associated with s^2. It determines the shape of the t distribution. The origin of the term degrees of freedom is theoretical and refers to the number of independent squared deviations in s^2 that are available for estimating σ^2. These degrees of freedom may change for different applications, and since they specify the correct t distribution to use, you need to remember to calculate the correct degrees of freedom for each application.

The table of normal probabilities for the standard normal z distribution is no longer useful in calculating critical values for the margin of error in your confidence interval formula. Instead, you will use the t-table (below). The table body lists critical values of t_c. The first column of the table is a particular number of degrees of freedom. The top row has a percentage area to the left of a vertical line at t_c.

Exercises:

1. For a t distribution with 10 degrees of freedom, the value of t that has an area 0.90 to its left is found in row 10 in the column marked "90%." You should verify that this is $t = 1.372$
2. Find the t value that represents the t-score in the 50th percentile.
3. Find the t value that represents the first quartile of t-scores. Assume $n = 8$.
4. Suppose you have a sample of size 15 from a normal distribution. Find a value of t such that only 2.5% of all values of t will be smaller.
5. Find the t value that represents the t-score in the 20th percentile for sample of size $n = 6$
6. Find the value of t_c needed to set up a 90% percent confidence interval estimate for μ. Assume the sample size is 23.

Formula for a c-percent Confidence Interval Estimate of a Population Mean μ (σ unknown)

$$\overline{X} - E \ < \mu < \ \overline{X} + E$$

where margin of error, E, is

$$E = t_c \cdot \frac{s}{\sqrt{n}}$$

with

$$\begin{aligned} n &= \text{sample size} \\ s &= \text{standard deviation of the sampled population} \\ c &= \text{the confidence level} \\ df &= n - 1 \end{aligned}$$

and t_c is the t value on the $(n-1)^{st}$ row of t-table that has an area of $\frac{1}{2}(1-c)+c = \frac{1}{2}(1+c)$, left of a vertical line at t_c.

For the TI84+ calculator, $t_c = invT\left(area,\ df\right)$, where

$$\begin{aligned} area &= \frac{1}{2}(1+c) \\ df &= n - 1 \end{aligned}$$

Sadly, there is no invT function on the TI83/83+, and some TI84+ don't have the invT command. When the sample size is large ($n \geq 30$) the critical values on the t-table approach the same critical values of z on the standard normal distribution table.

Assumptions:

1. The sample must be a random sample.
2. The population from which you are sampling must be normally distributed.
3. The value of the population standard deviation,σ, is unknown or not given along with the statement of the problem.

7. Find the value of t_c needed to set up a 95% percent confidence interval estimate for μ. Assume the sample size is 11.

8. Suppose that for a random sample of 5 computers at a certain electronics store, the mean repair cost was \$178. The sample standard deviation was \$32. Assume the population is normally distributed. Construct a 99% confidence interval estimate for the population mean repair cost.

   A. Press STAT and move the cursor to TESTS.
   B. Press 8 for TInterval.
   C. Move the cursor to Stats and press ENTER.
   D. Type in the appropriate values.
   E. Move the cursor to Calculate and press ENTER.

9. Find the 95% confidence interval for the mean using this sample.

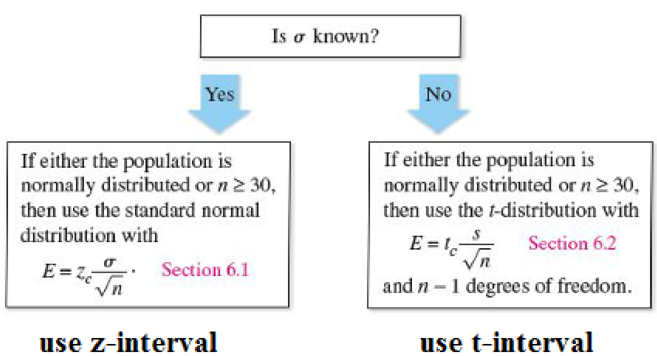
   625 675 535 406 512 680 483 522 619 575

   A. Enter the data into L1. (Press STAT $\rightarrow$ ENTER to access L1)
   B. Press STAT and move the cursor to TESTS.
   C. Press 8 for TInterval.
   D. Move the cursor to Data and press ENTER.
   F. Type in the appropriate values.
   E. Move the cursor to Calculate and press ENTER

Comparing the t and z Distributions

Look at one of the columns in the t-table. As the degrees of freedom increase, the critical value of t decreases until, when $df = \infty$; the critical t-value is the same as the critical z-value for the same tail area. For example, when the area right of t is 95minimum of $t_c = z_c = 1.645$. This helps to explain why we use $n = 30$ as the somewhat arbitrary dividing line between large and small samples. When $n = 30$ $(df = 29)$, the critical values of t are quite close to their normal counterparts. Notice that $t_c = 1.699$ is quite close to $z_c = 1.645$. Rather than produce a t table with rows for many more degrees of freedom, the critical values of z are sufficient when the sample size reaches $n = 30$.

When do I use t-interval and when do I use z-interval?



6.3 Confidence Intervals for a Population Proportion p

Section Learning objectives:

1. How to find a point estimate for the population proportion
2. How to construct and interpret confidence intervals for a population proportion
3. How to determine the minimum sample size required when estimating a population proportion

---

Many research experiments or sample surveys have as their objective the estimation of the proportion of people or objects in a large group that possess a certain, characteristic. Here are some examples:

- The proportion of Americans who have internet access
- The proportion of those who believe there is solid evidence that Earth is getting warmer
- The proportion of residents who closely follow the local news or who often discuss local crime

Each is a practical example of the binomial experiment, and the parameter to be estimated is the binomial proportion, $\hat{p}$. When the sample size is large,

$$\hat{p} = \frac{x}{n} = \frac{\text{total number of successes}}{\text{total number of trials}}$$

is the best point estimator for the population proportion p. Since its sampling distribution is approximately normal, with mean p and standard error $\text{SE} = \sqrt{\frac{pq}{n}}$, the sample percentage $\hat{p}$ can be used to construct a confidence interval according to the general approach given here.

Formula for a c-percent Confidence Interval Estimate of a Population Proportion, p.

$$\hat{p} - E < p < \hat{p} + E$$

where

$$E = z_c \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

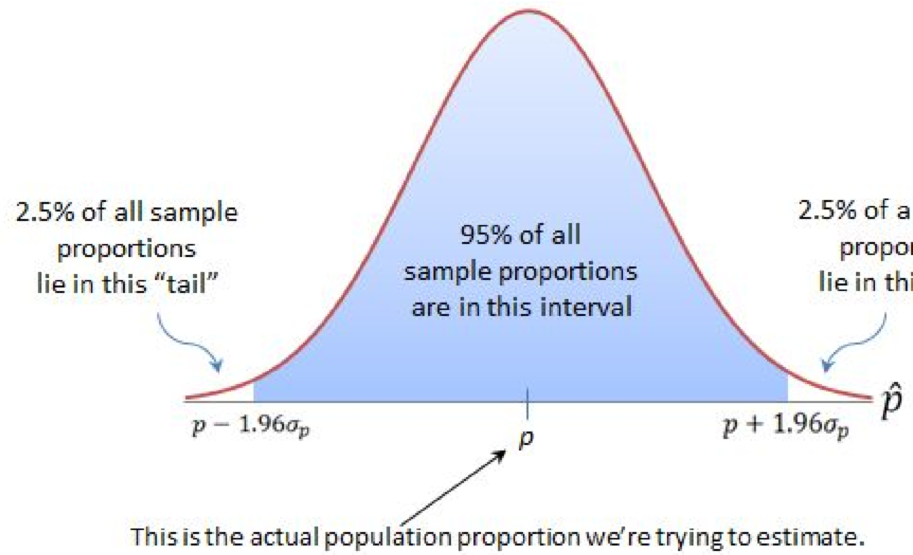
and

$$z_c = invnorm\left(\frac{1}{2}(1+c)\right)$$

z_c is the z value corresponding to an area $\frac{1}{2}(1-c)$ in the right tail of a standard normal z distribution. Since p and q are unknown, their values are estimated using the best point estimators: $\hat{p}$ and $\hat{q}$ (where $\hat{q} = 1 - \hat{p}$). The sample size is considered large when the normal approximation to the binomial distribution is adequate — when both

$$np > 5 \quad \text{and} \quad nq > 5$$

(but we don't have values for p and q, so we use $\hat{p}$ and $\hat{q}$ and check that both $n\hat{p} > 5$ and $n\hat{q} > 5$).



Exercises:

1. A reporter wants to estimate the percentage of U.S. households that have internet. A random sample of 4000 U.S. households found 2880 households that had internet. Construct a 98% confidence interval estimate of the percentage of U.S. households that have internet.

   Calculator Solution:
   A. Press the STAT button on your calculator and move the cursor to TESTS.
   B. Press A (ALPHA, MATH) for 1-PropZlnt.
   C. Type in the appropriate values.
   D. Move the cursor to Calculate and press ENTER

2. A reporter wants to estimate the percentage of residents in her city who would rate their city as an excellent place to live. A random sample of 120 individuals is drawn and suppose 40% indicated they would rate their city as an excellent place to live. Use this information to estimate the population proportion of city residents who would rate their city as an excellent place to live. Use a 99% level of confidence. (Source: journalism.org)

3. There were 200 nursing applications in a sample, and 12% of the applicants were male. Find the 99% confidence interval for the true proportion of male applicants.

Sample Size Determination

Find a Minimum Sample Size to Estimate a Binomial Population Proportion, p
Given a c-confidence level and a margin of error E, the minimum sample size n needed to estimate the population proportion, p is

$$n = \hat{p} \cdot \hat{q} \cdot \left(\frac{z_c}{E} \right)^2$$

where $\hat{q} = 1 - \hat{p}$ Always round the value up to the next whole number. This formula assumes you have preliminary or previous estimates of $\hat{p}$ and $\hat{q}$. If not, use $\hat{p} = 0.50$ and $\hat{q} = 0.50$ since the formula outputs its maximum value if we assume $\hat{p} = 0.50$ and $\hat{q} = 0.50$. Also note that for this formula the margin of error is a percentage (as a decimal), a unitless quantity.

4. You wish to estimate, with 90% confidence, the population proportion of U.S. adults who are confident in the stability of the U.S. banking system. Your estimate must be accurate to within 3% of the population proportion.
   (a) No preliminary or previous estimate is available. Find the minimum sample size needed.
   (b) Find the minimum sample size needed, using a prior study that found 43% of U.S. adults are confident in the stability of the system.
   (c) Compare the results from (a) and (b).

5. Determine the minimum sample size required when you want to be 95% confident that the sample proportion is within two percentage points of the population proportion. Assume the population is normally distributed.

6. A recent report by Pew Research Center estimated that 68% of Americans have smartphones and 45% have tablet computers. How large a sample is needed to estimate the true proportion of Americans with smartphones to within 4% with 86% confidence?

6.4 WEB LINK — Confidence Intervals for a Population Variance or Standard Deviation

http://timbusken.com/estimation.html#CIvar

6.5 WEB LINK — More about Confidence Intervals

7 Hypothesis Tests with One Sample

7.1 Introduction to Hypothesis Testing

Section Learning objectives:

1. How to state a null hypothesis and an alternative hypothesis
2. How to identify type I and type II errors and interpret the level of significance
3. How to know whether to use a one-tailed or two-tailed statistical test and find a p-value
4. How to make and interpret a decision based on the results of a statistical test
5. How to write a claim for a hypothesis test

Classroom Exercises

Stating Hypotheses **In exercises 1–4, the statement represents a claim. Write its complement and state which is H_0 and which is H_a.**

1. $\mu \leq 55$

2. $\sigma \neq 1.3$

3. $p < 0.45$

4. $\mu > 158$

Stating the Hypotheses **In Exercises 5–8, write the claim as a mathematical statement. State the null and alternative hypotheses, and identify which represents the claim.**

5. A laptop manufacturer claims that the mean life of the battery for a certain model of laptop is more than 6 hours.

6. As stated by a company's shipping department, the number of shipping errors per million shipments has a standard deviation that is less than 3.

7. An amusement park claims that the mean daily attendance at the park is at least 20,000 people.

8. According to a recent survey, 39% of college students own a credit card. (Source: Sallie Mae)

Interpreting a Decision **In Exercises 9-12, determine whether the claim represents the null hypothesis or the alternative hypothesis. If a hypothesis test is performed, how should you interpret a decision that**
(a) rejects the null hypothesis?
(b) fails to reject the null hypothesis?

9. A scientist claims that the mean incubation period for swan eggs is less than 40 days.

10. A sports drink maker claims that the mean calorie content of its beverages is 72 calories per serving

11. A government agency claims that more than 75% of full-time workers earn over $538 per week.

12. A researcher claims that the standard deviation of the life of a certain type of lawn mower is at most 2.8 years.

Identifying Tests **In Exercises 13–15, determine whether the hypothesis test is left-tailed, right-tailed, or two-tailed.**

13. $$H_0: \quad \mu \leq 5.6$$
$$H_a: \quad \mu > 5.6$$

14. $$H_0: \quad \sigma^2 = 125$$
$$H_a: \quad \sigma^2 \neq 125$$

15. $$H_0: \quad \sigma \geq 5.2$$
$$H_a: \quad \sigma < 5.2$$

Identifying Tests **In Exercises 16–17, state H_0 and H_a in words and in symbols. Then determine whether the hypothesis test is left-tailed, right-tailed, or two-tailed. Explain your reasoning.**

16. A manufacturer of grandfather clocks claims that the mean time its clocks lose is no more than 0.02 second per day.

17. A state claims that the mean tuition of its universities is no more than $25,000 per year.

Identifying Errors **In Exercises 18–19, describe type I and type II errors for a hypothesis test of the indicated claim.**

18. A garden hose manufacturer advertises that the mean flow rate of a certain type of hose is 16 gallons per minute.

19. A computer repairer advertises that the mean cost of removing a virus infection is less than $100.

7.2 Hypothesis Testing for the Mean (σ known)

Section Learning objectives:

1. How to find and interpret p-values
2. How to use p-values for a z-test for a μ when σ is known

For each hypothesis test question I give you, I will ask you to answer questions a through i listed below. I recommend you use the hypothesis testing handout in the appendix of this workbook

(a) Which test is appropriate here?

(b) Write the symbolic form of the claim.

(c) Write the null and alternative hypotheses.

(d) What number is the test statistic equal to?

(e) What formula should be used for the the standardized test statistic?

(f) What number is the standardized test statistic equal to? Round to the hundredths.

(g) What p-value do you obtain? Round to the ten-thousandths.

(h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

Classroom Exercises

1. **Sprinkler Systems** A manufacturer of sprinkler systems designed for fire protection claims that the average activating temperature is at least 135°F. To test this claim, you randomly select a sample of 32 systems and find the mean activation temperature to be 133°F. Assume the population standard deviation is 3.3°F. At $\alpha = 0.10$, do you have enough evidence to reject the manufacturer's claim?

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) State a full sentence conclusion stating the decision you made.

2. **Fast Food** A fast food restaurant estimates that the mean sodium content in one of its breakfast sandwiches is no more than 920 milligrams. A random sample of 44 breakfast sandwiches has a mean sodium content of 925 milligrams. Assume the population standard deviation is 18 milligrams. At $\alpha = 0.10$, do you have enough evidence to reject the restaurant's claim?

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) State a full sentence conclusion stating the decision you made.

3. **MCAT Scores** A random sample of 50 medical school applicants at a university has a mean raw score of 31 on the multiple choice portions of the Medical College Admission Test (MCAT). A student says that the mean raw score for the school's applicants is more than 30. Assume the population standard deviation is 2.5. At $\alpha = 0.01$, is there enough evidence to support the student's claim?

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) State a full sentence conclusion stating the decision you made.

4. **Cheddar Cheese Consumption** A consumer group claims that the mean annual consumption of cheddar cheese by a person in the United States is at most 10.3 pounds. A random sample of 100 people in the United States has a mean annual cheddar cheese consumption of 9.9 pounds. Assume the population standard deviation is 2.1 pounds. At $\alpha = 0.05$, can you reject the claim?

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) State a full sentence conclusion stating the decision you made.

5. **Caffeine Content in Colas** A company that makes cola drinks states that the mean caffeine content per 12-ounce bottle of cola is 40 milligrams. You want to test this claim. During your tests, you find that a random sample of twenty 12-ounce bottles of cola has a mean caffeine content of 39.2 milligrams. Assume the population is normally distributed and the population standard deviation is 7.5 milligrams. At $\alpha = 0.01$, can you reject the company's claim?

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) State a full sentence conclusion stating the decision you made.

7.3 Hypothesis Testing for the Mean (σ unknown)

Section Learning objectives:

1. How to find critical values in a t-distribution
2. How to use the t-test to test a μ when σ is not known

Classroom Exercises

1. **Credit Card Balances** A credit card company claims that the mean credit card debt for individuals is greater than \$5000. You want to test this claim. You find that a random sample of 37 cardholders has a mean credit card balance of \$5122 and a standard deviation of \$625. At $\alpha = 0.05$, can you support the claim?
   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

2. **Waste Generated** As part of your work for an environmental awareness group, you want to test a claim that the mean amount of waste generated by adults in the United States is less than 5 pounds per day. In a random sample of 19 adults in the United States, you find that the mean waste generated per person per day is 4.43 pounds with a standard deviation of 1.21 pounds. At $\alpha = 0.01$, can you support the claim?

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) State a full sentence conclusion stating the decision you made.

Class sizes					
35	28	29	33	32	40
26	25	29	28	30	36
33	29	27	30	28	25

3. **Faculty Classroom Hours** Class Size You receive a brochure from a large university. The brochure indicates that the mean class size for full-time faculty is fewer than 32 students. You want to test this claim. You randomly select 18 classes taught by full-time faculty and determine the class size of each. The results are shown in the table at the left. At $\alpha = 0.05$, can you support the university's claim?
   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) State a full sentence conclusion stating the decision you made.

7.4 Hypothesis Testing for Proportions

Section Learning objectives:

1. How to use the z-test to test a population proportion p

In Sections 7.2 and 7.3, you learned how to perform a hypothesis test for a population mean μ. In this section, you will learn how to test a population proportion (percentage), p. Hypothesis tests for proportions (percentages) can be used to test claims about the percentage of a population has a certain characteristic. Hypothesis tests for proportions (percentages) can be used whenever you want to test a person or organization's claim about a population percentage. Hypothesis tests for proportions (percentages) can be used of when politicians want to know the proportion of their constituents who favor a certain bill or when quality assurance engineers test the proportion of parts that are defective. If $np \geq 5$ and $nq \geq 5$ for a binomial distribution, then the sampling distribution for np is approximately normal with a mean of $\mu_p = p$. and a standard error of $\sigma_p = \sqrt{pq/n}$

Classroom Exercises

1. Surveillance Cameras A research center claims that 63% of U.S. adults support using surveillance cameras in public places. In a random sample of 300 U.S. adults, 70% say that they support using surveillance cameras in public places. At $\alpha = 0.05$, is there enough evidence to reject the research center's claim?
   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

(g) What p-value do you obtain? Round to the ten-thousandths.

(h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

2. **Smokers** A medical researcher says that less than 25% of U.S. adults are smokers. In a random sample of 200 U.S. adults, 19% say that they are smokers. At $\alpha = 0.05$, is there enough evidence to support the researcher's claim?

(a) Which test is appropriate here?

(b) Write the symbolic form of the claim.

(c) Write the null and alternative hypotheses.

(d) What number is the test statistic equal to?

(e) What formula should be used for the standardized test statistic?

(f) What number is the standardized test statistic equal to? Round to the hundredths.

(g) What p-value do you obtain? Round to the ten-thousandths.

(h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

3. **Internal Revenue Service** A research center claims that at least 46% of U.S. adults think that the IRS is not aggressive enough in pursuing people who cheat on their taxes. In a random sample of 600 U.S. adults, 246 say that the IRS is not aggressive enough in pursuing people who cheat on their taxes. At $\alpha = 0.01$, is there enough evidence to reject the center's claim?

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) State a full sentence conclusion stating the decision you made.

4. **Hands-Free Cell Phones** A research center claims that at most 75% of U.S. adults think that drivers are safer using hands-free cell phones instead of using hand-held cell phones. In a random sample of 150 U.S. adults, 77% think that drivers are safer using hands-free cell phones instead of hand-held cell phones. At $\alpha = 0.01$, is there enough evidence to reject the center's claim?

(a) Which test is appropriate here?

(b) Write the symbolic form of the claim.

(c) Write the null and alternative hypotheses.

(d) What number is the test statistic equal to?

(e) What formula should be used for the standardized test statistic?

(f) What number is the standardized test statistic equal to? Round to the hundredths.

(g) What p-value do you obtain? Round to the ten-thousandths.

(h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

7.5 WEB LINK — Hypothesis Tests for a Population Variance or Standard Deviation

http://timbusken.com/hypothesis-testing-standard-deviation.html

7.6 WEB LINK — More Hypothesis Testing Examples

Link to more hypothesis testing examples
http://timbusken.com/hypothesis_testing.html

8 Hypothesis Tests with Two Samples

8.1 Testing the Difference Between Means (Independent Samples, σ_1 and σ_2 known)

Section Learning objectives:

1. How to determine whether two samples are independent or dependent
2. An introduction to two-sample hypothesis testing for the difference between two population parameters
3. How to perform a two-sample 2-SampZTest for the difference between two means μ_1 and μ_2 using independent samples with σ_1 and σ_2 known

---

Definition: Two samples are **independent** when the sample selected from one population is not related to the sample selected from the second population. Two samples are **dependent** when each member of one sample corresponds to a member of the other sample. Dependent samples are also called **paired samples** or **matched samples**.

Dependent samples often involve before and after results for the same person or object (such as a person's weight before starting a diet and after 6 weeks), or results of individuals matched for specific characteristics (such as identical twins).

Dependent samples often come from applying a treatment to persons or objects. In this scenario, samples are gathered before and after the treatment is applied.

Classify each pair of samples as independent or dependent and justify your answer.

1. Sample 1: Weights of 65 college students before their freshman year begins

   Sample 2: Weights of the same 65 college students after their freshman year

2. Sample 1: Scores for 38 adult males on a psychological screening test for attention-deficit hyperactivity disorder

   Sample 2: Scores for 50 adult females on a psychological screening test for attention-deficit hyperactivity disorder

In this section, you will learn how to test a claim comparing the means of two different populations using independent samples.

It is important to remember that when you perform a two-sample hypothesis test using independent samples, you are testing a claim concerning the difference between the parameters in two populations, not the values of the parameters themselves

For a two-sample hypothesis test with independent samples,

1. **the null hypothesis** H_0 is a statistical hypothesis that usually states there is no difference between the parameters of two populations. The null hypothesis always contains the symbol $\leq, =$, or $\geq$

2.**the alternative hypothesis** H_a is a statistical hypothesis that is true when H_0 is false. The alternative hypothesis contains the symbol $>, \neq$ or $<$.

To write the null and alternative hypotheses for a two-sample hypothesis test with independent samples, translate the claim made about the population parameters from a verbal statement to a mathematical statement. Then, write its complementary statement. For instance, for a claim about two population parameters μ_1 and μ_1, some possible pairs of null and alternative hypotheses are

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_a: \mu_1 \neq \mu_2 \end{cases}, \quad \begin{cases} H_0: \mu_1 \leq \mu_2 \\ H_a: \mu_1 > \mu_2 \end{cases}, \quad \text{and} \quad \begin{cases} H_0: \mu_1 \geq \mu_2 \\ H_a: \mu_1 < \mu_2 \end{cases}.$$

Regardless of which hypotheses you use, at the outset of conducting the hypothesis test, you always assume there is no difference between the population means $\mu_1 = \mu_2$. This centers a sampling distribution of $\bar{x}_1 - \bar{x}_2$ at 0. Then you locate the test statistic $\bar{x}_1 - \bar{x}_2$ relative to this center, and measure the distance between 0 and the test statistic with the z score of the test statistic and the p-value of the test statistic. If the z-score is large in absolute value, then the p-value is small—smaller than α and we reject the null hypothesis. If, on the other hand, the z-score is small then the p-value is large, larger than α, and we fail to reject the null hypothesis.

You will now learn how to perform a ***2-SampZTest*** for the difference between two population means μ_1 and μ_1 when the samples are independent. **These conditions are necessary to perform such a test**:
1. The population standard deviations are known.
2. The samples are randomly selected.
3. The samples are independent.
4. The populations are normally distributed or each sample size is at least 30.

Classroom Exercise

1. **Meal-Replacement Diets** To compare the amounts spent in the first three months by clients of two meal-replacement diets, a researcher randomly selects 20 clients of each diet. The mean amount spent for Diet A is \$643. Assume the population standard deviation is \$89. The mean amount spent for Diet B is \$588. Assume the population standard deviation is \$75. At $\alpha = 0.01$, can the researcher support the claim that the mean amount spent in the first three months by clients of Diet A is greater than the mean amount spent in the first three months by clients of Diet B?

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) State a full sentence conclusion stating the decision you made.

8.2 Testing the Difference Between Means (Independent Samples, σ_1 and σ_2 unknown)

Section 8.1, you learned how to test the difference between means when both population standard deviations are known. In many real-life situations, both population standard deviations are not known. In this section, you will learn how to use a t-test to test the difference between two population means μ_1 and μ_2 using independent samples from each population when σ_1 and σ_2 are unknown.

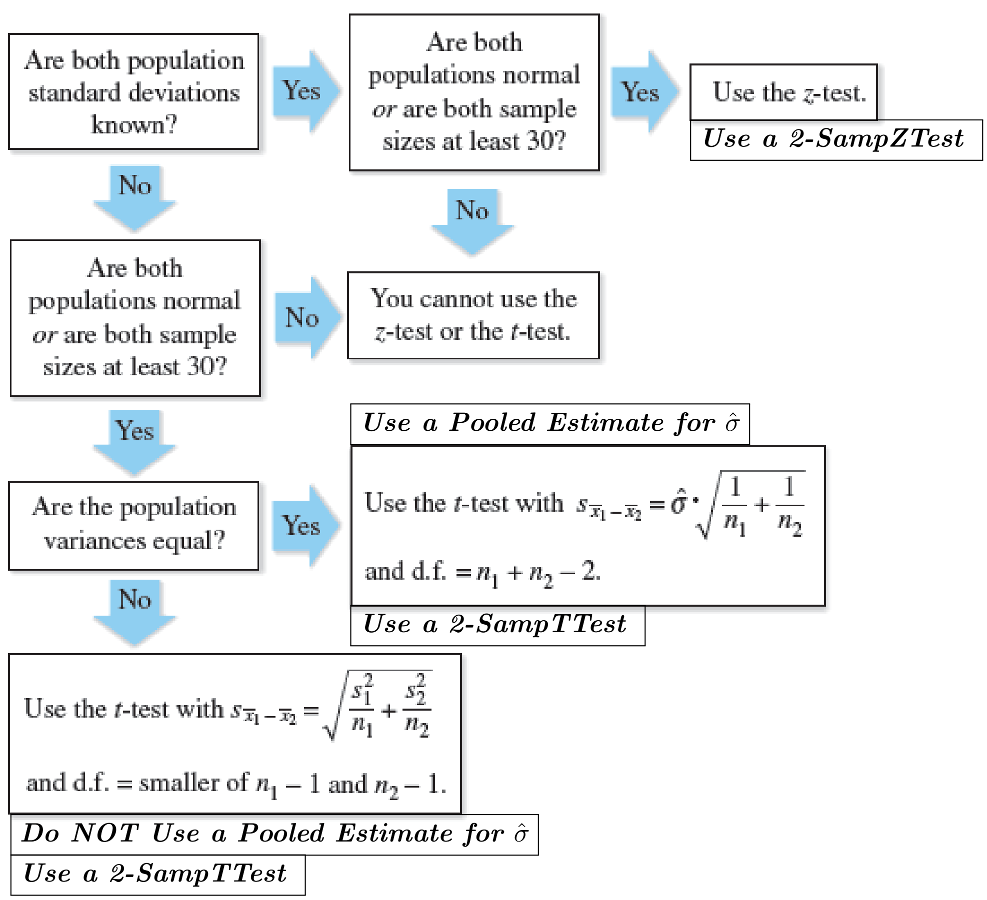
To use a 2-SampTTest, these conditions are necessary.

1. The population standard deviations are unknown.
2. The samples are randomly selected.
3. The samples are independent.
4. The populations are normally distributed or each sample size is at least 30.

When these conditions are met, the sampling distribution for the difference between the sample means $\bar{x}_1 - \bar{x}_2$ is approximated by a t-distribution with mean $\mu_1 - \mu_2$.

The requirements for the z-test described in Section 8.1 and the t-test described in this section are shown in the flowchart below.

Two-Sample Tests for Independent Samples



Sample Statistics for Amount Spent by Customers

Burger Stop	Fry World
$\bar{x}_1 = \$5.46$ $s_1 = \$0.89$ $n_1 = 22$	$\bar{x}_2 = \$5.12$ $s_2 = \$0.79$ $n_2 = 30$

Classroom Exercises

1. **Transactions** A magazine claims that the mean amount spent by a customer at Burger Stop is greater than the mean amount spent by a customer at Fry World. The results for samples of customer transactions for the two fast food restaurants are shown above. At $\alpha = 0.05$, can you support the magazine's claim? Assume the population variances are equal.

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

(h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

2. **Annual Income** A personnel director from Pennsylvania claims that the mean household income is greater in Allegheny County than it is in Erie County. In Allegheny County, a sample of 19 residents has a mean household income of $49,700 and a standard deviation of $8800. In Erie County, a sample of 15 residents has a mean household income of $42,000 and a standard deviation of $5100. At $\alpha = 0.05$, can you support the personnel director's claim? Assume the population variances are not equal.

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

   (e) What formula should be used for the standardized test statistic?

   (f) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) What p-value do you obtain? Round to the ten-thousandths.

   (h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

8.3 Testing the Difference Between Means (Dependent Samples)

In Sections 8.1 and 8.2, you performed two-sample hypothesis tests with *independent samples* using the test statistic $\bar{x}_1 - \bar{x}_2$ (the difference between the means of the two samples). To perform a two-sample hypothesis test with *dependent samples*, you will use a different technique. You will first find the difference d for each data pair.

$$d = (\text{data entry in first sample}) - (\text{corresponding data entry in second sample})$$

The ***test statistic*** is the mean $\bar{d}$ of these differences

$$\bar{d} = \frac{d}{n}$$

That is to say that d bar is the average of the differences between paired data entries in the dependent samples

The ***standardized test statistic*** is

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$$

The degrees of freedom are

$$d.f. = n - 1.$$

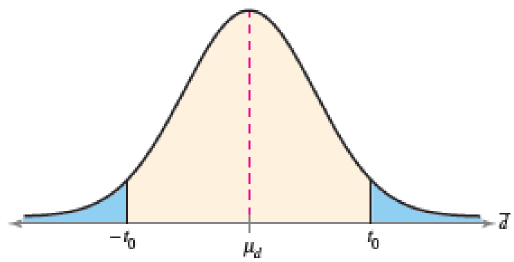
These conditions are necessary to conduct the test.

1. The samples are randomly selected.
2. The samples are dependent (paired).
3. The populations are normally distributed or the number n of pairs of data is at least 30.

When these conditions are met, the sampling distribution for $\bar{d}$, the mean of the differences of the paired data entries in the dependent samples, is approximated by a t-distribution with $n - 1$ degrees of freedom, where n is the number of data pairs.

As a result, we can <u>use the calculator's t-test function</u> to get the p-value and the value of the standardized test statistic.

The symbols listed in the table (below) are used for the t-test for μ_d. Although formulas are given for the mean and standard deviation of differences, you should use technology to calculate these statistics.



Symbol	Description
n	The number of pairs of data
d	The difference between entries in a data pair
μ_d	The hypothesized mean of the differences of paired data in the population
$\overline{d}$	The mean of the differences between the paired data entries in the dependent samples $\overline{d} = \frac{\Sigma d}{n}$
s_d	The standard deviation of the differences between the paired data entries in the dependent samples $s_d = \sqrt{\frac{\Sigma(d - \overline{d})^2}{n - 1}}$

Classroom Exercises

1. A medical researcher wants to determine whether a drug changes the body's temperature. Seven test subjects are randomly selected, and the body temperature (in degrees Fahrenheit) of each is measured. The subjects are then given the drug and, after 20 minutes, the body temperature of each is measured again. The results are listed below. At $\alpha = 0.05$, is there enough evidence to conclude that the drug changes the body's temperature? Assume the body temperatures are normally distributed.

Subject	1	2	3	4	5	6	7
Initial temperature	101.8	98.5	98.1	99.4	98.9	100.2	97.9
Second temperature	99.2	98.4	98.2	99.0	98.6	99.7	97.8

(a) Which test is appropriate here?

(b) Write the symbolic form of the claim.

(c) Write the null and alternative hypotheses.

(d) What number is the test statistic equal to?

(e) What formula should be used for the standardized test statistic?

(f) What number is the standardized test statistic equal to? Round to the hundredths.

(g) What p-value do you obtain? Round to the ten-thousandths.

(h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

2. ***Pneumonia*** A scientist claims that pneumonia causes weight loss in mice. The table below shows the weights (in grams) of six mice before infection and two days after infection. At $\alpha = 0.01$, is there enough evidence to support the scientist's claim?

Mouse	1	2	3	4	5	6
Weight (before)	19.8	20.6	20.3	22.1	23.4	23.6
Weight (after)	18.4	19.6	19.6	20.7	22.2	23.0

(a) Which test is appropriate here?

(b) Write the symbolic form of the claim.

(c) Write the null and alternative hypotheses.

(d) What number is the test statistic equal to?

(e) What formula should be used for the standardized test statistic?

(f) What number is the standardized test statistic equal to? Round to the hundredths.

(g) What p-value do you obtain? Round to the ten-thousandths.

(h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

8.4 Testing the Difference Between Proportions

In this section, you will learn how to use a 2-prop-z-test to test the difference between two population proportions p_1 and p_2 using a sample proportion from each population. If a claim is about two population parameters p_1 and p_2, then some possible pairs of null and alternative hypotheses are

$$\begin{cases} H_0\colon p_1 = p_2 \\ H_a\colon p_1 \neq p_2 \end{cases}, \qquad \begin{cases} H_0\colon p_1 \leq p_2 \\ H_a\colon p_1 > p_2 \end{cases}, \qquad \text{and} \qquad \begin{cases} H_0\colon p_1 \geq p_2 \\ H_a\colon p_1 < p_2 \end{cases}.$$

Regardless of which hypotheses you use, you always assume there is no difference between the population proportions $p_1 = p_2$. This centers the sampling distribution of the difference between sample percentages at zero. We then locate the test statistic, the difference between sample percentages $(\hat{p}_1 - \hat{p}_2)$ relative to the proposed center (of the sampling distribution), and measure this distance in two ways: with the z-score of the difference between sample percentages $(\hat{p}_1 - \hat{p}_2)$, and with the p-value associated with the difference between sample percentages $(\hat{p}_1 - \hat{p}_2)$. If the z-score and p-value are unusual then we reject the null hypothesis. Ordinary z-scores and probability values have us fail to reject the null hypothesis.

A two-sample 1-prop-z-test is used to test the difference between two population proportions p_1 and p_2 when these conditions are met.
1. The samples are random.
2. The samples are independent.
3. The quantities $n_1\bar{p}$, $n_1\bar{q}$, $n_2\bar{p}$, and $n_2\bar{q}$ are at least 5.

The test statistic is $\hat{p}_1 - \hat{p}_2$. The standardized test statistic is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

where

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

and $\bar{q} = 1 - \bar{p}$.

If the null hypothesis states $p_1 = p_2$, $p_1 \leq p_2$, or $p_1 \geq p_2$, then $p_1 = p_2$ is assumed and the expression $p_1 = p_2$ is equal to 0 in the preceding test.

Classroom Exercises

1. In a survey of 175 females ages 16 to 24 who have completed high school during the past 12 months, 72% were enrolled in college. In a survey of 160 males ages 16 to 24 who have completed high school during the past 12 months, 65% were enrolled in college. At $\alpha = 0.01$, can you reject the claim that there is no difference in the percentage of college enrollees between the two groups?

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

(d) What number is the test statistic equal to?

(e) What formula should be used for the standardized test statistic?

(f) What number is the standardized test statistic equal to? Round to the hundredths.

(g) What p-value do you obtain? Round to the ten-thousandths.

(h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

2. **Seat Belt Use** In a survey of 480 drivers from the South, 408 wear a seat belt. In a survey of 360 drivers from the Northeast, 288 wear a seat belt. At $\alpha = 0.05$, can you support the claim that the proportion of drivers who wear seat belts is greater in the South than in the Northeast?

   (a) Which test is appropriate here?

   (b) Write the symbolic form of the claim.

   (c) Write the null and alternative hypotheses.

   (d) What number is the test statistic equal to?

(e) What formula should be used for the standardized test statistic?

(f) What number is the standardized test statistic equal to? Round to the hundredths.

(g) What p-value do you obtain? Round to the ten-thousandths.

(h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) State a full sentence conclusion stating the decision you made.

9 Correlation and Regression

9.1 Correlation

In this section, you will study how to describe what type of relationship, or correlation, exists between two quantitative variables and how to determine whether a linear correlation is significant.

Definition: A **correlation** is a relationship between two variables. The data can be represented by the ordered pairs (x, y), where x is the independent (or explanatory) variable and y is the dependent (or response) variable.

A scatter plot can be used to determine whether a linear (straight line) correlation exists between two variables. The scatter plots below show several types of correlation.

2) For $P(x \leq b)$ or $P(x < b)$

use $normalCdf(-10^9, b, \mu, \sigma)$

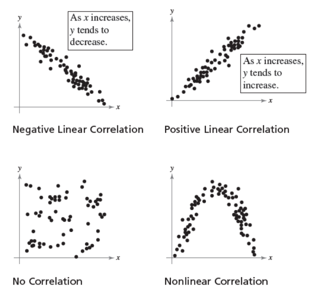


Figure 5: from Larson Statistics, 6E

9.2 Regression

9.3 WEBLINK — Prediction Intervals (Extra Credit Lab)

http://timbusken.com/Correlation.html

10 Chi-Square Tests and the F-Distribution

10.1 WEBLINK — Goodness of Fit

http://timbusken.com/gof.html

10.2 WEBLINK — Independence

http://timbusken.com/independence.html

10.3 Comparing Population Variances (Omitted from the Course Outline)

10.4 WEBLINK — ANOVA

http://timbusken.com/Anova.html

Appendix A --- Hypothesis Testing Handout

The 5 Steps of a Statistical Hypothesis Test

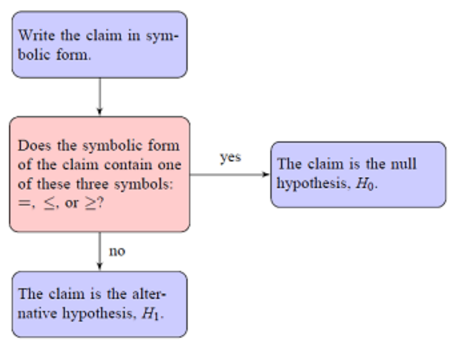
1. **State the null hypothesis, H_0**
2. **State the alternative hypothesis, H_a**
3. **Determine values for the test statistic and the standardized test statistic.**
4. **Find the P-value. Use the P-value to decide whether or not to reject the null hypothesis.**
5. **Write the full sentence conclusion (result) of the hypothesis test.**

Calculator Facts

- **(Section 7.2) If your hypothesis test is about a population mean μ, and the value of population standard deviation σ is known or given, then use your calculator's "z-test" to obtain the p-value**
- **(Section 7.3) If your hypothesis test is about a population mean μ, and the value of population standard deviation σ is unknown or not given, then use your calculator's "t-test" to obtain the p-value**
- **(Section 7.4) If your hypothesis test is about a population proportion (percentage) p, then use your calculator's "1-prop-z-test" to obtain the p-value**

The 5 Steps of a Statistical Hypothesis Test

1. **State the null hypothesis, H_0**
2. **State the alternative hypothesis, H_a**



	Two-Tailed Test	**Left-Tailed Test**	**Right-Tailed Test**
Sign in the null hypothesis, H_0	$=$	$\geq$	$\leq$
Sign in the alternative hypothesis, H_a	$\neq$	<	>

Note that the null hypothesis always has an equal to (=) or a greater than or equal to ($\geq$) or a less than or equal to ($\leq$) sign, and the alternative hypothesis always has a not equal to ($\neq$) or a less than (<) or a greater than (>) sign.

3. **Determine values for the test statistic and the standardized test statistic.**

What is a test statistic?

Definition

The test statistic is the sample mean, sample proportion or sample variance or standard deviation (depending on which one of the parameters, μ, p, σ^2, or σ, you are testing).

What is a standardized test statistic?

Definition

The standardized test statistic is the number of standard deviations that your sample statistic is above or below the hypothesized mean (of the sampling distribution).

The Empirical Rule tells us that 95% of our test statistics (and standardized test statistics) will be within 1.96 standard deviations of the mean. 5% of test statistics are either less than -1.96 or greater than 1.96.

Formulas used for the Standardized Test Statistic

- If the test is about a population mean, μ, with the value of σ given or known, the standardized test statistic is given by the formula $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$, where μ_0 is the value of the mean selected for the null hypothesis.
- If the test is about a population mean, μ, with the value of σ not given or unknown, the standardized test statistic is given by the formula $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
- If the test is about a population proportion, p, the standardized test statistic is given by the formula $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$, where p_0 is the value of the population proportion selected for the null hypothesis and $q_0 = 1 - p_0$.

4. **Find the P-value. Use the P-value to decide whether or not to reject the null hypothesis.**

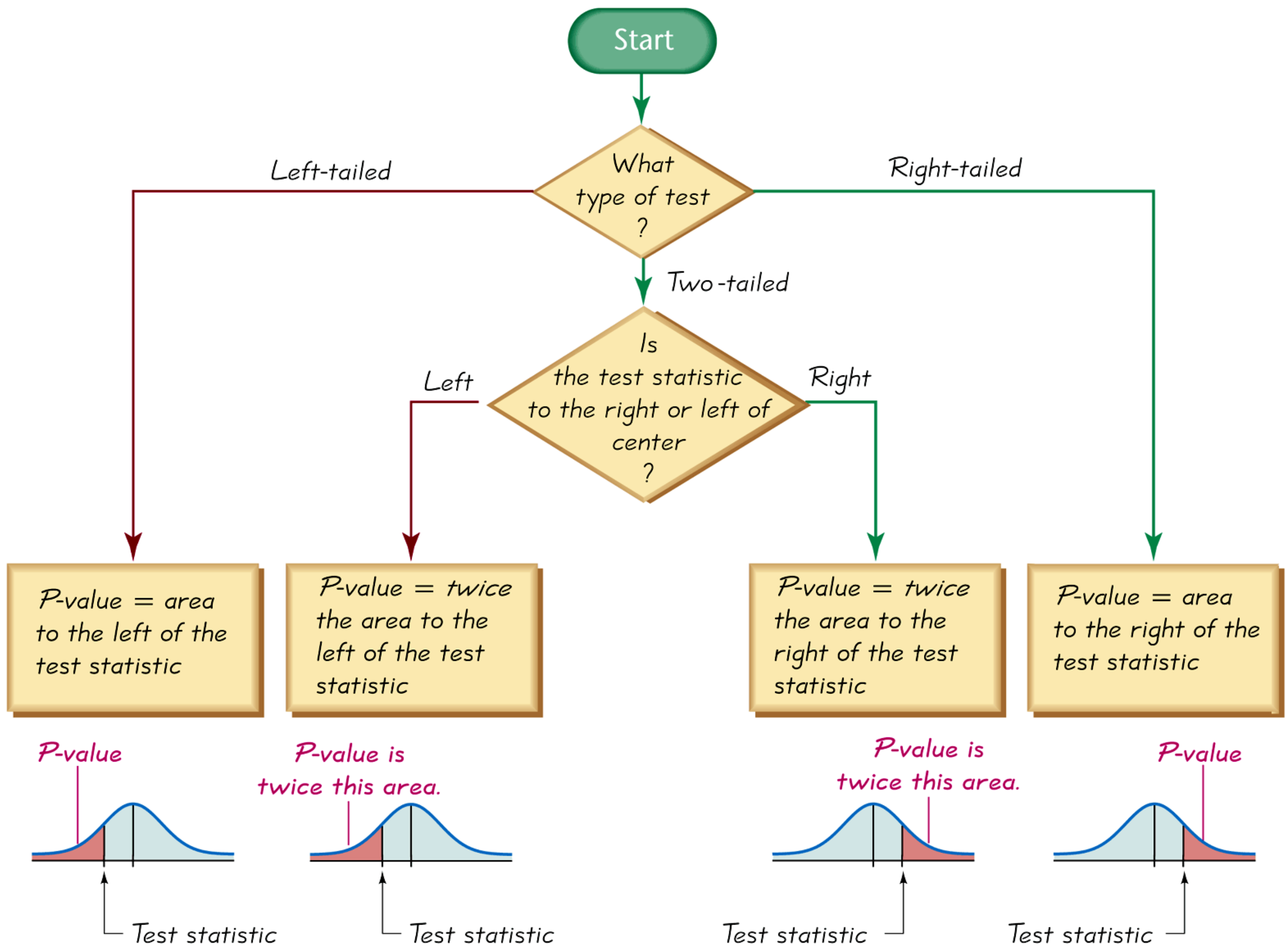
What is a p-value?

Definition

The P–value is the probability of getting a value of the test statistic (and standardized test statistic) that is at least as extreme as the one representing the given sample data, assuming that the null hypothesis is true. One often "rejects the null hypothesis" when the P–value is less than the predetermined significance level (α), indicating that the observed result would be highly unlikely under the null hypothesis.

How can I find the p-value?

Run the appropriate test (z-test, t-test or 1-prop-z-test) on your calculator and it gives you the P-value. Draw a picture of the appropriate sampling distribution. Center the distribution at the value used in the statement of your null hypothesis. Afterwards, label the location of the sample mean or sample proportion along the x axis. Finally, sketch the graph of the standardized sampling distribution (either the z or t distribution), and label the location of the test statistic. Draw a vertical line at that location. Then, use the flowchart below to help you sketch the correct area under the sampling distribution that represents the p-value.



How do I use the P-value to decide whether or not to reject the null hypothesis?

- **If the p-val $\leq \alpha$ then reject the null hypothesis.**
- **If the p-val > α then do not reject the null hypothesis.**

What is α (alpha)?

Definition

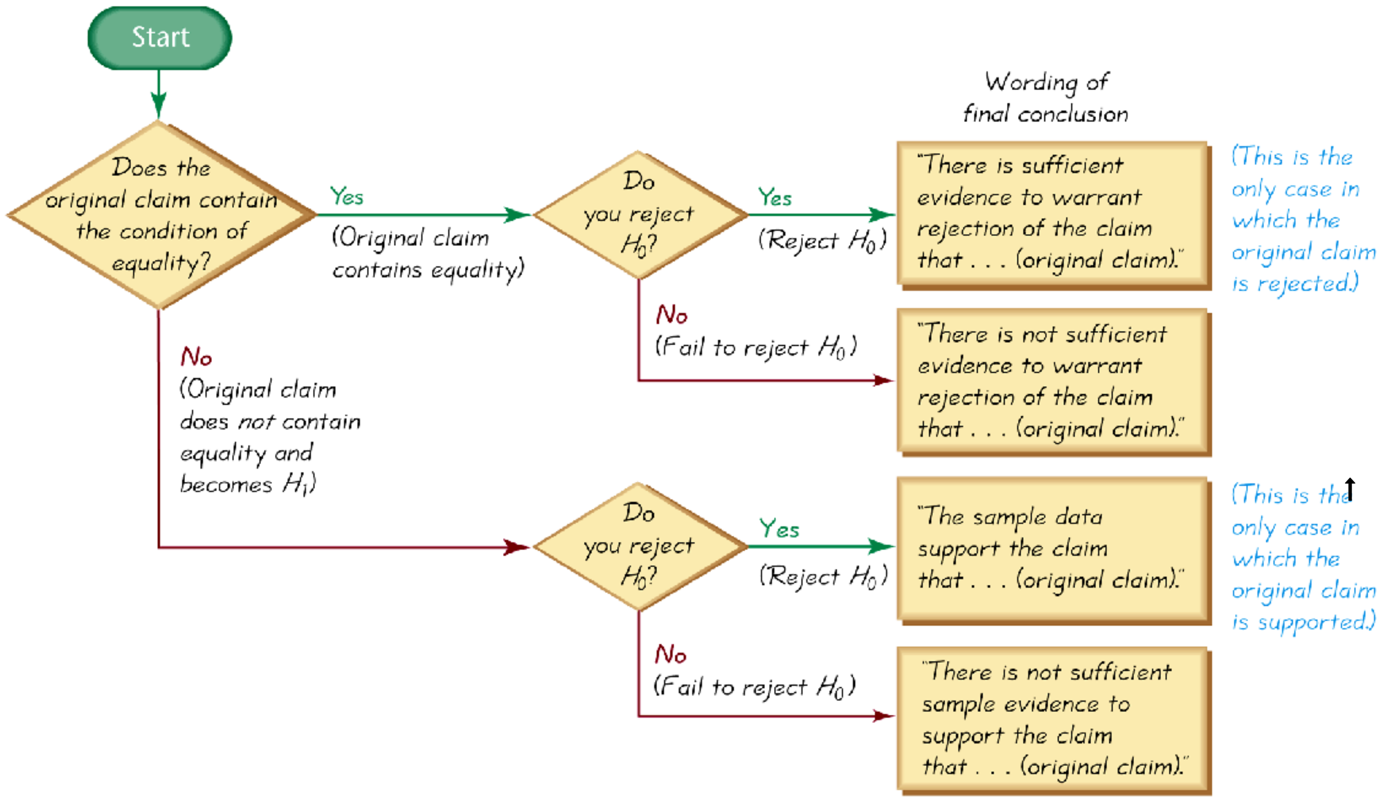
A type I (or α type) error occurs if you reject the null hypothesis (conclude the sample evidence suggests it is false), when it is really true.

Definition

The level of significance is the maximum probability of committing a type I error. This probability is symbolized by α (Greek letter alpha). That is, P(type I error) = α.

As a researcher you will preselect a value for the level of significance, α. For the problems we do in this class, you will be given a value for α in the statement of the problem. If you are not given a value of α in the statement of the problem then use a 5% significance level ($\alpha = 0.05$).

5. **Write the full sentence conclusion (result) of the hypothesis test.** Use the wording from the flow chart given below



Hypothesis Test Conclusion	This means that...
Reject H_0	There is convincing evidence against the null hypothesis. If H_0 were true, the sample data would be very surprising.
Fail to Reject H_0	There is not convincing evidence against the null hypothesis. If H_0 were true, the sample data would not be considered surprising.

Parameter being tested	test statistic	standardized test statistic formula	how to get pvalue and standardized test stat	book section
mean, μ (σ known)	$\bar{x}$	$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	z-test	7.2
mean, μ (σ unknown)	$\bar{x}$	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$	t-test	7.3
proportion (%), p	$\hat{p}$	$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$	1-prop-z-test	7.4
standard deviation, σ	s^2	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	χ^2 cdf	7.5
variance, σ^2	s^2	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	χ^2 cdf	7.5
diff between means, $\mu_1 - \mu_2$	$\bar{x_1} - \bar{x_2}$	$z = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	2-Samp-z-test	8.1
diff between props, $p_1 - p_2$	$\hat{p_1} - \hat{p_2}$	$z = \frac{(\hat{p_1} - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	2-prop-z-test	8.4

with

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} \qquad and \qquad \bar{q} = 1 - \bar{p}$$

Table 1—Binomial Distribution

This table shows the probability of x successes in n independent trials, each with probability of success p.

		p																			
n	**x**	**.01**	**.05**	**.10**	**.15**	**.20**	**.25**	**.30**	**.35**	**.40**	**.45**	**.50**	**.55**	**.60**	**.65**	**.70**	**.75**	**.80**	**.85**	**.90**	**.95**
2	0	.980	.902	.810	.723	.640	.563	.490	.423	.360	.303	.250	.203	.160	.123	.090	.063	.040	.023	.010	.002
	1	.020	.095	.180	.255	.320	.375	.420	.455	.480	.495	.500	.495	.480	.455	.420	.375	.320	.255	.180	.095
	2	.000	.002	.010	.023	.040	.063	.090	.123	.160	.203	.250	.303	.360	.423	.490	.563	.640	.723	.810	.902
3	0	.970	.857	.729	.614	.512	.422	.343	.275	.216	.166	.125	.091	.064	.043	.027	.016	.008	.003	.001	.000
	1	.029	.135	.243	.325	.384	.422	.441	.444	.432	.408	.375	.334	.288	.239	.189	.141	.096	.057	.027	.007
	2	.000	.007	.027	.057	.096	.141	.189	.239	.288	.334	.375	.408	.432	.444	.441	.422	.384	.325	.243	.135
	3	.000	.000	.001	.003	.008	.016	.027	.043	.064	.091	.125	.166	.216	.275	.343	.422	.512	.614	.729	.857
4	0	.961	.815	.656	.522	.410	.316	.240	.179	.130	.092	.062	.041	.026	.015	.008	.004	.002	.001	.000	.000
	1	.039	.171	.292	.368	.410	.422	.412	.384	.346	.300	.250	.200	.154	.112	.076	.047	.026	.011	.004	.000
	2	.001	.014	.049	.098	.154	.211	.265	.311	.346	.368	.375	.368	.346	.311	.265	.211	.154	.098	.049	.014
	3	.000	.000	.004	.011	.026	.047	.076	.112	.154	.200	.250	.300	.346	.384	.412	.422	.410	.368	.292	.171
	4	.000	.000	.000	.001	.002	.004	.008	.015	.026	.041	.062	.092	.130	.179	.240	.316	.410	.522	.656	.815
5	0	.951	.774	.590	.444	.328	.237	.168	.116	.078	.050	.031	.019	.010	.005	.002	.001	.000	.000	.000	.000
	1	.048	.204	.328	.392	.410	.396	.360	.312	.259	.206	.156	.113	.077	.049	.028	.015	.006	.002	.000	.000
	2	.001	.021	.073	.138	.205	.264	.309	.336	.346	.337	.312	.276	.230	.181	.132	.088	.051	.024	.008	.001
	3	.000	.001	.008	.024	.051	.088	.132	.181	.230	.276	.312	.337	.346	.336	.309	.264	.205	.138	.073	.021
	4	.000	.000	.000	.002	.006	.015	.028	.049	.077	.113	.156	.206	.259	.312	.360	.396	.410	.392	.328	.204
	5	.000	.000	.000	.000	.000	.001	.002	.005	.010	.019	.031	.050	.078	.116	.168	.237	.328	.444	.590	.774
6	0	.941	.735	.531	.377	.262	.178	.118	.075	.047	.028	.016	.008	.004	.002	.001	.000	.000	.000	.000	.000
	1	.057	.232	.354	.399	.393	.356	.303	.244	.187	.136	.094	.061	.037	.020	.010	.004	.002	.000	.000	.000
	2	.001	.031	.098	.176	.246	.297	.324	.328	.311	.278	.234	.186	.138	.095	.060	.033	.015	.006	.001	.000
	3	.000	.002	.015	.042	.082	.132	.185	.236	.276	.303	.312	.303	.276	.236	.185	.132	.082	.042	.015	.002
	4	.000	.000	.001	.006	.015	.033	.060	.095	.138	.186	.234	.278	.311	.328	.324	.297	.246	.176	.098	.031
	5	.000	.000	.000	.000	.002	.004	.010	.020	.037	.061	.094	.136	.187	.244	.303	.356	.393	.399	.354	.232
	6	.000	.000	.000	.000	.000	.000	.001	.002	.004	.008	.016	.028	.047	.075	.118	.178	.262	.377	.531	.735
7	0	.932	.698	.478	.321	.210	.133	.082	.049	.028	.015	.008	.004	.002	.001	.000	.000	.000	.000	.000	.000
	1	.066	.257	.372	.396	.367	.311	.247	.185	.131	.087	.055	.032	.017	.008	.004	.001	.000	.000	.000	.000
	2	.002	.041	.124	.210	.275	.311	.318	.299	.261	.214	.164	.117	.077	.047	.025	.012	.004	.001	.000	.000
	3	.000	.004	.023	.062	.115	.173	.227	.268	.290	.292	.273	.239	.194	.144	.097	.058	.029	.011	.003	.000
	4	.000	.000	.003	.011	.029	.058	.097	.144	.194	.239	.273	.292	.290	.268	.227	.173	.115	.062	.023	.004
	5	.000	.000	.000	.001	.004	.012	.025	.047	.077	.117	.164	.214	.261	.299	.318	.311	.275	.210	.124	.041
	6	.000	.000	.000	.000	.000	.001	.004	.008	.017	.032	.055	.087	.131	.185	.247	.311	.367	.396	.372	.257
	7	.000	.000	.000	.000	.000	.000	.000	.001	.002	.004	.008	.015	.028	.049	.082	.133	.210	.321	.478	.698
8	0	.923	.663	.430	.272	.168	.100	.058	.032	.017	.008	.004	.002	.001	.000	.000	.000	.000	.000	.000	.000
	1	.075	.279	.383	.385	.336	.267	.198	.137	.090	.055	.031	.016	.008	.003	.001	.000	.000	.000	.000	.000
	2	.003	.051	.149	.238	.294	.311	.296	.259	.209	.157	.109	.070	.041	.022	.010	.004	.001	.000	.000	.000
	3	.000	.005	.033	.084	.147	.208	.254	.279	.279	.257	.219	.172	.124	.081	.047	.023	.009	.003	.000	.000
	4	.000	.000	.005	.018	.046	.087	.136	.188	.232	.263	.273	.263	.232	.188	.136	.087	.046	.018	.005	.000
	5	.000	.000	.000	.003	.009	.023	.047	.081	.124	.172	.219	.257	.279	.279	.254	.208	.147	.084	.033	.005
	6	.000	.000	.000	.000	.001	.004	.010	.022	.041	.070	.109	.157	.209	.259	.296	.311	.294	.238	.149	.051
	7	.000	.000	.000	.000	.000	.000	.001	.003	.008	.016	.031	.055	.090	.137	.198	.267	.336	.385	.383	.279
	8	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.004	.008	.017	.032	.058	.100	.168	.272	.430	.663
9	0	.914	.630	.387	.232	.134	.075	.040	.021	.010	.005	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000
	1	.083	.299	.387	.368	.302	.225	.156	.100	.060	.034	.018	.008	.004	.001	.000	.000	.000	.000	.000	.000
	2	.003	.063	.172	.260	.302	.300	.267	.216	.161	.111	.070	.041	.021	.010	.004	.001	.000	.000	.000	.000
	3	.000	.008	.045	.107	.176	.234	.267	.272	.251	.212	.164	.116	.074	.042	.021	.009	.003	.001	.000	.000
	4	.000	.001	.007	.028	.066	.117	.172	.219	.251	.260	.246	.213	.167	.118	.074	.039	.017	.005	.001	.000
	5	.000	.000	.001	.005	.017	.039	.074	.118	.167	.213	.246	.260	.251	.219	.172	.117	.066	.028	.007	.001
	6	.000	.000	.000	.001	.003	.009	.021	.042	.074	.116	.164	.212	.251	.272	.267	.234	.176	.107	.045	.008
	7	.000	.000	.000	.000	.000	.001	.004	.010	.021	.041	.070	.111	.161	.216	.267	.300	.302	.260	.172	.063
	8	.000	.000	.000	.000	.000	.000	.000	.001	.004	.008	.018	.034	.060	.100	.156	.225	.302	.368	.387	.299
	9	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.005	.010	.021	.040	.075	.134	.232	.387	.630

Table 1—Binomial Distribution *(continued)*

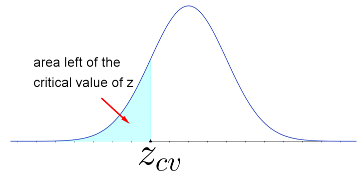
		p																			
n	***x***	**.01**	**.05**	**.10**	**.15**	**.20**	**.25**	**.30**	**.35**	**.40**	**.45**	**.50**	**.55**	**.60**	**.65**	**.70**	**.75**	**.80**	**.85**	**.90**	**.95**
10	0	.904	.599	.349	.197	.107	.056	.028	.014	.006	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.091	.315	.387	.347	.268	.188	.121	.072	.040	.021	.010	.004	.002	.000	.000	.000	.000	.000	.000	.000
	2	.004	.075	.194	.276	.302	.282	.233	.176	.121	.076	.044	.023	.011	.004	.001	.000	.000	.000	.000	.000
	3	.000	.010	.057	.130	.201	.250	.267	.252	.215	.166	.117	.075	.042	.021	.009	.003	.001	.000	.000	.000
	4	.000	.001	.011	.040	.088	.146	.200	.238	.251	.238	.205	.160	.111	.069	.037	.016	.006	.001	.000	.000
	5	.000	.000	.001	.008	.026	.058	.103	.154	.201	.234	.246	.234	.201	.154	.103	.058	.026	.008	.001	.000
	6	.000	.000	.000	.001	.006	.016	.037	.069	.111	.160	.205	.238	.251	.238	.200	.146	.088	.040	.011	.001
	7	.000	.000	.000	.000	.001	.003	.009	.021	.042	.075	.117	.166	.215	.252	.267	.250	.201	.130	.057	.010
	8	.000	.000	.000	.000	.000	.000	.001	.004	.011	.023	.044	.076	.121	.176	.233	.282	.302	.276	.194	.075
	9	.000	.000	.000	.000	.000	.000	.000	.000	.002	.004	.010	.021	.040	.072	.121	.188	.268	.347	.387	.315
	10	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.006	.014	.028	.056	.107	.197	.349	.599
11	0	.895	.569	.314	.167	.086	.042	.020	.009	.004	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.099	.329	.384	.325	.236	.155	.093	.052	.027	.013	.005	.002	.001	.000	.000	.000	.000	.000	.000	.000
	2	.005	.087	.213	.287	.295	.258	.200	.140	.089	.051	.027	.013	.005	.002	.001	.000	.000	.000	.000	.000
	3	.000	.014	.071	.152	.221	.258	.257	.225	.177	.126	.081	.046	.023	.010	.004	.001	.000	.000	.000	.000
	4	.000	.001	.016	.054	.111	.172	.220	.243	.236	.206	.161	.113	.070	.038	.017	.006	.002	.000	.000	.000
	5	.000	.000	.002	.013	.039	.080	.132	.183	.221	.236	.226	.193	.147	.099	.057	.027	.010	.002	.000	.000
	6	.000	.000	.000	.002	.010	.027	.057	.099	.147	.193	.226	.236	.221	.183	.132	.080	.039	.013	.002	.000
	7	.000	.000	.000	.000	.002	.006	.017	.038	.070	.113	.161	.206	.236	.243	.220	.172	.111	.054	.016	.001
	8	.000	.000	.000	.000	.000	.001	.004	.010	.023	.046	.081	.126	.177	.225	.257	.258	.221	.152	.071	.014
	9	.000	.000	.000	.000	.000	.000	.001	.002	.005	.013	.027	.051	.089	.140	.200	.258	.295	.287	.213	.087
	10	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.005	.013	.027	.052	.093	.155	.236	.325	.384	.329
	11	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.004	.009	.020	.042	.086	.167	.314	.569
12	0	.886	.540	.282	.142	.069	.032	.014	.006	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.107	.341	.377	.301	.206	.127	.071	.037	.017	.008	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000
	2	.006	.099	.230	.292	.283	.232	.168	.109	.064	.034	.016	.007	.002	.001	.000	.000	.000	.000	.000	.000
	3	.000	.017	.085	.172	.236	.258	.240	.195	.142	.092	.054	.028	.012	.005	.001	.000	.000	.000	.000	.000
	4	.000	.002	.021	.068	.133	.194	.231	.237	.213	.170	.121	.076	.042	.020	.008	.002	.001	.000	.000	.000
	5	.000	.000	.004	.019	.053	.103	.158	.204	.227	.223	.193	.149	.101	.059	.029	.011	.003	.001	.000	.000
	6	.000	.000	.000	.004	.016	.040	.079	.128	.177	.212	.226	.212	.177	.128	.079	.040	.016	.004	.000	.000
	7	.000	.000	.000	.001	.003	.011	.029	.059	.101	.149	.193	.223	.227	.204	.158	.103	.053	.019	.004	.000
	8	.000	.000	.000	.000	.001	.002	.008	.020	.042	.076	.121	.170	.213	.237	.231	.194	.133	.068	.021	.002
	9	.000	.000	.000	.000	.000	.000	.001	.005	.012	.028	.054	.092	.142	.195	.240	.258	.236	.172	.085	.017
	10	.000	.000	.000	.000	.000	.000	.000	.001	.002	.007	.016	.034	.064	.109	.168	.232	.283	.292	.230	.099
	11	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.008	.017	.037	.071	.127	.206	.301	.377	.341
	12	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.006	.014	.032	.069	.142	.282	.540
15	0	.860	.463	.206	.087	.035	.013	.005	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.130	.366	.343	.231	.132	.067	.031	.013	.005	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	2	.009	.135	.267	.286	.231	.156	.092	.048	.022	.009	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000
	3	.000	.031	.129	.218	.250	.225	.170	.111	.063	.032	.014	.005	.002	.000	.000	.000	.000	.000	.000	.000
	4	.000	.005	.043	.116	.188	.225	.219	.179	.127	.078	.042	.019	.007	.002	.001	.000	.000	.000	.000	.000
	5	.000	.001	.010	.045	.103	.165	.206	.212	.186	.140	.092	.051	.024	.010	.003	.001	.000	.000	.000	.000
	6	.000	.000	.002	.013	.043	.092	.147	.191	.207	.191	.153	.105	.061	.030	.012	.003	.001	.000	.000	.000
	7	.000	.000	.000	.003	.014	.039	.081	.132	.177	.201	.196	.165	.118	.071	.035	.013	.003	.001	.000	.000
	8	.000	.000	.000	.001	.003	.013	.035	.071	.118	.165	.196	.201	.177	.132	.081	.039	.014	.003	.000	.000
	9	.000	.000	.000	.000	.001	.003	.012	.030	.061	.105	.153	.191	.207	.191	.147	.092	.043	.013	.002	.000
	10	.000	.000	.000	.000	.000	.001	.003	.010	.024	.051	.092	.140	.186	.212	.206	.165	.103	.045	.010	.001
	11	.000	.000	.000	.000	.000	.000	.001	.002	.007	.019	.042	.078	.127	.179	.219	.225	.188	.116	.043	.005
	12	.000	.000	.000	.000	.000	.000	.000	.000	.002	.005	.014	.032	.063	.111	.170	.225	.250	.218	.129	.031
	13	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.009	.022	.048	.092	.156	.231	.286	.267	.135
	14	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.002	.005	.013	.031	.067	.132	.231	.343	.366
	15	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.002	.005	.013	.035	.087	.206	.463

Table 1—Binomial Distribution *(continued)*

		p																			
n	***x***	**.01**	**.05**	**.10**	**.15**	**.20**	**.25**	**.30**	**.35**	**.40**	**.45**	**.50**	**.55**	**.60**	**.65**	**.70**	**.75**	**.80**	**.85**	**.90**	**.95**
16	0	.851	.440	.185	.074	.028	.010	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.138	.371	.329	.210	.113	.053	.023	.009	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	2	.010	.146	.275	.277	.211	.134	.073	.035	.015	.006	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000
	3	.000	.036	.142	.229	.246	.208	.146	.089	.047	.022	.009	.003	.001	.000	.000	.000	.000	.000	.000	.000
	4	.000	.006	.051	.131	.200	.225	.204	.155	.101	.057	.028	.011	.004	.001	.000	.000	.000	.000	.000	.000
	5	.000	.001	.014	.056	.120	.180	.210	.201	.162	.112	.067	.034	.014	.005	.001	.000	.000	.000	.000	.000
	6	.000	.000	.003	.018	.055	.110	.165	.198	.198	.168	.122	.075	.039	.017	.006	.001	.000	.000	.000	.000
	7	.000	.000	.000	.005	.020	.052	.101	.152	.189	.197	.175	.132	.084	.044	.019	.006	.001	.000	.000	.000
	8	.000	.000	.000	.001	.006	.020	.049	.092	.142	.181	.196	.181	.142	.092	.049	.020	.006	.001	.000	.000
	9	.000	.000	.000	.000	.001	.006	.019	.044	.084	.132	.175	.197	.189	.152	.101	.052	.020	.005	.000	.000
	10	.000	.000	.000	.000	.000	.001	.006	.017	.039	.075	.122	.168	.198	.198	.165	.110	.055	.018	.003	.000
	11	.000	.000	.000	.000	.000	.000	.001	.005	.014	.034	.067	.112	.162	.201	.210	.180	.120	.056	.014	.001
	12	.000	.000	.000	.000	.000	.000	.000	.001	.004	.011	.028	.057	.101	.155	.204	.225	.200	.131	.051	.006
	13	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.009	.022	.047	.089	.146	.208	.246	.229	.142	.036
	14	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.006	.015	.035	.073	.134	.211	.277	.275	.146
	15	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.009	.023	.053	.113	.210	.329	.371
	16	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.010	.028	.074	.185	.440
20	0	.818	.358	.122	.039	.012	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.165	.377	.270	.137	.058	.021	.007	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	2	.016	.189	.285	.229	.137	.067	.028	.010	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	3	.001	.060	.190	.243	.205	.134	.072	.032	.012	.004	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
	4	.000	.013	.090	.182	.218	.190	.130	.074	.035	.014	.005	.001	.000	.000	.000	.000	.000	.000	.000	.000
	5	.000	.002	.032	.103	.175	.202	.179	.127	.075	.036	.015	.005	.001	.000	.000	.000	.000	.000	.000	.000
	6	.000	.000	.009	.045	.109	.169	.192	.171	.124	.075	.036	.015	.005	.001	.000	.000	.000	.000	.000	.000
	7	.000	.000	.002	.016	.055	.112	.164	.184	.166	.122	.074	.037	.015	.005	.001	.000	.000	.000	.000	.000
	8	.000	.000	.000	.005	.022	.061	.114	.161	.180	.162	.120	.073	.035	.014	.004	.001	.000	.000	.000	.000
	9	.000	.000	.000	.001	.007	.027	.065	.116	.160	.177	.160	.119	.071	.034	.012	.003	.000	.000	.000	.000
	10	.000	.000	.000	.000	.002	.010	.031	.069	.117	.159	.176	.159	.117	.069	.031	.010	.002	.000	.000	.000
	11	.000	.000	.000	.000	.000	.003	.012	.034	.071	.119	.160	.177	.160	.116	.065	.027	.007	.001	.000	.000
	12	.000	.000	.000	.000	.000	.001	.004	.014	.035	.073	.120	.162	.180	.161	.114	.061	.022	.005	.000	.000
	13	.000	.000	.000	.000	.000	.000	.001	.005	.015	.037	.074	.122	.166	.184	.164	.112	.055	.016	.002	.000
	14	.000	.000	.000	.000	.000	.000	.000	.001	.005	.015	.037	.075	.124	.171	.192	.169	.109	.045	.009	.000
	15	.000	.000	.000	.000	.000	.000	.000	.000	.001	.005	.015	.036	.075	.127	.179	.202	.175	.103	.032	.002
	16	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.005	.014	.035	.074	.130	.190	.218	.182	.090	.013
	17	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.004	.012	.032	.072	.134	.205	.243	.190	.060
	18	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.010	.028	.067	.137	.229	.285	.189
	19	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.002	.007	.021	.058	.137	.270	.377
	20	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.012	.039	.122	.358

The Standard Normal Distribution of z scores

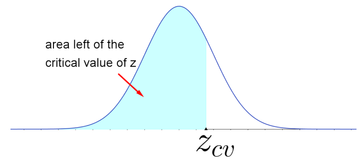
Table values represent the AREA to the LEFT of the z score value.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.5^-	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

The Standard Normal Distribution of z scores

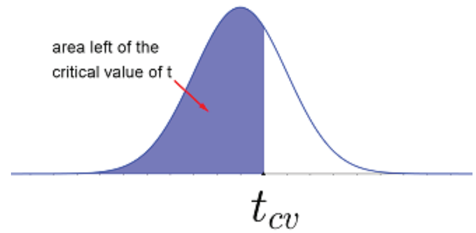
Table values represent the AREA to the LEFT of the z score value.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

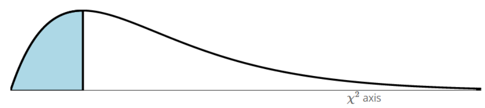
The Student t Distribution

The percentage at the top of the table is equal to the AREA to the LEFT of the t score value.



df	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.434	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.434	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.434	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.433	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.255	0.433	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.254	0.433	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090

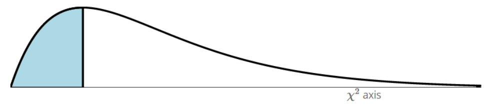
The χ^2 Distribution Table



The percentage at the top of the table is equal to the AREA to the LEFT of χ^2 — where χ^2 is located both along the horizontal axis and in the table body.

df	0.1%	0.5%	1.0%	2.5%	5.0%	10.0%	12.5%	20.0%	25.0%	33.3%	50.0%
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335
55	28.173	31.735	33.570	36.398	38.958	42.060	43.220	46.036	47.610	49.972	54.335
60	31.738	35.534	37.485	40.482	43.188	46.459	47.680	50.641	52.294	54.770	59.335

The χ^2 Distribution Table



The percentage at the top of the table is equal to the AREA to the LEFT of χ^2 — where χ^2 is located both along the horizontal axis and in the table body.

df	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661
55	57.016	58.945	61.665	63.577	67.211	68.796	73.311	77.380	82.292	85.749	93.168
60	62.135	64.147	66.981	68.972	72.751	74.397	79.082	83.298	88.379	91.952	99.607

Critical Values for the Pearson Correlation Coefficient

The correlation is significant when the absolute value of *r* is greater than the value in the table.

n	$\alpha = 0.05$	$\alpha = 0.01$
4	0.950	0.990
5	0.878	0.959
6	0.811	0.917
7	0.754	0.875
8	0.707	0.834
9	0.666	0.798
10	0.632	0.765
11	0.602	0.735
12	0.576	0.708
13	0.553	0.684
14	0.532	0.661
15	0.514	0.641
16	0.497	0.623
17	0.482	0.606
18	0.468	0.590
19	0.456	0.575
20	0.444	0.561
21	0.433	0.549
22	0.423	0.537
23	0.413	0.526
24	0.404	0.515
25	0.396	0.505
26	0.388	0.496
27	0.381	0.487
28	0.374	0.479
29	0.367	0.471
30	0.361	0.463
35	0.334	0.430
40	0.312	0.403
45	0.294	0.380
50	0.279	0.361
55	0.266	0.345
60	0.254	0.330
65	0.244	0.317
70	0.235	0.306
75	0.227	0.296
80	0.220	0.286
85	0.213	0.278
90	0.207	0.270
95	0.202	0.263
100	0.197	0.256

Statistics Test 1 **Name:** ____________________

1. ***Determine whether the given value is a statistic or a parameter.***

   (a) Burrito Bell fast food restaurant is interested in finding the average ages of its employees. The company gets the age of each employee from its Human Resources department. Those ages are entered into a list and the mean age is found to be 34 years old.

   (b) Burrito Bell fast food restaurant is interested in finding the average ages of its employees. A random sample of 50 employee ages are taken from company records. Those ages are entered into a list and the mean age is found to be 34 years old.

   (c) Krusty-O Cereal Manufacturing is interested in finding the average number of boxes of Krusty-0s that each household of its customers consume each week. Surveys forms were printed on the back of each box and among those who responded, the average number of boxes consumed was 2.2.

   (d) 80% of the different kinds of elements on the Periodic table naturally occur in nature. The rest are man made (synthetic).

2. ***In each of these statements, tell whether descriptive or inferential statistics have been used.***

   (a) The report by the Medicare Office of the Actuary estimated that health spending will grow by an average of 5.8 percent a year through 2020, compared to 5.7 percent without the health overhaul.

   (b) Expenditures for the cable industry were $5.66 billion in 1996 (Source: USA TODAY).

   (c) The mean travel time to work (years 2008–2012) for San Diego County workers age 16+ is 24.2 minutes. (Source: census.gov).

   (d) Allergy therapy makes bees go away (Source: Prevention).

3. ***Classify each variable as discrete or continuous.***

   (a) Number of consumers in a poll who prefer Krusty-O cereal over Cheerios

   (b) Heights of buildings in New York City

   (c) The ages of Burger World employees

4. ***Classify each variable as qualitative or quantitative.***

(a) A Gallup poll asked Americans, "Which subject (Math, English, History,...), if any, has been the most valuable in your life?"

(b) The number of donuts sold each week at Dunkin Donuts

(c) The different flavors of donuts at Dunkin Donuts

(d) Heights of NBA athletes

5. ***Classify each variable as discrete or continuous.***

(a) The number of people who buy a coffee from Starbucks today.

(b) The volume of water (in cubic feet) of each of the Great Lakes.

(c) The number of drones Amazon.com wants to have delivering orders by 2020.

(d) The amount of time it takes to answer this question.

6. ***Classify each sample as a simple random sample, systematic, stratified, cluster or convenience.***

(a) To determine his blood pressure, Stan divides up his day into three parts: morning, afternoon and evening. He then measures his blood pressure at 3 randomly selected times during each part of the day.

(b) To estimate the percentage of defects in a box of Krusty-O cereal, a quality control engineer inspects every 120^{th} box of Krusty-Os that is produced by the factory.

(c) To determine customer opinion of their inflight service (for all of its airplanes) on a particular day, LeastWorst Airlines selects 100 of its flights leaving San Diego International Airport and surveys all the participants of each flight.

(d) To determine customer opinion of their inflight service (for all of its airplanes) on a particular day, LeastWorst Airlines randomly selects 2 people from every LeastWorst flight scheduled for that day to participate in their survey.

(e) A man experienced a tax audit. The tax department claimed that the man was audited because he was randomly selected from all men in his age group.

(f) A San Diego television station asks its viewers to call in their opinion regarding the weather programming.

7. ***Classify each sample as a simple random sample, systematic, stratified, cluster or convenience.***

   (a) Each employee of Nerdstroms Department Store has a 7-digit employee number. The district manager wants to find out how happy the employees who work there are with their jobs. She randomly selects 30 employee numbers and surveys the corresponding employees.

   (b) A researcher interviews four guys who are sitting in a booth at Burger World.

   (c) To determine customer opinion of their inflight service for Flight 3568 from San Diego to Vancouver, Canada, LeastWorst Airlines selects every fifth person on Flight 3568 to participate in their survey.

   (d) In order to find out which job title Burger World employees prefer to be called ("associate" or "server"), emails are sent inviting employees to click on a link to participate in an online survey.

   (e) The school needs 50 students to participate in a survey. A random number generator is used to select 50 student identification numbers.

   (f) Mail carriers of a large city are divided into four groups according to gender (male or female) and according to whether they walk or ride on their routes. Then 10 are selected from each group and interviewed to determine whether they have been bitten by a dog in the last year.

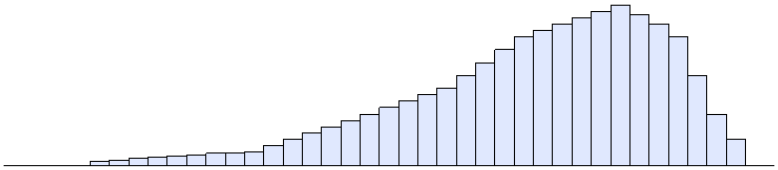
8. ***Identify each study as being an observational study or experimental.***

   (a) Subjects were randomly assigned to two groups, and one group was given an herb and the other group a placebo. After 6 months, the numbers of respiratory tract infections each group had were compared.

   (b) In a study sponsored by LeastWorst Airlines, 1200 customers were asked which food item from the breakfast menu they liked the most, and 32% of respondents said it was the Krusty-O's.

   (c) Subjects are randomly assigned to four groups. Each group is placed on one of four special diets— a low-fat diet, a high-fish diet, a combination of low-fat diet and high-fish diet, and a regular diet. After 6 months, the blood pressures of the groups are compared to see if diet has any effect on blood pressure.

   (d) Athletes who had suffered hamstring injuries were randomly assigned to one of two exercise programs. Those who engaged in static stretching returned to sports activity in a mean of 37.5 days (SD=27.2 days). Those assigned to a program of agility and trunk stabilization exercised returned to sports in a mean of 22.3 days (SD=8.4 days).



9. (1 point) Is the distribution in the histogram above skewed left, skewed right, or is it normal?

9. ____________

10. ***Find the class boundaries, midpoints and widths for the numeric class 23—30.***

   (a) class boundaries (a) ____________

   (b) class midpoint (b) ____________

   (c) class width (c) ____________

11. ***Find the class boundaries, midpoints and widths for the numeric class 3.2—5.2.***

   (a) class boundaries (a) ____________

   (b) class midpoint (b) ____________

   (c) class width (c) ____________

12. ***Find the class boundaries, midpoints and widths for the numeric class 0.25—0.30.***

   (a) class boundaries (a) ____________

   (b) class midpoint (b) ____________

   (c) class width (c) ____________

13. List the original data from the stemplot in an ordered list sorted in a descending fashion.(Assume the leaf units are in the ones decimal place)

Stem	Leaves
22	2 4
23	6 8 9 9
24	1 4 8
25	1 3

14. Sketch a histogram below for which the median would be ***greater*** than the mean.

15. ***Construct the dotplot for the given data.*** Attendance records for a preschool class with 15 students show the number of days each student was absent during the year. The days absent for each student were as follows.

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

16. A questionnaire on political affiliation showed information obtained from 25 respondents. Construct a relative frequency distribution for the data (I = independant, D = democrat, R = republican) given below.

R D I R R I D D I R R D R D I D R D I D R I D R R

17. Construct a pie chart for the data given in the previous question.

18. The data shown (in millions of dollars) are the values of the 30 National Football League franchises. Construct a frequency distribution for the data using 8 classes.

170	191	171	235	173	187	181	191	211	186
200	218	243	200	182	320	184	239	197	204
186	199	186	210	209	240	204	193	188	242

19. Construct a histogram for the data given in the previous question.

20. Construct a frequency polygon for the data used in the previous two questions.

21. ***Use the data to create a stemplot.*** The midterm test scores for the university's seventh-period graduate chemistry class are listed below.

92 73 82 95 77 84 65 67 78 88 56 75 86 72 92 79 77

22. Sample data is summarized in the frequency distribution below. Construct a cumulative frequency distribution.

class	frequency
2—6	13
7—11	64
12—16	49
17—21	23

22. ____________

23. How many measurements were given in the distribution table given in the previous question?

23. ____________

24. Make an ogive for the distribution used in the previous two questions.

25. Sample data is summarized in the frequency distribution below. Find the mean of the distribution.

25. ____________

class	frequency
2—6	13
7—11	64
12—16	49
17—21	23

26. Sample data is summarized in the frequency distribution below. Find the standard deviation of the distribution.

26. ____________

class	frequency
2—6	13
7—11	64
12—16	49
17—21	23

Use this census data to find the values of each parameter.

92 73 82 95 77 84 65 67 78 88 56 75 86 72 92 79 77

27. (1 point) mean 27. ____________

28. (1 point) median 28. ____________

29. (1 point) mode 29. ____________

30. (1 point) standard deviation 30. ____________

31. (1 point) variance 31. ____________

32. (1 point) range 32. ____________

33. (1 point) minimum 33. ____________

34. (1 point) quartile 1 34. ____________

35. (1 point) quartile 2 35. ____________

36. (1 point) quartile 3 36. ____________

37. (1 point) maximum 37. ____________

38. (2 points) What is the interquartile range? 38. ____________

39. (2 points) What two values provide the lower and upper fences for the data?

39. ____________

40. (1 point) Are there any outliers? Explain your reasoning to obtain full credit.

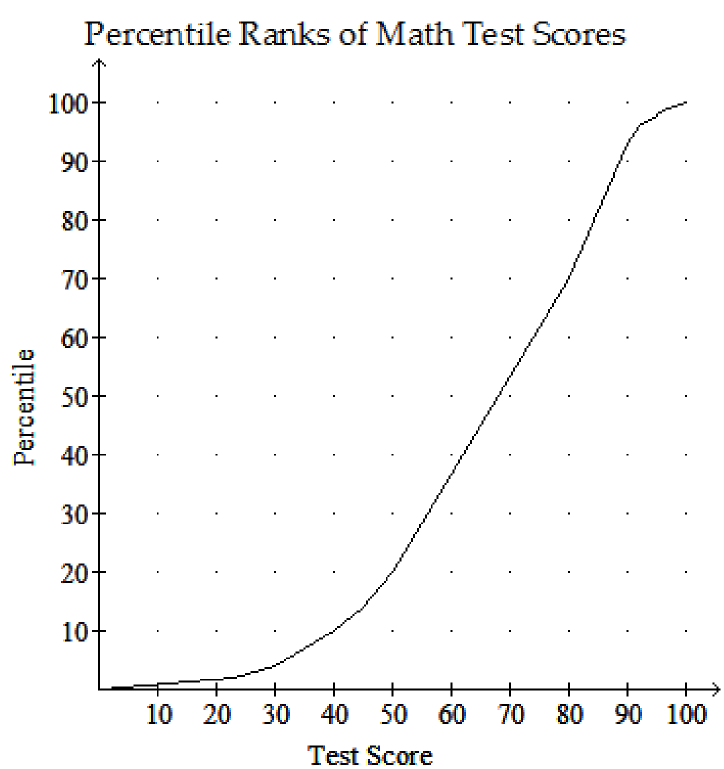
41. (2 points) Construct a boxplot for the data.

Use the census data on the previous page to answer questions 43 and 44.

42. (2 points) 86 represents what percentile?

43. (2 points) 67 represents what percentile?

44. (2 points) The graph below is an ogive of scores on a math test. Use the graph to approximate the percentile rank of an individual whose test score is 70.



45. The average labor charge for automobile mechanics is \$54 per hour. The standard deviation is \$4. Find the minimum percentage of data values that will fall within the range of \$48 to \$60. Use Chebyshev's theorem.

46. A survey of car rental agencies shows that the average cost of a car rental is \$0.32 per mile. The standard deviation is \$0.03. Using Chebyshev's theorem, find the range in which at least 75% of the data values will fall.

47. The mean of the times it takes a commuter to get to work in Baltimore is 29.7 minutes. Assume the distribution of commuter times is approximately bellshaped.
    (a) If the standard deviation is 6 minutes, within what limits would you expect approximately 68% of the times to fall?
    (b) Within what limits would you expect approximately 95% of the times to fall?
    (c) Within what limits would you expect approximately 99.7% of the times to fall?

48. Identify the population and the sample. A survey of 1009 U.S. adults found that 26% think higher education is affordable for everyone who needs it.

49. A television station claims that the amount of advertising per hour of broadcast time has an average of 16 minutes and a standard deviation equal to 2.9 minutes. You watch the station for 1 hour, at a randomly selected time, and carefully observe that the amount of advertising time is equal to 9 minutes. Calculate the z-score for this amount of advertising time. Round to the nearest hundredth.

50. Using your z-score from the previous question, is 9 minutes of advertising time considered usual, unusual or very unusual. Explain your reasoning.

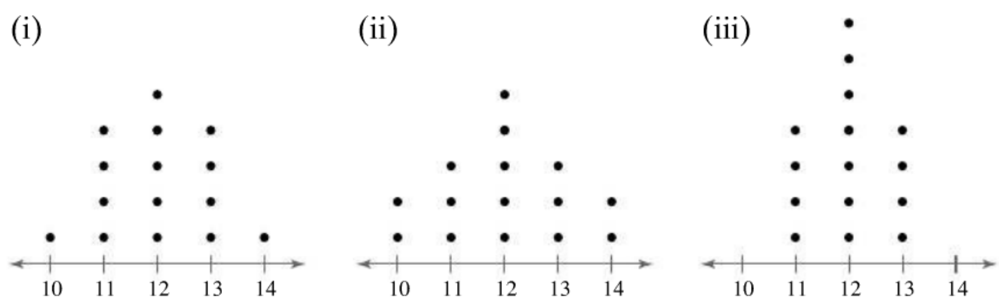
51. The mean age of lead actresses from the top ten grossing movies of 2010 was 29.6 years with a standard deviation of 6.35 years. Assume the distribution of the actresses' ages is approximately bell-shaped. In 2010, popular actress Jennifer Aniston was 41-years-old. What is Jennifer Aniston's age if it is standardized? Round to the nearest hundredth.

52. Would it be usual, unusual or very unusual for a 41-year-old actress to be in a top-grossing film of 2010? Assume the Empirical Rule applies. Explain your reasoning.

53. A television station claims that the amount of advertising per hour of broadcast time has an average of 17 minutes and a standard deviation equal to 1.1 minutes. You watch the station for 1 hour, at a randomly selected time, and carefully observe that the amount of advertising time is equal to 9 minutes. Calculate the z-score for this amount of advertising time.

54. Using your z-score from the previous question, is 9 minutes of advertising time considered usual, unusual or very unusual. Explain your reasoning.

55. Without calculating, which of the three data sets have the greatest standard deviation, which has the smallest standard deviation, and which has the second largest standard deviation?



56. ***Classify each statement as an example of classical probability, empirical probability, or subjective probability.***

(a) An executive for the Krusty-O cereal factory makes an educated guess as to how well a new flavor of Krusty-Os will sell.

(b) The probability of getting a 0 in roulette.

(c) The probability that a randomly selected driver having a California drivers license will have brown eyes.

(d) In a recent year, among 135,933,000 registered passenger cars in the U.S., there were 10,427,000 crashes. So, the probability that a randomly selected passenger car in the U.S. will crash this year is $\frac{10{,}427{,}000}{135{,}933{,}000} \doteq 0.0767$.

57. In 2013, 32.3% of LeastWorst Airlines customers who purchased a ticket spent an additional \$20 to be in the first boarding group. Choose one LeastWorst customer at random. What is the probability that the customer *didn't* spend the additional \$20 to be in the first boarding group?

57. ____________

58. ***True or False.*** Two events are mutually exclusive if they cannot both occur.

58. ____________

59. ***Determine whether these events are mutually exclusive.***

(a) Roll a die: Roll an even number, and roll a 3.

(b) A single card is drawn from a standard deck: the card is a heart, and the card is a ace of clubs.

(c) Select a student in your class: the student is texting, the student is not a math major.

(d) Select a pair of shoes: the shoes are brown colored, the shoes are made in the USA.

60. A candy dish contains four red candies, seven yellow candies and fourteen blue candies. You close your eyes, choose two candies one at a time (without replacement) from the dish, and record their colors.

(a) Find the probability that both candies are red.

(a) ____________

(b) Find the probability that the first candy is red and the second candy is blue.

(b) ____________

61. Suppose that $P(A) = 0.2$ and $P(B) = 0.3$. If events A and B are independent, find $P(A \text{ and } B)$.

61. ____________

62. Suppose that $P(A) = 0.2$ and $P(B) = 0.3$. If events A and B are mutually exclusive, find $P(A \text{ or } B)$.

62. ____________

63. Toss a single, six-sided die three times. Find the probability that all three rolls are fives.

63. ____________

64. If 28% of U.S. medical degrees are conferred to women, find the probability that 3 randomly selected medical school graduates are men. Would you consider this event likely or unlikely to occur?

64. ____________

65. The human resources division at the Krusty-O cereal factory reports a breakdown of employees by job type and sex, summarized in the table below.

	Sex		
Job Type	Male	Female	total
Management	7	6	13
Supervision	8	12	20
Production	45	72	117
total	60	90	150

One of these workers is randomly selected.

(a) (2 points) Find the probability that the worker is a female.

(b) (2 points) Find the probability that the worker is a female or a supervisor.

(c) (2 points) Find the probability that the worker is male with the Supervision job type.

(d) (2 points) Find the probability that the worker is female, given that the person works in production.

(e) (2 points) Find the probability that the worker works in production and is a female.

(f) (2 points) Find the probability that the worker works in production or is a female.

66. Voter Support for political term limits is strong in many parts of the U.S. A poll conducted by the Field Institute in California showed support for term limits by a 2–1 margin. The results of this poll of $n = 347$ registered voters are given in the table.

	For (F)	**Against (A)**	**No Opinion (N)**	**Total**
Republican (R)	0.28	0.10	0.02	0.40
Democrat (D)	0.31	0.16	0.03	0.50
Other (O)	0.06	0.04	0.00	0.10
Total	0.65	0.30	0.05	1.00

If one individual in drawn at random from this group of 347 people, calculate the following probabilities:

(a) $P(N)$ (a) ____________

(b) $P(D \text{ and } A)$ (b) ____________

(c) $P(D \text{ or } A)$ (c) ____________

(d) $P(A \text{ or } O)$ (d) ____________

(e) $P(A \text{ and } O)$ (e) ____________

(f) $P(\overline{N})$ (f) ____________

(g) $P(R|N)$ (g) ____________

(h) $P(A|D)$ (h) ____________

(i) $P(D|A)$ (i) ____________

	Nonsmoker	Light Smoker	Heavy Smoker	Total
Men	306	74	66	446
Women	345	68	81	494
Total	651	142	147	940

Consider the following events:

Event N: The person selected is a nonsmoker
Event L: The person selected is a light smoker
Event H: The person selected is a heavy smoker
Event M: The person selected is a male
Event F: The person selected is a female

67. Suppose one of the 940 subjects is chosen at random. Determine the following probabilities:

a. $P(N|F)$

b. $P(F|N)$

c. $P(H \cup M)$

d. $P(M \cap L)$

e. $P(\text{the person selected is a smoker})$

f. $P(F \cap \overline{H})$

	Nonsmoker	Light Smoker	Heavy Smoker	Total
Men	306	74	66	446
Women	345	68	81	494
Total	651	142	147	940

Consider the following events:

Event N: The person selected is a nonsmoker
Event L: The person selected is a light smoker
Event H: The person selected is a heavy smoker
Event M: The person selected is a male
Event F: The person selected is a female

68. Now suppose that two people are selected from the group, ***without replacement***. Let A be the event "the first person selected is a nonsmoker," and let B be the event "the second person is a light smoker." What is $P(A \cap B)$?

69. Two people are selected from the group, ***with replacement***. What is the probability that both people are nonsmokers?

70. Three cards are drawn, without replacement, from an ordinary deck. Find the probability of these events.

(a) Getting 3 aces

(b) Getting a 5, a queen, and a 10 in that order

(c) Getting a club, a diamond, and a heart in that order

71. An urn contains 4 red balls, 3 blue balls, and 7 white balls. A ball is selected and its color noted. Then it is replaced (or put back). A second ball is selected and its color noted. Find the probability of each of these events.

(a) Selecting 2 red balls

(b) P(Selecting 1 white ball and then 1 red ball)

72. An urn contains 4 red balls, 3 blue balls, and 7 white balls. A ball is selected and its color noted. It is **not** replaced (or put back). Then a second ball is selected and its color noted. Find the probability of each of these events.

(a) P(Selecting 2 red balls)

(b) P(Selecting 1 white ball and then 1 red ball)

73. It is reported that 16% of households regularly eat Krusty-O cereal. Choose 4 households at random. Find the probability that

(a) none regularly eat Krusty-O cereal

(b) all of them regularly eat Krusty-O cereal

(c) at least one regularly eats Krusty-O cereal

74. It is reported that 82% of LeastWorst Airline flights arrive on time. Choose 5 LeastWorst flights at random. Find the probability that

(a) none arrive on time

(b) all of them arrive on time

(c) at least one arrives on time

75. A medication is 75% effective against a bacterial infection. Find the probability that if 12 people take the medication, at least 1 persons infection will not improve.

76. How many 5-digit zip codes are possible if digits can be repeated?

77. How many 5-digit zip codes are possible if digits cannot be repeated?

78. How many ways can a baseball manager arrange a batting order of 9 players?

79. How many different ways can 7 different deans be seated in a row on a stage?

80. How many ways can a hiring committee of 7 math teachers be formed if the 7 are to be selected from a group of 22 full-time math teachers?

81. In a board of directors composed of 8 people, how many ways can one chief executive officer, one director, and one treasurer be selected?

Statistics Test 1
Tim Busken **Name:** KEY

1. ***Determine whether the given value is a statistic or a parameter.***

   (a) Burrito Bell fast food restaurant is interested in finding the average ages of it's employees. The company gets the age of each employee from its Human Resources department. Those ages are entered into a list and the mean age is found to be 34 years old. PARAMETER

   (b) Burrito Bell fast food restaurant is interested in finding the average ages of its employees. A random sample of 50 employee ages are taken from company records. Those ages are entered into a list and the mean age is found to be 34 years old. STATISTIC

   (c) Krusty-O Cereal Manufacturing is interested in finding the average number of boxes of Krusty-0s that each household of its customers consume each week. Surveys forms were printed on the back of each box and among those who responded, the average number of boxes consumed was 2.2. STATISTIC

   (d) 80% of the different kinds of elements on the Periodic table naturally occur in nature. The rest are man made (synthetic). PARAMETER

2. ***In each of these statements, tell whether descriptive or inferential statistics have been used.***

   (a) The report by the Medicare Office of the Actuary estimated that health spending will grow by an average of 5.8 percent a year through 2020, compared to 5.7 percent without the health overhaul. INFERENTIAL

   (b) Expenditures for the cable industry were \$5.66 billion in 1996 (Source: USA TODAY). DESCRIPTIVE

   (c) The mean travel time to work (years 2008–2012) for San Diego County workers age 16+ is 24.2 minutes. (Source: census.gov). DESCRIPTIVE

   (d) Allergy therapy makes bees go away (Source: Prevention). INFERENTIAL

3. ***Classify each variable as as discrete or continuous.***

   (a) Number of consumers in a poll who prefer Krusty-O cereal over Cheerios DISCRETE

   (b) Heights of buildings in New York City CONTINUOUS

   (c) The ages of Burger World employees CONTINUOUS

4. ***Classify each variable as qualitative or quantitative.***

   (a) A Gallup poll asked Americans, "Which subject (Math, English, History,...), if any, has been the most valuable in your life?" QUALITATIVE

   (b) The number of donuts sold each week at Dunkin Donuts QUANTITATIVE

   (c) The different flavors of donuts at Dunkin Donuts QUALITATIVE

   (d) Heights of NBA athletes QUANTITATIVE

5. ***Classify each variable as discrete or continuous.***

   (a) The number of people who buy a coffee from Starbucks today. DISCRETE

   (b) The volume of water (in cubic feet) of each of the Great Lakes. CONTINUOUS

   (c) The number of drones Amazon.com wants to have delivering orders by 2020. DISCRETE

   (d) The amount of time it takes to answer this question.CONTINUOUS

6. ***Classify each sample as a simple random sample, systematic, stratified, cluster or convenience.***

   (a) To determine his blood pressure, Stan divides up his day into three parts: morning, afternoon and evening. He then measures his blood pressure at 3 randomly selected times during each part of the day. STRATIFIED

   (b) To estimate the percentage of defects in a box of Krusty-O cereal, a quality control engineer inspects every 120^{th} box of Krusty-Os that is produced by the factory. SYSTEMATIC

   (c) To determine customer opinion of their inflight service (for all of its airplanes) on a particular day, LeastWorst Airlines selects 100 of its flights leaving San Diego International Airport and surveys all the participants of each flight.

   CLUSTER

   (d) To determine customer opinion of their inflight service (for all of its airplanes) on a particular day, LeastWorst Airlines randomly selects 2 people from every LeastWorst flight scheduled for that day to participate in their survey.

   STRATIFIED

   (e) A man experienced a tax audit. The tax department claimed that the man was audited because he was randomly selected from all men in his age group. STRATIFIED

   (f) A San Diego television station asks its viewers to call in their opinion regarding the weather programming. CONVENIENCE

7. ***Classify each sample as a simple random sample, systematic, stratified, cluster or convenience.***

   (a) Each employee of Nerdstroms Department Store has a 7-digit employee number. The district manager wants to find out how happy the employees who work there are with their jobs. She randomly selects 30 employee numbers and surveys the corresponding employees. SIMPLE RANDOM SAMPLE

   (b) A researcher interviews four guys who are sitting in a booth at Burger World. CONVENIENCE

   (c) To determine customer opinion of their inflight service for Flight 3568 from San Diego to Vancouver, Canada, LeastWorst Airlines selects every fifth person on Flight 3568 to participate in their survey. SYSTEMATIC

   (d) In order to find out which job title Burger World employees prefer to be called ("associate" or "server"), emails are sent inviting employees to click on a link to participate in an online survey. CONVENIENCE

   (e) The school needs 50 students to participate in a survey. A random number generator is used to select 50 student identification numbers. SIMPLE RANDOM SAMPLE

   (f) Mail carriers of a large city are divided into four groups according to gender (male or female) and according to whether they walk or ride on their routes. Then 10 are selected from each group and interviewed to determine whether they have been bitten by a dog in the last year. STRATIFIED

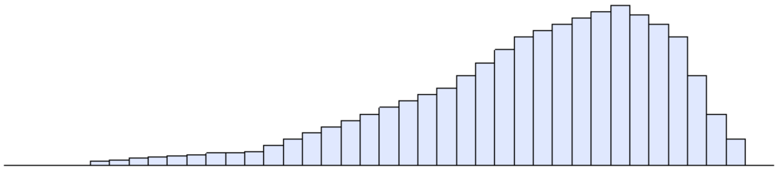
8. ***Identify each study as being an observational study or experimental.***

   (a) Subjects were randomly assigned to two groups, and one group was given an herb and the other group a placebo. After 6 months, the numbers of respiratory tract infections each group had were compared. EXPERIMENTAL

   (b) In a study sponsored by LeastWorst Airlines, 1200 customers were asked which food item from the breakfast menu they liked the most, and 32% of respondents said it was the Krusty-O's. OBSERVATIONAL

   (c) Subjects are randomly assigned to four groups. Each group is placed on one of four special diets — a low-fat diet, a high-fish diet, a combination of low-fat diet and high-fish diet, and a regular diet. After 6 months, the blood pressures of the groups are compared to see if diet has any effect on blood pressure. EXPERIMENTAL

   (d) Athletes who had suffered hamstring injuries were randomly assigned to one of two exercise programs. Those who engaged in static stretching returned to sports activity in a mean of 37.5 days (SD=27.2 days). Those assigned to a program of agility and trunk stabilization exercised returned to sports in a mean of 22.3 days (SD=8.4 days). EXPERIMENTAL



9. (1 point) Is the distribution in the histogram above skewed left, skewed right, or is it normal?

9. SKEWED LEFT

10. ***Find the class boundaries, midpoints and widths for the numeric class 23—30.***

    (a) class boundaries

    (b) class midpoint

    (c) class width

RULE OF THUMB: CLASS LIMITS SHOULD HAVE THE SAME NUMBER OF DECIMAL PLACES AS THE NUMBERS IN THE RAW DATA. CLASS BOUNDARIES WILL ALWAYS BE NUMBERS THAT OCCUPY ONE DECIMAL PLACE MORE THAN THE RAW DATA, AND END IN A 5.

(a) 22.5 — 30.5

(b) $\frac{23+30}{2} = 26.5$

(c) 8

CLASS WIDTH = DIFF BETWEEN THE UPPER AND LOWER CLASS BOUNDARIES OF ANY CLASS. SO, CLASS WIDTH$= 30.5 - 22.5 = 8$.

11. ***Find the class boundaries, midpoints and widths for the numeric class 3.2—5.2.***

    (a) class boundaries

    (b) class midpoint $= \frac{3.2+5.2}{2} = 4.2$

    (c) class width

(a) 3.15 — 5.25

(b) 4.2

(c) 2.1

CLASS WIDTH = DIFF BETWEEN THE UPPER AND LOWER CLASS BOUNDARIES OF ANY CLASS. SO, CLASS WIDTH$= 5.25 - 3.15 = 2.1$.

12. ***Find the class boundaries, midpoints and widths for the numeric class 0.25—0.30.***

    (a) class boundaries

    (b) class midpoint $= \frac{0.25+0.30}{2} = 0.275$

    (c) class width

(a) 0.245 — 0.305

(b) 0.275

(c) 0.06

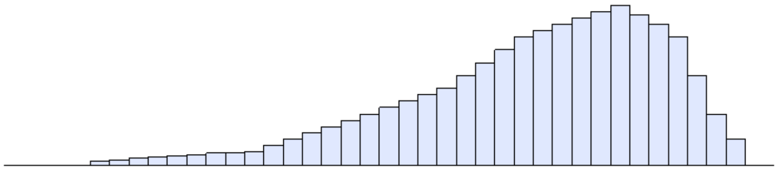
CLASS WIDTH = DIFF BETWEEN THE UPPER AND LOWER CLASS BOUNDARIES OF ANY CLASS. SO, CLASS WIDTH$= 0.305 - 2.45 = 0.06$.

13. List the original data from the stemplot in an ordered list sorted in a descending fashion. (Assume the leaf units are in the ones decimal place)

Stem	Leaves
22	2 4
23	6 8 9 9
24	1 4 8
25	1 3

25.3, 25.1, 24.8, 24.4, 24.1, 23.9, 23.9, 23.8, 23.6, 22.4, 22.2

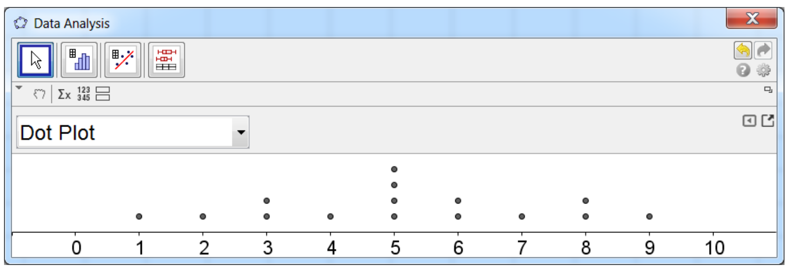
14. Sketch a histogram below for which the median would be ***greater*** than the mean.



mean median mode

15. ***Construct the dotplot for the given data.*** Attendance records for a preschool class with 15 students show the number of days each student was absent during the year. The days absent for each student were as follows.

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

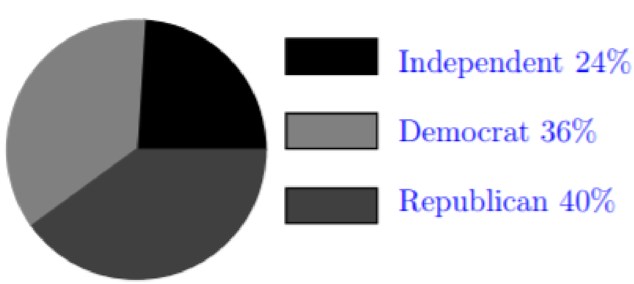


16. A questionnaire on political affiliation showed information obtained from 25 respondents. Construct a relative frequency distribution for the data (I = independant, D = democrat, R = republican) given below.

R D I R R I D D I R R D R D I D R D I D R I D R R

class	frequency	rel. freq.	pie chart angle
I	6	0.24	$0.24 \cdot 360° \approx 86°$
D	9	0.36	$0.36 \cdot 360° \approx 130°$
R	10	0.40	$0.40 \cdot 360° \approx 144°$

17. Construct a pie chart for the data given in the previous question.



18. The data shown (in millions of dollars) are the values of the 30 National Football League franchises. Construct a frequency distribution for the data using 8 classes.

170	191	171	235	173	187	181	191	211	186
200	218	243	200	182	320	184	239	197	204
186	199	186	210	209	240	204	193	188	242

We are told to use 8 classes. So, class width

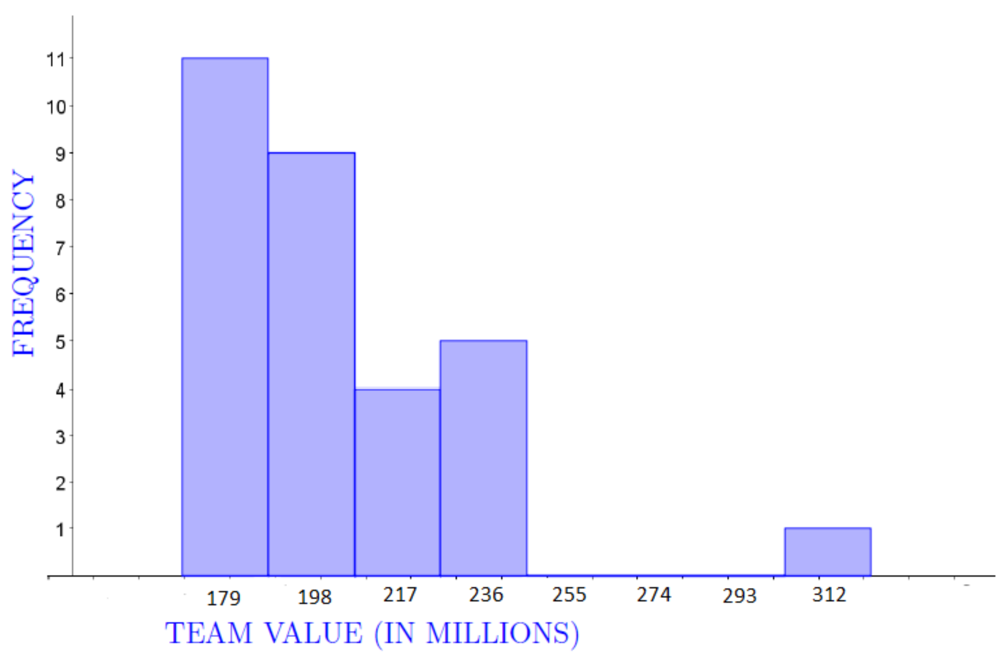
$$= \frac{\text{max} - \text{min}}{\text{the number of classes}} = \frac{320 - 170}{8} = 18.75$$

We always round this number up, so class width is 19 (or 20 might be more convenient).

We select a starting point for the lower class limit of the first class. We take the lowest value, 170, or a convenient number less than that. Add the width (19) to 170 to obtain the lower class limit of the next class. Keep adding until there are 8 classes. Subtract one from the lower limit of the 2nd class to get the upper class limit of the first class. Then add the class width to each upper limit to get all of the upper limits. Afterwards, fill out the frequency column of your table. Sort your data set first, then count the number of entries that belong to each class.

CLASS	FREQ.
170 – 188	11
189 – 207	9
208 – 226	4
227 – 245	5
246 – 264	0
265 – 283	0
284 – 302	0
303 – 321	1

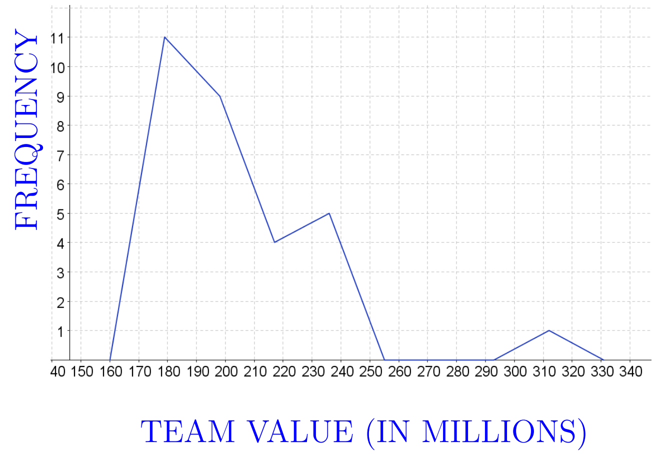
19. Construct a histogram for the data given in the previous question.



Your histogram bars should connect unless you have a class with a frequency that is zero. Histogram bars are centered at class midpoints. Histogram bars should connect at class boundaries. Your graph needs to have either midpoint values or boundary values labeled along the x axis.

20. Construct a frequency polygon for the data used in the previous two questions.

CLASS MIDPOINT	FREQ.
160	0
179	11
198	9
217	4
236	5
255	0
274	0
293	0
312	1
331	0



Plot the points (x, y) = (midpoint, frequency) for each class. Then connect the points with line segments. Afterwards, locate the left x-intercept by subtracting class width from the midpoint of your first class. Then locate the right x-intercept by adding class width to the midpoint of your last class.

21. ***Use the data to create a stemplot.*** The midterm test scores for the university's seventh-period graduate chemistry class are listed below.

92 73 82 95 77 84 65 67 78 88 56 75 86 72 92 79 77

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

Key : 3|1 means 31

22. Sample data is summarized in the frequency distribution below. Construct a cumulative frequency distribution.

class	frequency
2—6	13
7—11	64
12—16	49
17—21	23

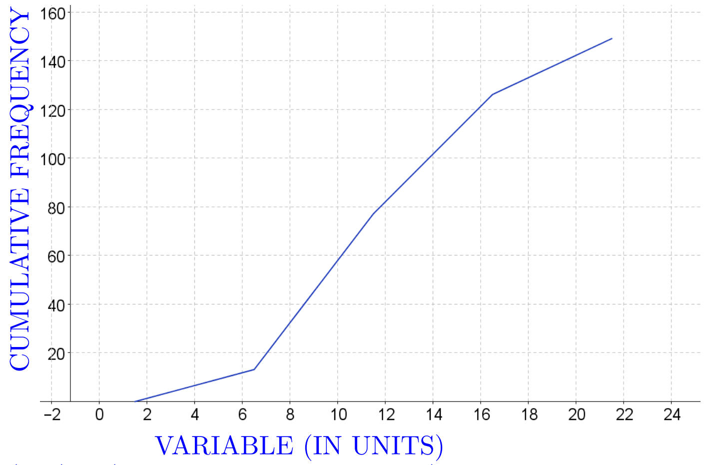
upper boundary	cumulative frequencies
less than 6.5	13
less than 11.5	77
less than 16.5	126
less than 21.5	149

23. How many measurements were given in the distribution table given in the previous question?

23. 149

The sum of the frequencies, $\sum f = 149$, is the number of measurements.

24. Make an ogive for the distribution used in the previous two questions.



Plot the points $(x, y) =$ (upper boundary, frequency) for each class. Then connect the points with line segments. Afterwards, locate the left x-intercept by subtracting class width (5) from the upper boundary of your first class ($6.5 - 5 = 1.5$). There is no right x-intercept on an ogive graph.

25. Sample data is summarized in the frequency distribution below. Find the mean of the distribution.

25. 11.8

Enter the class midpoints into your calculator's list 1 and the class frequencies in list 2. Then use the 1-var-stats command in the calculator (with both lists) to find the mean. Alternatively, you could use the formula (below)

class	frequency, f	class midpoint, X_m	$f \cdot X_m$
2—6	13	4	52
7—11	64	9	576
12—16	49	14	686
17—21	23	19	437

$$\bar{x} = \frac{\sum f \cdot X_m}{\sum f} = \frac{1751}{149} = 11.8$$

$$\sum f \cdot X_m = 1751$$

26. Sample data is summarized in the frequency distribution below. Find the standard deviation of the distribution.

26. $s = \sqrt{s^2} \doteq 4.3$

Enter the class midpoints into your calculator's list 1 and the class frequencies in list 2. Then use the 1-var-stats command in the calculator (with both lists) to find the standard deviation. Alternatively, you could use the formula (below)

class	f	X_m	$f \cdot X_m$	$f \cdot X_m^2$
2—6	13	4	52	208
7—11	64	9	576	5184
12—16	49	14	686	9604
17—21	23	19	437	8303

$$\sum f \cdot X_m = 1751$$

$$\sum f \cdot X_m^2 = 23299$$

$$s^2 = \frac{n \cdot \left(\sum f \cdot X_m^2\right) - \left(\sum f \cdot X_m\right)^2}{n(n-1)}$$

$$= \frac{149 \cdot (23{,}299) - 1751^2}{149(149-1)}$$

$$= \frac{405{,}550}{22{,}052}$$

$$\doteq 18.39062217$$

Use this census data to find the values of each parameter.

92 73 82 95 77 84 65 67 78 88 56 75 86 72 92 79 77

27. (1 point) mean — 27. 78.7

28. (1 point) median — 28. 78

29. (1 point) mode — 29. 77, 92

30. (1 point) standard deviation, $\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$ — 30. 10.1

31. (1 point) variance, σ^2 — 31. 102.1

32. (1 point) range $= \text{max} - \text{min} = 95 - 56 = 39$ — 32. 39

33. (1 point) minimum — 33. 56

34. (1 point) quartile 1 — 34. 72.5

35. (1 point) quartile 2 — 35. 78

36. (1 point) quartile 3 — 36. 87

37. (1 point) maximum — 37. 95

38. (2 points) What is the interquartile range? — 38. 14.5

    $\text{IQR} = Q3 - Q1 = 87 - 72.5 = 14.5$

39. (2 points) What two values provide the lower and upper fences for the data?

    $x_{min} = Q1 - 1.5 \cdot IQR = 72.5 - 1.5 \cdot 14.5 = 50.75$

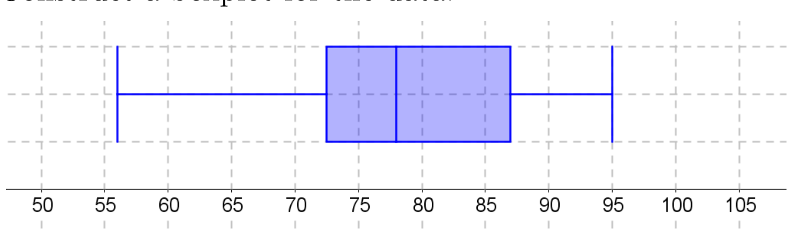
    $x_{max} = Q3 + 1.5 \cdot IQR = 87 + 1.5 \cdot 14.5 = 108.75$

    39. 50.75, 108.75

40. (1 point) Are there any outliers? Explain your reasoning to obtain full credit.

    There are no data values less than $x_{min} = 50.75$. There are no data values greater than $x_{max} = 108.75$. Therefore, there are no outliers in the census data.

41. (2 points) Construct a boxplot for the data.



Use the census data on the previous page to answer questions 43 and 44.

42. (2 points) 86 represents what percentile? 71st percentile

There are two formulas commonly used for percentile, see this note. I will use this formula:

$$\text{percentile of } 86 = \frac{(\text{number of values below } X)}{\text{total number of values}} \cdot 100\% = \frac{12}{17} \cdot 100\% \doteq 71\%$$

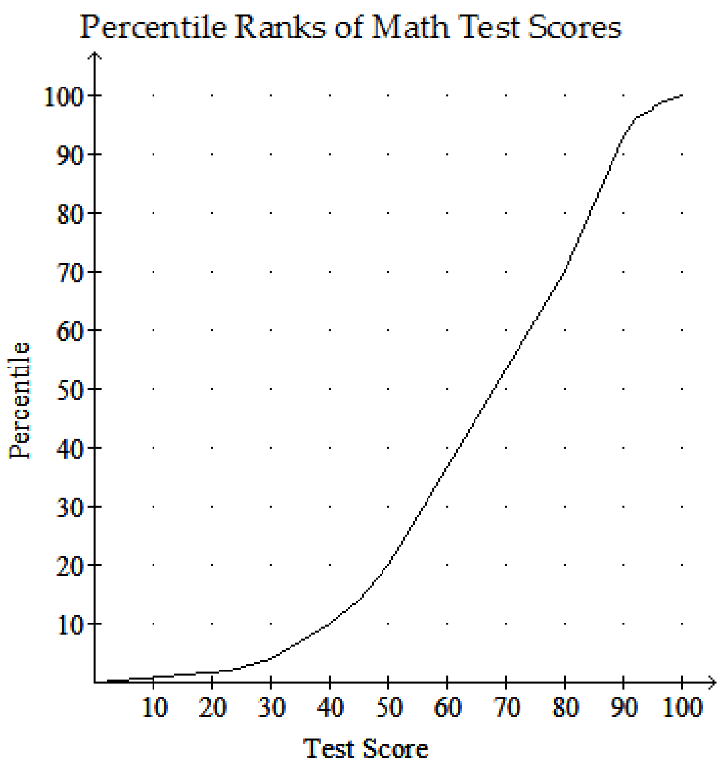
43. (2 points) 67 represents what percentile? 12th percentile

There are two formulas commonly used for percentile, see this note. I will use this formula:

$$\text{percentile of } 67 = \frac{(\text{number of values below } X)}{\text{total number of values}} \cdot 100\% = \frac{2}{17} \cdot 100\% \doteq 12\%$$

44. (2 points) The graph below is an ogive of scores on a math test. Use the graph to approximate the percentile rank of an individual whose test score is 70.

53rd percentile



45. The average labor charge for automobile mechanics is \$54 per hour. The standard deviation is \$4. Find the minimum percentage of data values that will fall within the range of \$48 to \$60. Use Chebyshev's theorem.

Solve $\bar{x} + k \cdot s = 60$ for k. Then evaluate $\left(1 - \frac{1}{k^2}\right) \cdot 100\%$ with k.

$\bar{x} + k \cdot s = 60$

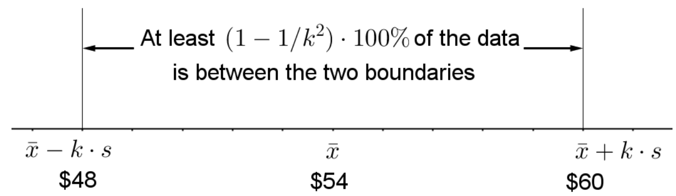
$54 + k \cdot (4) = 60$

$4k = 60 - 54$

$4k = 6$

$\frac{4k}{4} = \frac{6}{4}$

$\boxed{k=1.5}$



So, $1 - \frac{1}{k^2} = 1 - \frac{1}{1.5^2} \doteq 0.56 = \boxed{56\%}$

46. A survey of car rental agencies shows that the average cost of a car rental is \$0.32 per mile. The standard deviation is \$0.03. Using Chebyshev's theorem, find the range in which at least 75% of the data values will fall. $\boxed{[\$0.26, \$0.38]}$

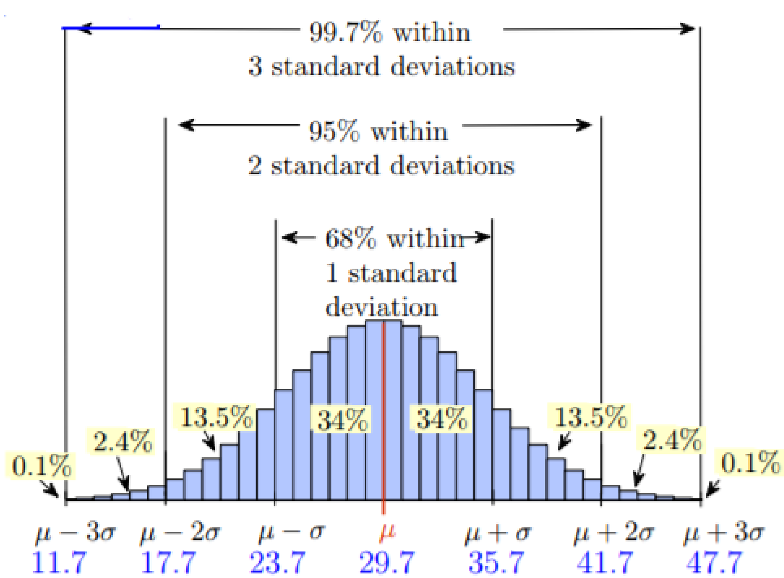
We know from the theorem that when $k = 2$ standard deviations, $1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 75\%$. This means that approximately 75% of the costs per mile will be within $k = 2$ standard deviation lengths away from the mean. Adding (subtracting) 2 standard deviations lengths ($2s = 2 \cdot \$0.03 = \$0.06$) from $\bar{x} = 0.32$ gives us the upper (lower) estimate for the middle 75%.

47. The mean of the times it takes a commuter to get to work in Baltimore is 29.7 minutes. Assume the distribution of commuter times is approximately bellshaped.

(a) If the standard deviation is 6 minutes, within what limits would you expect approximately 68% of the times to fall? $\boxed{[23.7min, 35.7min]}$

(b) Within what limits would you expect approximately 95% of the times to fall? $\boxed{[17.7min, 41.7min]}$

(c) Within what limits would you expect approximately 99.7% of the times to fall? $\boxed{[11.7min, 47.7min]}$



48. Identify the population and the sample. A survey of 1009 U.S. adults found that 26% think higher education is affordable for everyone who needs it.

Population: Collection of opinions of all adults in the United States
Sample: Collection of opinions of the 1009 U.S. adults surveyed

49. A television station claims that the amount of advertising per hour of broadcast time has an average of 16 minutes and a standard deviation equal to 2.9 minutes. You watch the station for 1 hour, at a randomly selected time, and carefully observe that the amount of advertising time is equal to 9 minutes. Calculate the z-score for this amount of advertising time. Round to the nearest hundredth.

$$z = \frac{x - \mu}{\sigma} = \frac{9 - 16}{2.9} \doteq -2.41$$

50. Using your z-score from the previous question, is 9 minutes of advertising time considered usual, unusual or very unusual. Explain your reasoning.

Since 9 minutes is more than 2 standard deviation lengths below the mean, it is an unusual amount of advertising time.

51. The mean age of lead actresses from the top ten grossing movies of 2010 was 29.6 years with a standard deviation of 6.35 years. Assume the distribution of the actresses' ages is approximately bell-shaped. In 2010, popular actress Jennifer Aniston was 41-years-old. What is Jennifer Aniston's age if it is standardized? Round to the nearest hundredth.

$$z = \frac{x - \mu}{\sigma} = \frac{41 - 29.6}{6.35} \doteq 1.80$$

52. Would it be usual, unusual or very unusual for a 41-year-old actress to be in a top-grossing film of 2010? Assume the Empirical Rule applies. Explain your reasoning.

Since an age of 41 years is within 2 standard deviation lengths from the mean, it is a usual age relative to the population of lead actresses. It would not be unusual for a 41-year-old actress to be in a top-grossing film of 2010.

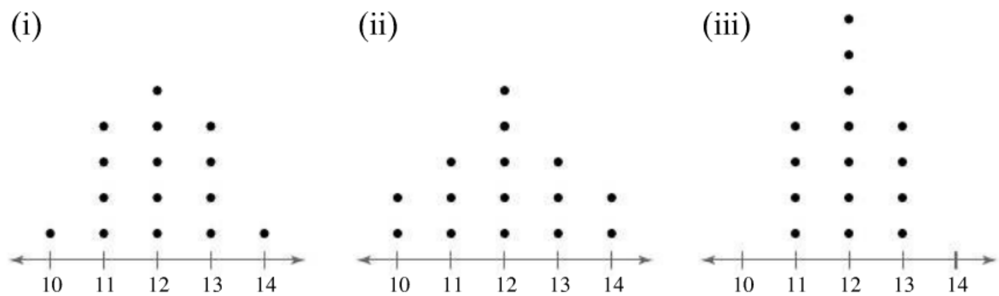
53. A television station claims that the amount of advertising per hour of broadcast time has an average of 17 minutes and a standard deviation equal to 1.1 minutes. You watch the station for 1 hour, at a randomly selected time, and carefully observe that the amount of advertising time is equal to 9 minutes. Calculate the z-score for this amount of advertising time.

$$z = \frac{x-\mu}{\sigma} = \frac{9-17}{1.1} \doteq -7.27$$

54. Using your z-score from the previous question, is 9 minutes of advertising time considered usual, unusual or very unusual. Explain your reasoning.

Since 9 minutes is more than 3 standard deviation lengths below the mean, it is a very unusual amount of advertising time.

55. Without calculating, which of the three data sets have the greatest standard deviation, which has the smallest standard deviation, and which has the second largest standard deviation?



Since all three graphs are symmetric, the mean is 12 for all three data sets.

Largest Standard Deviation: (ii) the middle graph has the most values farthest from the mean.

Smallest Standard Deviation: (iii) the third graph has the most values closest to the mean.

Second largest Standard Deviation: (i)

56. ***Classify each statement as an example of classical probability, empirical probability, or subjective probability.***

(a) An executive for the Krusty-O cereal factory makes an educated guess as to how well a new flavor of Krusty-Os will sell. SUBJECTIVE

(b) The probability of getting a 0 in roulette. CLASSICAL

(c) The probability that a randomly selected driver having a California drivers license will have brown eyes. EMPIRICAL

(d) In a recent year, among 135,933,000 registered passenger cars in the U.S., there were 10,427,000 crashes. So, the probability that a randomly selected passenger car in the U.S. will crash this year is $\frac{10{,}427{,}000}{135{,}933{,}000} \doteq 0.0767$. EMPIRICAL

57. In 2013, 32.3% of LeastWorst Airlines customers who purchased a ticket spent an additional \$20 to be in the first boarding group. Choose one LeastWorst customer at random. What is the probability that the customer *didn't* spend the additional \$20 to be in the first boarding group?

Define event A as the event "a randomly selected customer spent \$20 to be in the first boarding group." We are given $P(A) = 0.323$ and asked to find $P(\overline{A})$. Using the complement rule, $P(\overline{A}) = 1 - P(A) = 1 - 0.323 = \boxed{0.677}$.

58. ***True or False.*** Two events are mutually exclusive if they cannot both occur.

58. TRUE

59. ***Determine whether these events are mutually exclusive.***

(a) Roll a die: Roll an even number, and roll a 3. MUTUALLY EXCLUSIVE

(b) A single card is drawn from a standard deck: the card is a heart, and the card is an ace of clubs. MUTUALLY EXCLUSIVE

(c) Select a student in your class: the student is texting, the student is not a math major.NOT MUTUALLY EXCLUSIVE

(d) Select a pair of shoes: the shoes are brown colored, the shoes are made in the USA. NOT MUTUALLY EXCLUSIVE

60. A candy dish contains four red candies, seven yellow candies and fourteen blue candies. You close your eyes, choose two candies one at a time (without replacement) from the dish, and record their colors.

(a) Find the probability that both candies are red.
Use the multiplication rule for dependent events since the candies are selected without replacement. Let A be the event the first candy selected is red and B be the event the second candy selected is red. Then

$$P(\text{1st Candy Red AND 2nd Candy Red}) = P(A \text{ AND } B)$$
$$= P(A) \cdot P(B|A) = \frac{4}{25} \cdot \frac{3}{24} = \boxed{0.02}$$

(b) Find the probability that the first candy is red and the second candy is blue.
Use the multiplication rule for dependent events since the candies are selected without replacement. Let A be the event the first candy selected is red and B be the event the second candy selected is blue. Then

$$P(\text{1st Candy Red AND 2nd Candy Blue}) = P(A \text{ AND } B)$$
$$= P(A) \cdot P(B|A) = \frac{4}{25} \cdot \frac{14}{24} = \boxed{0.09\bar{3}}$$

61. Suppose that $P(A) = 0.2$ and $P(B) = 0.3$. If events A and B are independent, find $P(A \text{ and } B)$.
Use the multiplication rule for independent events:
$P(A \text{ AND } B) = P(A) \cdot P(B) = 0.2 \cdot 0.3 = \boxed{0.06}$

62. Suppose that $P(A) = 0.2$ and $P(B) = 0.3$. If events A and B are mutually exclusive, find $P(A \text{ or } B)$.

Use addition rule for mutually exclusive events:
$P(A \text{ OR } B) = P(A) + P(B) = 0.2 + 0.3 = \boxed{0.50}$

63. Toss a single, six-sided die three times. Find the probability that all three rolls are fives. Use the multiplication rule for independent events:

$$P(\text{1st roll a 5 AND 2nd roll a 5 AND 3rd roll a 5}) = P(A \text{ AND } B \text{ AND } C)$$
$$= P(A) \cdot P(B) \cdot P(C)$$
$$= \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \doteq \boxed{0.00463}$$

64. If 28% of U.S. medical degrees are conferred to women, find the probability that 3 randomly selected medical school graduates are men. Would you consider this event likely or unlikely to occur?

Define event A to be the event a randomly selected medical school graduate is a female. The complement of event A is a randomly selected medical school graduate is a male. Then $P(A) = 0.28$ and $P(\overline{A}) = 1 - P(A) = 1 - 0.28 = 0.72$. Now use multiplication rule for independent events (assume "with replacement" conditions). Then, $P(\text{1st person is M AND 2nd person is M AND 3rd person is M}) =$
$= P(A \text{ AND } B \text{ AND } C) = P(A) \cdot P(B) \cdot P(C) = 0.72 \cdot 0.72 \cdot 0.72 = (0.72)^3 \doteq \boxed{0.3732}$.
This probability is not unlikely since it is not less than 5%.

65. The human resources division at the Krusty-O cereal factory reports a breakdown of employees by job type and sex, summarized in the table below.

Job Type	Sex: Male	Female	total
Management	7	6	13
Supervision	8	12	20
Production	45	72	117
total	60	90	150

One of these workers is randomly selected.

(a) (2 points) Find the probability that the worker is a female.

$P(F) = \frac{90}{150} = \boxed{0.60}$

(b) (2 points) Find the probability that the worker is a female or a supervisor.

$P(F \text{ OR sup}) = P(F) + P(\text{sup}) - P(F \text{ AND sup}) = \frac{90}{150} + \frac{20}{150} - \frac{12}{150} = \boxed{0.65\overline{3}}$

(c) (2 points) Find the probability that the worker is male with the Supervision job type.

$P(M \text{ AND in Sup}) = \frac{8}{150} = \boxed{0.05\overline{3}}$

(d) (2 points) Find the probability that the worker is female, given that the person works in production.

$P(F|\text{prod}) = \frac{72}{117}$
$\doteq \boxed{0.6154}$

(e) (2 points) Find the probability that the worker works in production and is a female.

$P(\text{prod AND } F) = \frac{72}{150} = \boxed{0.48}$

(f) (2 points) Find the probability that the worker works in production or is a female.

$$P(\text{prod OR F}) = P(\text{prod}) + P(F) - P(\text{prod AND } F)$$

$$= \frac{117}{150} + \frac{90}{150} - \frac{72}{150} = \frac{135}{150} = \boxed{0.90}$$

66. Voter Support for political term limits is strong in many parts of the U.S. A poll conducted by the Field Institute in California showed support for term limits by a 2–1 margin. The results of this poll of $n = 347$ registered voters are given in the table.

	For (F)	**Against (A)**	**No Opinion (N)**	**Total**
Republican (R)	0.28	0.10	0.02	0.40
Democrat (D)	0.31	0.16	0.03	0.50
Other (O)	0.06	0.04	0.00	0.10
Total	0.65	0.30	0.05	1.00

If one individual in drawn at random from this group of 347 people, calculate the following probabilities:

(a) $P(N)$ (a) 0.05

(b) $P(D \text{ and } A)$ (b) 0.16

(c) $P(D \text{ or } A) = P(D) + P(A) - P(D \text{ and } A)$ (c) 0.64

$= 0.50 + 0.30 - 0.16 = 0.64$

(d) $P(A \text{ or } O) = P(A) + P(O) - P(A \text{ and } O)$ (d) 0.36

$= 0.30 + 0.10 - 0.04 = 0.36$

(e) $P(A \text{ and } O)$ (e) 0.04

(f) $P(\overline{N}) = 1 - P(N) = 1 - 0.05 = 0.95$ (f) 0.95

(g) $P(R|N) = \dfrac{0.02}{0.05} = 0.40$ (g) 0.40

(h) $P(A|D) = \dfrac{0.16}{0.50} = 0.32$ (h) 0.32

(i) $P(D|A) = \dfrac{0.16}{0.30} = 0.5\overline{3}$ (i) $0.5\overline{3}$

	Nonsmoker	Light Smoker	Heavy Smoker	Total
Men	306	74	66	446
Women	345	68	81	494
Total	651	142	147	940

Consider the following events:

Event N:	The person selected is a nonsmoker
Event L:	The person selected is a light smoker
Event H:	The person selected is a heavy smoker
Event M:	The person selected is a male
Event F:	The person selected is a female

67. Suppose one of the 940 subjects is chosen at random. Determine the following probabilities:

a. $P(N|F) = \dfrac{345}{494} \doteq \boxed{0.6984}$

b. $P(F|N) = \dfrac{345}{651} \doteq \boxed{0.5300}$

c. $P(H \cup M) = P(H) + P(M) - P(H \text{ and } M) = \dfrac{147}{940} + \dfrac{446}{940} - \dfrac{66}{940} = \dfrac{527}{940} \doteq \boxed{0.5606}$

d. $P(M \cap L) = \dfrac{74}{940} \doteq \boxed{0.0787}$

e. $P(\text{the person selected is a smoker}) = 1 - P(\text{the person selected is a nonsmoker})$

$$= 1 - \frac{651}{940} \doteq \boxed{0.3074}$$

f. $P(F \cap \overline{H}) = \dfrac{345 + 68}{940} = \dfrac{413}{940} \doteq \boxed{0.4394}$

68. Now suppose that two people are selected from the group, ***without replacement***. Let A be the event "the first person selected is a nonsmoker," and let B be the event "the second person is a light smoker." What is $P(A \cap B)$?

$P(A \cap B) = P(A) \cdot P(B|A) = \dfrac{651}{940} \cdot \dfrac{142}{939} \doteq \boxed{0.1047}$

69. Two people are selected from the group, ***with replacement***. What is the probability that both people are nonsmokers?

Let A be the event "the first person selected is a nonsmoker,"
and let B be the event "the second person is a nonsmoker."

Then $P(A \cap B) = P(A) \cdot P(B) = \dfrac{651}{940} \cdot \dfrac{651}{940} \doteq \boxed{0.4796}$

70. Three cards are drawn, without replacement, from an ordinary deck. Find the probability of these events. We use the multiplication rule for dependent events.

 (a) Getting 3 aces

$$
\begin{aligned}
&P(\text{3 aces}) = \\
&= P(\text{1st card Ace AND 2nd card ace AND 3rd card ace}) \\
&= P(\text{1st card Ace}) \cdot P(\text{2nd card ace} \mid \text{1st card Ace}) \cdot P(\text{3rd card ace} \mid \text{1st 2 cards Ace}) \\
&= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \doteq \boxed{0.0001809}
\end{aligned}
$$

 (b) Getting a 5, a queen, and a 10 in that order $= \frac{4}{52} \cdot \frac{4}{51} \cdot \frac{4}{50} \doteq \boxed{0.000483}$

 (c) Getting a club, a diamond, and a heart in that order $= \frac{13}{52} \cdot \frac{13}{51} \cdot \frac{13}{50} \doteq \boxed{0.0166}$

71. An urn contains 4 red balls, 3 blue balls, and 7 white balls. A ball is selected and its color noted. Then it is replaced (or put back). A second ball is selected and its color noted. Find the probability of each of these events. We use the multiplication rule for independent events.

 (a) Selecting 2 red balls

$$
\begin{aligned}
P(\text{2 reds}) &= P(\text{1st ball red AND 2nd ball red}) \\
&= P(\text{1st ball red}) \cdot P(\text{2nd ball red}) = \frac{4}{14} \cdot \frac{4}{14} \doteq \boxed{0.0816}
\end{aligned}
$$

 (b) P(Selecting 1 white ball and then 1 red ball) $= \frac{7}{14} \cdot \frac{4}{14} \doteq \boxed{0.1429}$

72. An urn contains 4 red balls, 3 blue balls, and 7 white balls. A ball is selected and its color noted. It is **not** replaced (or put back). Then a second ball is selected and its color noted. Find the probability of each of these events. We use the multiplication rule for dependent events.

 (a) P(Selecting 2 red balls)

$$
\begin{aligned}
&= P(\text{1st ball red AND 2nd ball red}) \\
&= P(\text{1st ball red}) \cdot P(\text{2nd ball red} \mid \text{1st ball red}) \\
&= \frac{4}{14} \cdot \frac{3}{13} \doteq \boxed{0.0659}
\end{aligned}
$$

 (b) P(Selecting 1 white ball and then 1 red ball)

$$
\begin{aligned}
&= P(\text{1st ball white AND 2nd ball red}) \\
&= P(\text{1st ball white}) \cdot P(\text{2nd ball red} \mid \text{1st ball white}) \\
&= \frac{7}{14} \cdot \frac{4}{13} \doteq \boxed{0.1538}
\end{aligned}
$$

73. It is reported that 16% of households regularly eat Krusty-O cereal. Choose 4 households at random. Find the probability that

(a) none regularly eat Krusty-O cereal

(b) all of them regularly eat Krusty-O cereal

(c) at least one regularly eats Krusty-O cereal

Let A be the event "a randomly selected household regularly eats Krusty-O cereal." Then $P(A) = 0.16$ and the complement of A (the event "a randomly selected household does not regularly eat Krusty-O cereal"), $P(\overline{A}) = 1 - P(A) = 1 - 0.16 = 0.84$.

(a) P(none regularly eat Krusty-O cereal)

= P(1st does not AND 2nd does not AND 3rd does not AND 4th does not)

$= (0.84) \cdot (0.84) \cdot (0.84) \cdot (0.84) = (0.84)^4 = \boxed{0.4979}$

(b) P(all 4 of them regularly eat Krusty-O cereal)

= P(1st does AND 2nd does AND 3rd does AND 4th does)

$= (0.16) \cdot (0.16) \cdot (0.16) \cdot (0.16) = (0.16)^4 = \boxed{0.000655}$

(c) P(at least one regularly eats Krusty-O cereal)

= 1 − P(none regularly eat Krusty-O cereal) = $1 - 0.4979 = \boxed{0.5021}$.

74. It is reported that 82% of LeastWorst Airline flights arrive on time. Choose 5 LeastWorst flights at random. Find the probability that

(a) none arrive on time

(b) all of them arrive on time

(c) at least one arrives on time

Let A be the event "a randomly selected LeastWorst Airline flight arrives on time." Then $P(A) = 0.82$ and the complement of A (the event "a randomly selected LeastWorst Airline flight does not arrive on time") is $P(\overline{A}) = 1 - P(A) = 1 - 0.82 = 0.18$.

(a) P(none arrive on time)

= P(1st late AND 2nd late AND 3rd late AND 4th late AND 5th late)

$= (0.18) \cdot (0.18) \cdot (0.18) \cdot (0.18) \cdot (0.18) = (0.18)^5 = \boxed{0.000189}$

(b) P(all 5 of them arrive on time)

= P(1st on time AND 2nd on time AND 3rd on time AND 4th on time AND 5th on time)

$= (0.82) \cdot (0.82) \cdot (0.82) \cdot (0.82) \cdot (0.82) = (0.82)^5 \doteq \boxed{0.3707}$

(c) P(at least one arrives on time)

= 1 − P(none arrive on time) = $1 - 0.000189 = \boxed{0.9998}$.

75. A medication is 75% effective against a bacterial infection. Find the probability that if 12 people take the medication, at least 1 person's infection will not improve.

P(at least one will not improve)=1 − P(none will not improve)

$= 1 - \text{P(all will improve)} = 1 - (0.75)^{12} \doteq \boxed{0.9683}$

76. How many 5-digit zip codes are possible if digits can be repeated?

Each of the five digits has can take on one of ten digits from zero through nine. Using the fundamental (multiplication) rule there are $\underline{10} \cdot \underline{10} \cdot \underline{10} \cdot \underline{10} \cdot \underline{10} = \boxed{100{,}000}$ different possible zip codes.

77. How many 5-digit zip codes are possible if digits cannot be repeated?

The first of the five digits can take on one of ten digits from zero through nine, the second of the five digits can take on one of nine digits, the third of the five digits can take on one of eight digits, and so on. Using the fundamental (multiplication) rule there are $\underline{10} \cdot \underline{9} \cdot \underline{8} \cdot \underline{7} \cdot \underline{6} = \boxed{30{,}240}$ different possible zip codes.

78. How many ways can a baseball manager arrange a batting order of 9 players?

The manager has 9 options for the first batter, 8 options for the 2nd batter, 7 options for the third batter, and so on. Using the fundamental (multiplication) rule there are $9! = \underline{9} \cdot \underline{8} \cdot \underline{7} \cdot \underline{6} \cdot \underline{5} \cdot \underline{4} \cdot \underline{3} \cdot \underline{2} \cdot \underline{1} = \boxed{362{,}880}$ different batting orders.

79. How many different ways can 7 different deans be seated in a row on a stage?

The first chair can seat one of 7 deans, the second chair can seat one of six deans, the third chair can seat one of five deans, and so on. Using the fundamental (multiplication) rule there are $7! = \underline{7} \cdot \underline{6} \cdot \underline{5} \cdot \underline{4} \cdot \underline{3} \cdot \underline{2} \cdot \underline{1} = \boxed{5040}$ different seating orders.

80. How many ways can a hiring committee of 7 math teachers be formed if the 7 are to be selected from a group of 22 full-time math teachers? Since the ordering of the members on the committee does not matter we use the combination rule.

$$_nC_r = {}_{22}C_7 = \frac{22!}{(22-7)! \cdot 7!} = \frac{22!}{15! \cdot 7!} = \boxed{170{,}544}$$

81. In a board of directors composed of 8 people, how many ways can one chief executive officer, one director, and one treasurer be selected?

Since the ordering of the arrangements matter we use the permutation rule.

$$_nP_r = {}_8P_3 = \frac{8!}{(8-3)!} = \frac{8 \cdot 7 \cdot 6 \cdot 5!}{5!} = \boxed{336}$$

Statistics Practice Test 2
(Chs. 9, 4 and 5) Name: ____________________

1. Baiers Electronics manufactures computer parts that are supplied to many computer companies. Despite the fact that two quality control inspectors at Baiers Electronics check every part for defects before it is shipped to another company, a few defective parts do pass through these inspections undetected. Let x denote the number of defective computer parts in a shipment. The following table gives the probability distribution of x. Use the table to answer parts a through g.

x	0	1	2	3	4	5	6	7	8	9
$p(x)$	.02	.20	.25	.20	.10	.08	.06	.04	.03	.02

   (a) $P(X = 5)$
   (b) $P(3)$
   (c) $P(X \leq 3)$
   (d) $P(X \geq 1)$
   (e) $P(X < 8)$
   (f) $P(X > 3)$
   (g) $P(2 \leq X \leq 4)$

2. The probabilities that a bakery has a demand for 2, 3, 5, or 7 birthday cakes on any given day are 0.35, 0.41, 0.15, and 0.09, respectively. *Construct a probability distribution for the data and draw a graph for the distribution.*

3. How many cakes can the bakery expect to sell (on average) on any given day?

4. A lottery offers one \$1000 prize, one \$500 prize, and five \$100 prizes. One thousand tickets are sold at \$3 each. Find the expected net gain if a person buys one ticket.

5. A bank vice president feels that each savings account customer has, on average, three credit cards. The following probability distribution represents the number of credit cards people own. Find the mean, variance, and standard deviation. Is the vice president correct?

Number of cards X	0	1	2	3	4
Probability $P(X)$	0.18	0.44	0.27	0.08	0.03

6. The local animal shelter adopts out cats and dogs each week with the following probabilities listed in the table below. How many animals each week would you expect the shelter to adopt out?

Number of animals adopted each week, X	5	6	7	8	9	10
$P(X)$	0.15	0.3	0.25	0.18	0.1	0.02

7. Assume that a procedure yields a binomial probability model with a trial repeated $n = 5$ times. Suppose the probability of success on a single trial is $p = 0.47$. Then X counts the number of successes among 5 trials. Describe the probability distribution by filling out the table below. Round calculations to five decimal places. In addition, graph the distribution.

x	$P(X = x)$
0	
1	
2	
3	
4	
5	

8. Find the mean, variance and standard deviation of the binomial probability distribution in the last question.

9. Let X be a binomial random variable with $n = 12$ and $p = 0.3$. Find the following probabilities:
   (a) $P(X = 5)$
   (b) $P(X = 8)$
   (c) $P(X \leq 3)$
   (d) $P(X \leq 5)$
   (e) $P(X < 2)$
   (f) $P(X > 9)$
   (g) $P(X \geq 4)$
   (h) $P(3 \leq X \leq 8)$
   (i) $P(6 < X < 10)$

10. Assume that 13% of people are left-handed. If we select 42 students at random, and ask the students if they are left-handed, then we are conducting a binomial experiment, where X counts the number of left-handed people in the group of 42. Find the probability of each outcome described below.
    (a) There is at least one lefty in the group
    (b) There are exactly 3 lefties in the group
    (c) There are not more than 3 lefties in the group
    (d) How many left-handed people would you expect there to be in a group of 42 students?

11. People with type O-negative blood are said to be "universal donors." About 7% of the U.S. population has this blood type. Suppose that 335 people show up at a blood drive, and ask the 335 people if they are type O-negative, then we are conducting a binomial experiment, where X counts the number of "universal donors." in the group of 335. Find the probability of each outcome described below.
    (a) What is the expected number of universal donors in the group?
    (b) What is the probability that exactly 30 universal donors are in the group?
    (c) What is the probability that more than 3 universal donors are in the group?

12. Nearly 40 percent of working-aged Americans now hold a college degree, according to a new report from the Lumina Foundation. Suppose a group of 32 Americans are randomly selected.

(a) What is the expected number of college graduates in the group?

(b) What is the standard deviation of the number of college graduates in the group?

(c) What is the probability there are 2 college graduates in the group?

(d) What is the probability there are 2 or fewer college graduates in the group?

13. The percentage of Americans who use the Internet grew to 84 percent from 52 percent at the turn of the century, according to data from the Pew Research Center. (source: *NY Times*, July 28, 2015) A random sample of 75 Americans have participated in a survey.

(a) What is the expected number of Americans in the sample that do not use the internet?

(b) What is the variance of the number of Americans that do not use the internet?

14. *Find the area under the standard normal distribution curve for each.*

(a) Between $z = -0.19$ and $z = 1.23$

(b) To the left of $z = -1.56$

(c) To the right of $z = -0.38$

15. *Find each probability using the standard normal distribution curve for each.*

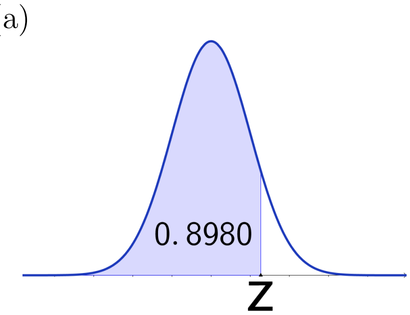
(a) $P(-0.09 < z < 2.42)$

(b) $P(z > -1.68)$

(c) $P(z < 0.23)$

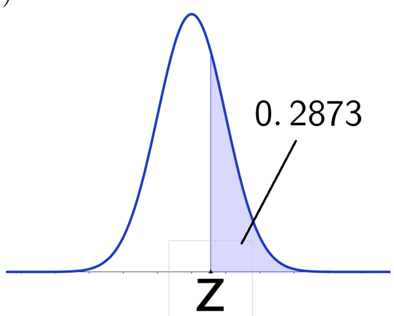
16. *Find the indicated z score. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.*

(a)



(a) ____________

(b)



(b) ____________

17. ***Use the Standard Normal Distribution to find the requested z value.***

(a) Find the z that separates the lower 88% of z scores from the top 12%

(b) Find the value of z that represents the 92nd percentile.

18. Find the value of z that represents the 45th percentile.

19. The waiting time in line at a Starschmuchs Coffee is normally distributed with a mean of 3.2 minutes and a standard deviation of 1.3 minutes. Find the probability that a randomly selected customer has to wait

(a) Less than 1 minute.

(b) more than 2 minutes.

(c) between 0.75 minutes and 2 minutes.

20. The average yearly precipitation in San Diego is 9.62 inches with a standard deviation of 4.42 inches and precipitation amounts are normally distributed.

(a) Find the probability that a randomly selected year will have precipitation greater than 12 inches.

(b) Find the probability that five randomly selected years will have an average precipitation greater than 8 inches.

(c) Find the precipitation amount from the distribution of precipitations that represents the 75th percentile.

21. Some passengers died when a water taxi sank in Baltimore's inner harbor. Men are typically heavier than women and children, so when loading a water taxi, let's assume a worst-case scenario in which all passengers are men. Based on data from the National Health and Nutrition Survey, assume that weights of men are normally distributed with a mean of 172 lb. and a standard deviation of 29 lb.

(a) Find the probability that if an individual man is randomly selected, his weight will be greater than 175 lb.

(b) Find the probability that 20 men will have a mean weight that is greater than 175 lb. (so that their total weight exceeds the safe capacity of 3500 lb.

22. The average per capita spending on health care in the United States is \$5274. The standard deviation is \$600 and the distribution of health care spending is approximately normal. Find the limits of the middle 50% of individual health care expenditures.

23. A prestigious college decides to only take applications from students who have scored in the top 5% on the SAT test. The SAT scores are approximately normally distributed with a mean of 490 and a standard deviation of 70. Find the score that is necessary to obtain in order to qualify for applying to this college.

24. Americans ate an average of 25.7 pounds of Krusty-O Cereal each last year and spent an average of \$61.50 per person doing so. If the standard deviation for consumption is 3.75 pounds and the standard deviation for the amount spent is \$5.89, find the following:

(a) The probability that the sample mean Krusty-O cereal consumption for a random sample of 40 American consumers exceeded 27 pounds.

(b) The probability that for a random sample of 50, the the average yearly amount spent on Krusty-O Cereal was between \$60.00 and \$100.

25. Use the information in the table below to answer parts a through f. The ages (in years) of 10 men and their systolic blood pressures (in millimeters of mercury) are listed in the table.

Age, x	16	25	39	45	49	64	70	29	57	22
Systolic Blood Pressure, y	109	122	143	132	199	185	199	130	175	118

(a) What is the sample correlation coefficient, r?

(b) Describe the type of correlation and interpret the correlation in the context of the data.

(c) Find the equation of the regression line for the data.

(d) Use the regression equation to predict the value of y for $x = 42$.

(e) Use the regression equation to predict the blood pressure for a man aged 67.

Statistics Practice Test 2
(Chs. 9, 4 and 5) **Name:** KEY

1. Baiers Electronics manufactures computer parts that are supplied to many computer companies. Despite the fact that two quality control inspectors at Baiers Electronics check every part for defects before it is shipped to another company, a few defective parts do pass through these inspections undetected. Let x denote the number of defective computer parts in a shipment. The following table gives the probability distribution of x. Use the table to answer parts a through g.

x	0	1	2	3	4	5	6	7	8	9
$p(x)$	.02	.20	.25	.20	.10	.08	.06	.04	.03	.02

(a) $P(X = 5) = \boxed{0.08}$

(b) $P(3) = \boxed{0.20}$

(c) $P(X \leq 3) = P(0) + P(1) + P(2) + P(3) = 0.02 + 0.20 + 0.25 + 0.20 = \boxed{0.67}$

(d)
$$\begin{aligned} P(X \geq 1) &= P(1) + P(2) + \cdots + P(9) \\ &= 1 - P(0) \\ &= 1 - 0.02 \\ &= \boxed{0.98} \end{aligned}$$

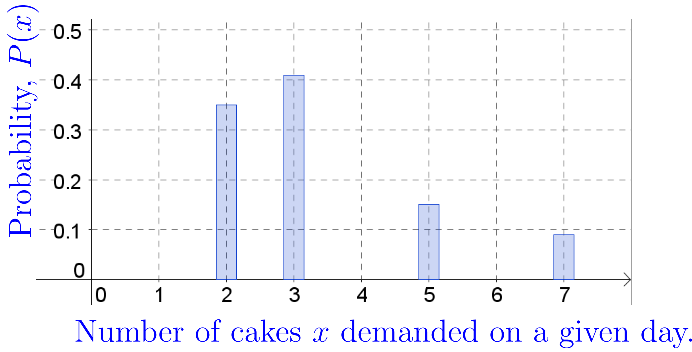
(e)
$$\begin{aligned} P(X < 8) &= P(0) + P(1) + P(2) + \cdots + P(7) \\ &= 1 - [P(8) + P(9)] \\ &= 1 - [0.03 + 0.02] \\ &= \boxed{0.95} \end{aligned}$$

(f)
$$\begin{aligned} P(X > 3) &= P(4) + P(5) + \cdots + P(9) \\ &= 1 - P(x \leq 3) \\ &= 1 - 0.67 \\ &= \boxed{0.33} \end{aligned}$$

(g) $P(2 \leq X \leq 4) = P(2) + P(3) + P(4) = 0.25 + 0.20 + 0.10 = \boxed{0.55}$

2. The probabilities that a bakery has a demand for 2, 3, 5, or 7 birthday cakes on any given day are 0.35, 0.41, 0.15, and 0.09, respectively. *Construct a probability distribution for the data and draw a graph for the distribution.*

x	$P(x)$
2	0.35
3	0.41
5	0.15
7	0.09



3. How many cakes can the bakery expect to sell (on average) on any given day?

We know that the expected number of cakes, $E(x)$, is just the mean, μ, of the random variable x, given by the formula $E(x) = \sum x \cdot P(x)$. We can use the calculator command 1-var-stats(L1,L2) with L1 and L2 being the x and $P(x)$ lists. (For some calculators this means your frequency list should be set to L2). After running 1-var-stats the calculator gives you 3.3 as the value for the average ($\bar{x}$). Therefore, the bakery can expect to sell $\boxed{\text{3.3 cakes}}$ on any given day.

Alternative Solution Path:
You could use the formula for the mean: $\mu = \sum x \cdot P(x)$, and the following tabular method of evaluating the formula. To write the solution this way you need to make a third column labeled $x \cdot P(x)$. To fill out that column we multiply straight across the rows. Afterwards we sum the numbers in the third column to get the mean.

x	$P(x)$	$x \cdot P(x)$
2	0.35	0.70
3	0.41	1.23
5	0.15	0.75
7	0.09	0.63
		3.3

$$E(x) = \sum x \cdot P(x) = 0.70 + 1.23 + 0.75 + 0.63 = \boxed{\text{3.3 cakes}}$$

4. A lottery offers one \$1000 prize, one \$500 prize, and five \$100 prizes. One thousand tickets are sold at \$3 each. Find the expected net gain if a person buys one ticket.

Let x represent the amount gained or lost purchasing a single ticket. The probability distribution of x is in the table below.

event	x	$P(x)$
grand prize	\$997	$1/1000 = 0.001$
first prize	\$497	$1/1000 = 0.001$
second prize	\$97	$5/1000 = 0.005$
first prize	$-\$3$	$993/1000 = 0.993$

We know that the expected net gain, $E(x)$, is just the mean, μ, of the random variable x, given by the formula $E(x) = \sum x \cdot P(x)$. We can use the calculator command 1-var-stats(L1,L2) with L1 and L2 being the x and $P(x)$ lists. The average ($\bar{x}$) from running 1-var-stats will give you $\bar{x} =$ $\boxed{\text{-\$1.00}}$.

Alternative Solution Path: you could use the formula for the mean: $\mu = \sum x \cdot P(x)$, and the tabular method (below) of evaluating the formula. To write the solution this way you need to make a fourth column in your table, and label it '$x \cdot P(x)$' (table below). To fill out that column we multiply straight across the rows. Afterwards we sum the numbers in the fourth column to get the mean.

event	x	$P(x)$	$x \cdot P(x)$
grand prize	\$997	$1/1000 = 0.001$	\$.997
first prize	\$497	$1/1000 = 0.001$	\$.497
second prize	\$97	$5/1000 = 0.005$	\$.485
first prize	−\$3	$993/1000 = 0.993$	−\$2.979
			−\$1.00

5. A bank vice president feels that each savings account customer has, on average, three credit cards. The following probability distribution represents the number of credit cards people own. Find the mean, variance, and standard deviation. Is the vice president correct?

Number of cards X	0	1	2	3	4
Probability $P(X)$	0.18	0.44	0.27	0.08	0.03

Put the distribution in your calculator's list environment and then call on

$$1 - var - stats \quad L1,\ L2$$

on the calculator to obtain values for μ and σ. Then square σ to get the variance.

mean, $\mu = 1.34 \doteq \boxed{1.3}$ credit cards.
The vice president's estimate for the average number of credit cards is incorrect.
standard deviation, $\sigma = 0.9615 \doteq \boxed{1.0}$ credit cards
variance, $\sigma^2 = 0.9244 \doteq \boxed{0.9}$ credit cards squared

---

Alternative Solution Path:

x L1	$P(x)$ L2	$x \cdot P(x)$ L1·L2	$x^2 \cdot P(x)$ L1·L3
0	0.18	0	0
1	0.44	0.44	0.44
2	0.27	0.54	1.08
3	0.08	0.24	0.72
4	0.03	0.12	0.48
totals		1.34	2.72

You could evaluate the formulas for the mean and variance. To do the problem this way you need to add two more columns to your table, labeled '$x \cdot P(x)$' and '$x^2 \cdot P(x)$.' To get the numbers in the third column, we multiply each x by its $P(x)$. We then sum the numbers in the third column to get the mean.

To get the numbers in the fourth column, we multiply the square of each x by its $P(x)$. To get the variance, we then sum the numbers in the fourth column to get 2.72. Afterwards we subtract the value of the mean squared from this sum.

mean, $\mu = \sum x \cdot P(x) = 1.34 \doteq \boxed{1.3}$ credit cards.
The vice president's estimate for the average number of credit cards is incorrect.
variance, $\sigma^2 = \sum x^2 \cdot P(x) - \mu^2 = 2.72 - 1.34^2 = 0.9244 \doteq \boxed{0.9}$ credit cards squared
standard deviation, $\sigma = \sqrt{\sigma^2} \doteq 0.9615 \approx \boxed{1.0}$ credit cards

6. The local animal shelter adopts out cats and dogs each week with the following probabilities listed in the table below. How many animals each week would you expect the shelter to adopt out?

Number of animals adopted each week, X	5	6	7	8	9	10
$P(X)$	0.15	0.3	0.25	0.18	0.1	0.02

The expected number of animals is the mean of X, which is 6.84 animals.

7. Assume that a procedure yields a binomial probability model with a trial repeated $n = 5$ times. Suppose the probability of success on a single trial is $p = 0.47$. Then X counts the number of successes among 5 trials. Describe the probability distribution by filling out the table below. Round calculations to five decimal places.

x	$P(X = x)$
0	0.04182
1	0.18543
2	0.32887
3	0.29164
4	0.12931
5	0.02293

Step 1 $seq(X, X, 0, 5) \to L1$ Put the numbers 0 through 5 in L1 with the "seq" (sequence) command. To access the "seq" command, press [2nd] [STAT], then arrow over to the OPS drop down menu. The sequence command is entry number 5.

Step 2 $binomPdf(5, 0.47) \to L2$ Use the calculator's binomPdf command with the two input arguments, $n = 5$ and $p = 0.47$ to store the six probabilities in your calculator's list two.

8. Find the mean, variance and standard deviation of the binomial probability distribution in the last question.

mean, $\mu = n \cdot p = 5 \cdot (0.47) = 2.35 \doteq \boxed{2.4}$ and

variance, $\sigma^2 = n \cdot p \cdot q = 2.35 \cdot (0.53) = 1.2455 \doteq \boxed{1.2}$ and

standard deviation, $\sigma = \sqrt{\sigma^2} \doteq \boxed{1.1}$.

9. Let X be a binomial random variable with $n = 12$ and $p = 0.3$. Find the following:

We obtain the binomial probability distribution table for this experiment using either the calculator or the binomial tables. The table below came from the binomial tables. Each probability in the table is approximated to the ten-thousandths place-value column. You should get the same answers for parts a through i even if you use the calculator and not the binomial tables.

x	0	1	2	3	4	5	6
$P(x)$	0.0138	0.0712	0.1678	0.2397	0.2311	0.1585	0.0792

x	7	8	9	10	11	12
$P(x)$	0.0291	0.0078	0.0015	0.0002	0.0000	0.0000

(a) $P(X = 5) \doteq \boxed{0.1585}$

(b) $P(X = 8) \doteq \boxed{0.0078}$

(c) $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$$\doteq 0.0138 + 0.0712 + 0.1678 + 0.2397 = \boxed{0.4925}$$

(d) $P(X \leq 5)$

$$= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$

$$\doteq 0.0138 + 0.0712 + 0.1678 + 0.2397 + 0.2311 + 0.1585 \doteq \boxed{0.8822}$$

(e) $P(X < 2) = P(X = 0) + P(X = 1) \doteq 0.0138 + 0.0712 = \boxed{0.0850}$

(f) $P(X > 9) = P(X = 10) + P(X = 11) + P(X = 12)$

$$= 0.0002 + 0.0000 + 0.0000 \doteq \boxed{0.0002}$$

(g) $P(X \geq 4) = P(X = 4) + P(X = 5) + \cdots + P(X = 11) + P(X = 12)$

using the complement rule, this is also equal to

$$= 1 - \Big[P(0) + P(1) + P(2) + P(3)\Big]$$

$$= 1 - \Big[0.0138 + 0.0712 + 0.1678 + 0.2397\Big]$$

$$= 1 - 0.4925 \doteq \boxed{0.5075}$$

(h) $P(3 \leq X \leq 8) = P(X=3) + P(X=4) + \cdots + P(X=8)$

$$= 0.2397 + 0.2311 + 0.1585 + 0.0792 + 0.0291 + 0.0078 \doteq \boxed{0.7455}$$

(i) $P(6 < X < 10) = P(X=7) + P(X=8) + P(X=9)$

$$= 0.0291 + 0.0078 + 0.0015 = \boxed{0.0384}$$

10. Assume that 13% of people are left-handed. If we select 42 students at random, and ask the students if they are left-handed, then we are conducting a binomial experiment, where X counts the number of left-handed people in the group of 42. Find the probability of each outcome described below.

(a) There is at least one lefty in the group

Let X = the number of people in the sample of $n = 42$ who are left-handed. Then X follows a binomial probability distribution with $n = 42$ and $p = 0.13$.

$$P(X \geq 1) = P(X=1) + P(X=2) + \cdots + P(X=42)$$

using the complement rule, this is also equal to

$$= 1 - P(X=0) = 1 - 0.0029 = \boxed{0.9971}$$

(b) There are exactly 3 lefties in the group

$$P(X=3) = \boxed{0.1104}$$

(c) There are not more than 3 lefties in the group

$$P(X \ngtr 3) = P(X \leq 3)$$

$$= P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

$$= 0.00288 + 0.01809 + 0.05542 + 0.11041 = \boxed{0.1868}$$

(d) How many left-handed people would you expect there to be in a group of 42 students?

$$E(x) = \mu = n \cdot p = 42 \cdot (0.13) = 5.46 \text{ people}$$

11. People with type O-negative blood are said to be "universal donors." About 7% of the U.S. population has this blood type. Suppose that 335 people show up at a blood drive, and ask the 335 people if they are type O-negative, then we are conducting a binomial experiment, where X counts the number of "universal donors." in the group of 335. Find the probability of each outcome described below.

(a) What is the expected number of universal donors in the group?

Let X represent the number of universal donors in the sample of 335 blood donors. Then X follows a binomial probability distribution with $n = 335$ and $p = 0.07$. $E(X) = \mu = n \cdot p = 335 \cdot 0.07 \doteq \boxed{23.5 \text{ universal donors}}$

(b) What is the probability that exactly 30 universal donors are in the group?

$P(X = 30) = \boxed{0.0307}$

(c) What is the probability that more than 3 universal donors are in the group?

$$P(X > 3) = P(X = 4) + P(X = 5) + \cdots + P(X = 335)$$

using the complement rule, this is also equal to

$$= 1 - P(X \leq 3) = 1 - \Big[P(0) + P(1) + P(2) + P(3)\Big]$$

$$= 1 - \Big[0.0138 + 0.0712 + 0.1678 + 0.2397\Big]$$

$$= 1 - 0.4925 \doteq \boxed{0.5075}$$

12. Nearly 40 percent of working-aged Americans now hold a college degree, according to a new report from the Lumina Foundation. Suppose a group of 32 Americans are randomly selected.

(a) What is the expected number of college graduates in the group?

Let X represent the number of college graduates in the sample of 32 Americans. Then X follows a binomial probability distribution with $n = 32$ and $p = 0.4$. $E(X) = \mu = n \cdot p = 32 \cdot 0.4 = \boxed{12.8 \text{ college graduates}}$

(b) What is the standard deviation of the number of college graduates in the group?

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{32 \cdot 0.4 \cdot 0.6} = \sqrt{7.68} \doteq \boxed{2.8 \text{ college graduates}}$

(c) What is the probability there are 2 college graduates in the group?

$P(X = 2) = 1.7544E - 5 \doteq \boxed{0.00002}$

(d) What is the probability there are 2 or fewer college graduates in the group?

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 8e - 8 + 1.7e - 6 + 1.8e - 5 = 1.978e - 5 \doteq \boxed{0.00002}$$

13. The percentage of Americans who use the Internet grew to 84 percent from 52 percent at the turn of the century, according to data from the Pew Research Center. (source: *NY Times*, July 28, 2015) A random sample of 75 Americans have participated in a survey.

(a) What is the expected number of Americans in the sample that do not use the internet?

Let X represent the number of Americans in the sample of do not use the internet. Then X follows a binomial probability distribution with $n = 75$ and $p = 0.16$. $E(X) = \mu = n \cdot p = 75 \cdot 0.16 = \boxed{12.0 \text{ Americans}}$

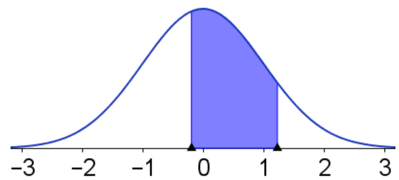
(b) What is the variance of the number of Americans that do not use the internet?

$\sigma^2 = n \cdot p \cdot q = (75) \cdot (0.16) \cdot (0.84) = 10.08 \doteq \boxed{10.1 \text{ Americans squared}}$

14. ***Find the area under the standard normal distribution curve for each.***

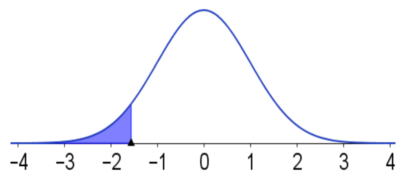
(a) Between $z = -0.19$ and $z = 1.23$.

$normalCdf(-0.19, 1.23) \doteq \boxed{0.4660}$



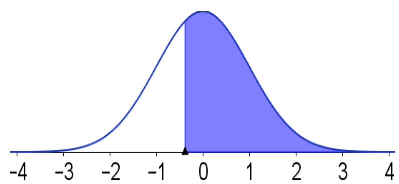
(b) To the left of $z = -1.56$.

$normalCdf(-100, -1.56) \doteq \boxed{0.0594}$



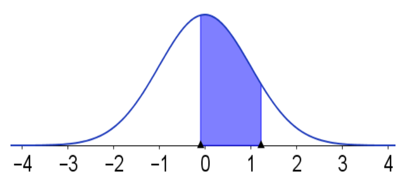
(c) To the right of $z = -0.38$.

$normalCdf(-0.38, 100) \doteq \boxed{0.6480}$

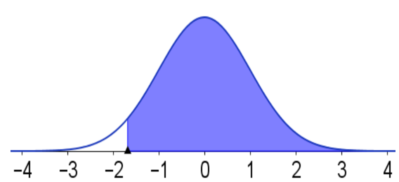


15. ***Find each probability using the standard normal distribution curve for each.***

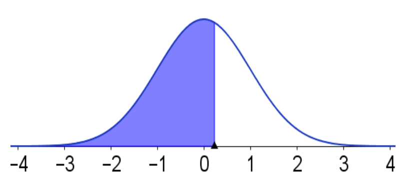
(a) $P(-0.09 < z < 2.42)$
$= normalCdf(-0.09, 2.42) \doteq \boxed{0.5281}$



(b) $P(z > -1.68)$
$= normalCdf(-1.68, 100) \doteq \boxed{0.9535}$

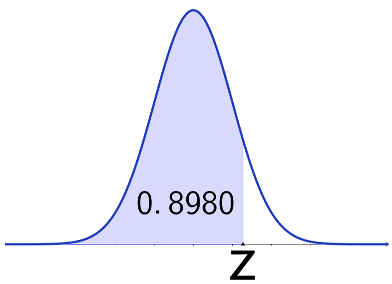


(c) $P(z < 0.23)$
$= normalCdf(-100, 0.23) \doteq \boxed{0.5910}$



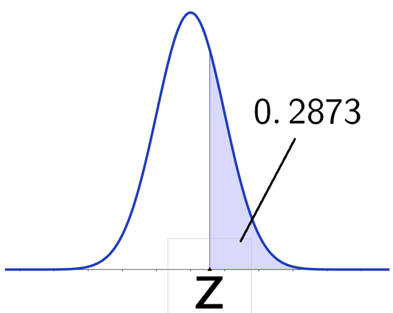
16. ***Find the indicated z score. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.***

(a)



$z = invnorm(0.8980) \doteq \boxed{1.27}$

(b)

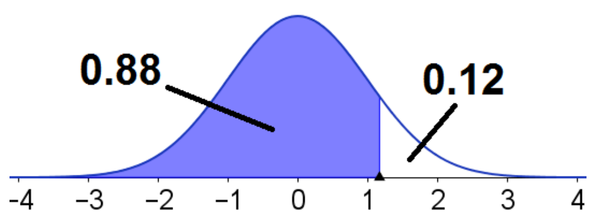


$z = invnorm(1 - 0.2873) \doteq \boxed{0.56}$

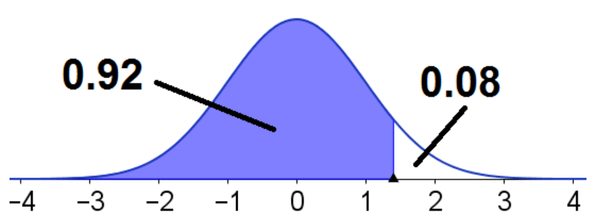
17. ***Use the Standard Normal Distribution to find the requested z value.***

(a) Find the z that separates the lower 88% of z scores from the top 12%

$z = invnorm(0.88) \doteq \boxed{1.17}$

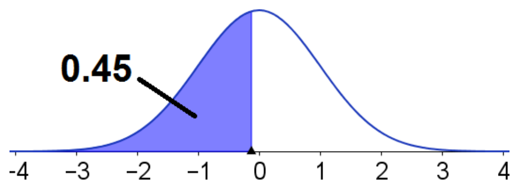


(b) $z = invnorm(0.92) \doteq \boxed{1.41}$



18. Find the critical value of z that represents the 45th percentile.

$z = invnorm(0.45) \doteq \boxed{\text{-0.13}}$



19. The waiting time in line at a Starschmuchs Coffee is normally distributed with a mean of 3.2 minutes and a standard deviation of 1.3 minutes. Find the probability that a randomly selected customer has to wait

(a) Less than 1 minute.

Calculator solution: $\mu = 3.2$ and $\sigma = 1.3$. Let X = the continuous random variable (CRV) representing a randomly selected wait time. We find the area under the distribution of wait times, X, with

$$P(X < 1) = normalCdf(-\infty, X, \mu, \sigma) = normalCdf(-10^9, 1, 3.2, 1.3) \doteq \boxed{0.0453}$$

---

z-table solution: We convert the distribution of X values over to its distribution of z scores, and find the corresponding area under the standard(ized) normal curve.

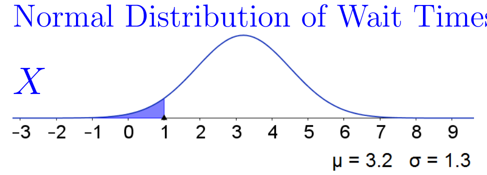
$$P(X < 1) = P\left(z < \frac{X-\mu}{\sigma}\right)$$

$$= P\left(z < \frac{1-3.2}{1.3}\right)$$

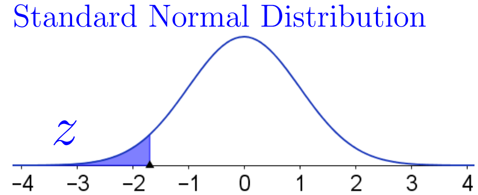
$$\doteq P(z < -1.69)$$

$$\doteq \boxed{0.0455}$$

Normal Distribution of Wait Times



Standard Normal Distribution



(b) more than 2 minutes.

Calculator solution: We find the area under the distribution of wait times, X, with

$$P(X > 2) = normalCdf(X, \infty, \mu, \sigma) = normalCdf(2, 10^9, 3.2, 1.3) \doteq \boxed{0.8220}$$

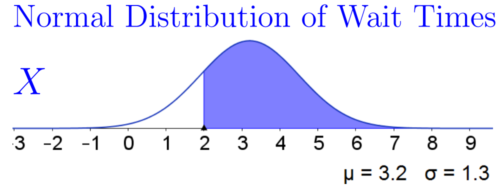
---

z-table solution:

$$P(X > 2) = P\left(z > \frac{X-\mu}{\sigma}\right)$$

$$= P\left(z > \frac{2-3.2}{1.3}\right)$$

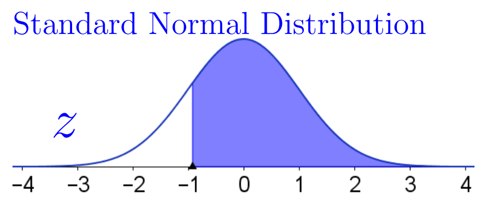
Normal Distribution of Wait Times



$\doteq P(z > -0.92)$

$= 1 - P(z < -0.92)$

$\doteq \boxed{0.8212}$



(c) between 0.75 minutes and 2 minutes.

Calculator solution:
$P(0.75 < X < 2) = normalCdf(0.75, 2, 3.2, 1.3) \doteq \boxed{0.1482}$

---

z-table solution:

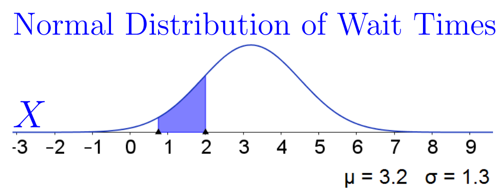
$P(0.75 < X < 2)$

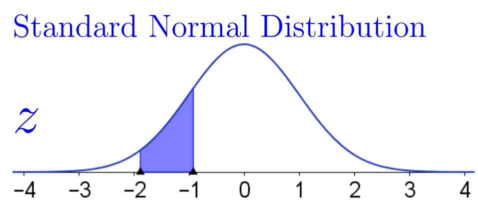
$= P\left(\frac{0.75 - 3.2}{1.3} < z < \frac{2 - 3.2}{1.3}\right)$

$\doteq P(-1.88 < z < -0.92)$

$= P(z < -0.92) - P(z < -1.88)$

$\doteq \boxed{0.1487}$





20. The average yearly precipitation in San Diego is 9.62 inches with a standard deviation of 4.42 inches and precipitation amounts are normally distributed.

(a) Find the probability that a randomly selected year will have precipitation greater than 12 inches.

Calculator solution:
$\mu = 9.62$ in. and $\sigma = 4.42$ in. Let X = the continuous random variable (CRV) representing a randomly selected yearly precipitation amount.
We find the area under the distribution of precipitations, X, with

$$P(X > 12) = normalCdf(X, \infty, \mu, \sigma) = normalCdf(12, 10^9, 9.62, 4.42) \doteq \boxed{0.2951}$$

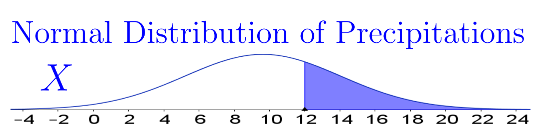
---

z-table solution:

$P(X > 12) = P\left(z > \frac{X - \mu}{\sigma}\right)$

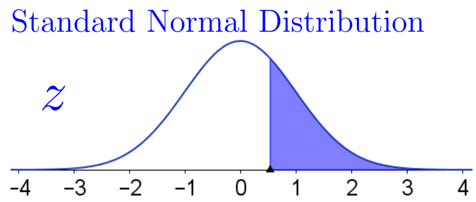
$= P\left(z > \frac{12 - 9.62}{4.42}\right)$

$\doteq P(z > 0.54)$



$= 1 - P(z < 0.54)$

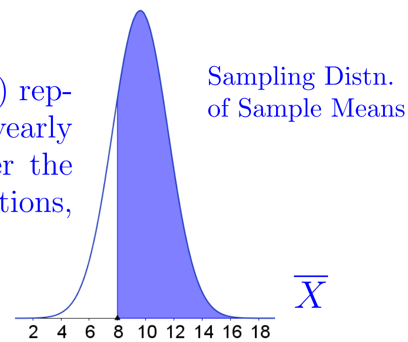
$\doteq \boxed{0.2946}$



(b) Find the probability that five randomly selected years will have an average precipitation greater than 8 inches.

Calculator solution:

Let $\overline{X}$ = the continuous random variable (CRV) representing a randomly selected sample mean yearly precipitation amount. We find the area under the sampling distribution of sample mean precipitations, $\overline{X}$, with



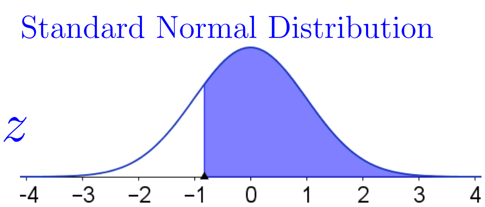
$$P(\overline{X} > 8) = normalCdf(\overline{X}, \infty, \mu_{\overline{X}}, \sigma_{\overline{X}})$$

$$= normalCdf(8, 10^9, 9.62, 4.42/\sqrt{5})$$

$$\doteq \boxed{0.7938}$$

$\mu_{\overline{X}} = 9.62$ and $\sigma_{\overline{X}} = 4.42/\sqrt{5} \doteq 1.98$

z-table solution:

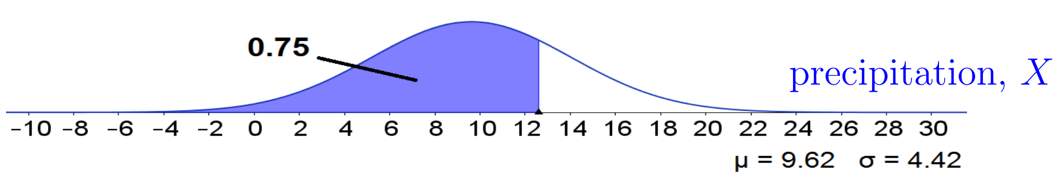


$$P(\overline{X} > 8) = P\left(z > \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(z > \frac{8 - 9.62}{4.42/\sqrt{5}}\right)$$

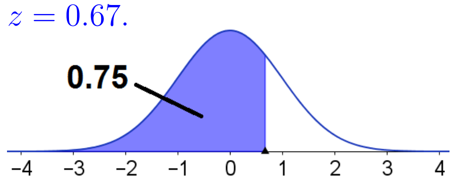
$$= P(z > -0.82) = 1 - P(z < -0.82) \doteq \boxed{0.7939}$$

(c) Find the precipitation amount from the distribution of precipitations that represents the 75th percentile.



$$X = invnorm(percentile, \mu, \sigma) = invnorm(0.75, 9.62, 4.42) \doteq \boxed{12.6 \text{ in}}$$

z-table solution:
We first find the z value from the standard normal distribution that corresponds to the 75th percentile. We find this value to be $z = 0.67$.



Second, we solve $z = \dfrac{X - \mu}{\sigma}$ for X to get

$$X = \mu + z \cdot \sigma$$

Afterwards, replace μ with 9.62, z with 0.67 and σ with 4.42 so that

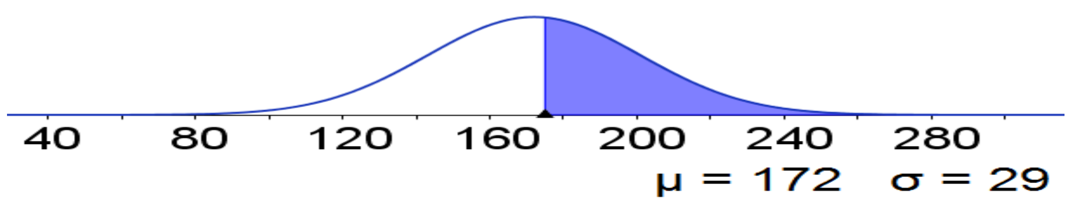
$$X = 9.62 + (0.67) \cdot (4.42) \doteq \boxed{12.6 \text{ in.}}$$

21. Some passengers died when a water taxi sank in Baltimore's inner harbor. Men are typically heavier than women and children, so when loading a water taxi, let's assume a worst-case scenario in which all passengers are men. Based on data from the National Health and Nutrition Survey, assume that weights of men are normally distributed with a mean of 172 lb. and a standard deviation of 29 lb.

(a) Find the probability that if an individual man is randomly selected, his weight will be greater than 175 lb.

Calculator solution: $\mu = 172$ lb. and $\sigma = 29$ lb. Let $X =$ the continuous random variable (CRV) representing the weight of a randomly selected man. We find the area under the distribution of weights, X, with

$$P(X > 175) = normalCdf(X, \infty, \mu, \sigma) = normalCdf(175, 10^9, 172, 29) \doteq \boxed{0.4588}$$

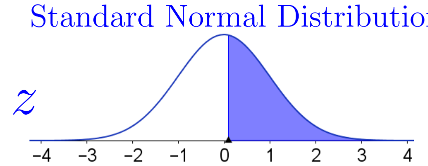


Distn. of Men's Weights, X

z-table solution:

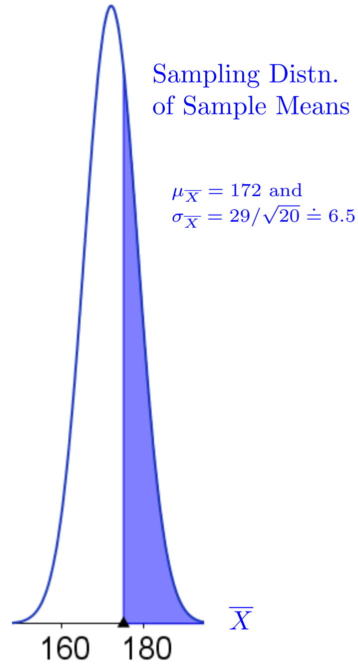
$$\begin{aligned} P(X > 175) &= P\left(z > \frac{X - \mu}{\sigma}\right) \\ &= P\left(z > \frac{175 - 172}{29}\right) \\ &\doteq P(z > 0.10) \\ &= 1 - P(z < 0.10) \\ &\doteq \boxed{0.4602} \end{aligned}$$

Standard Normal Distribution



(b) Find the probability that 20 men will have a mean weight that is greater than 175 lb. (so that their total weight exceeds the safe capacity of 3500 lb.

Let $\overline{X}$ = the continuous random variable (CRV) representing a randomly selected sample mean weight. We find the area under the sampling distribution of sample mean weights, $\overline{X}$, with



$$P(\overline{X} > 175) = normalCdf(\overline{X}, \infty, \mu_{\overline{X}}, \sigma_{\overline{X}})$$
$$= normalCdf(175, 10^9, 172, 29/\sqrt{20})$$
$$\doteq \boxed{0.3218}$$

z-table solution:

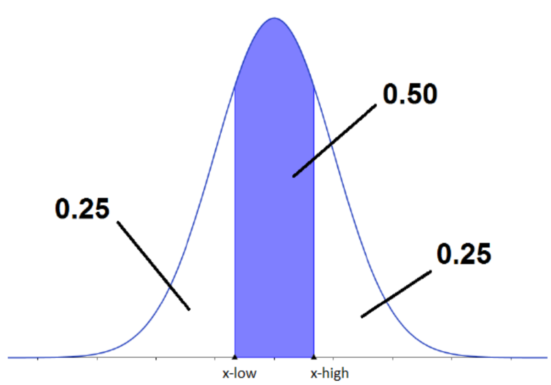
$$P(\overline{X} > 175) = P\left(z > \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)$$
$$= P\left(z > \frac{175 - 172}{29/\sqrt{20}}\right)$$
$$\doteq P(z > 0.46) = 1 - P(z < 0.46) \doteq \boxed{0.3228}$$

22. The average per capita spending on health care in the United States is \$5274. The standard deviation is \$600 and the distribution of health care spending is approximately normal. Find the limits of the middle 50% of individual health care expenditures.

Let X be the continuous random variable (CRV) representing a randomly selected individual health care expenditure. The leftmost value in the middle 50%, x_{low} (picture below), has 25% of heath care expenditures below it, while the rightmost value in the middle 50%, x_{high} (picture below), has 75% of heath care expenditures below it

$$x_{low} = invnorm(percentile, \mu, \sigma) = invnorm(0.25, 5274, 600) \doteq \boxed{\$4869.31}$$

$$x_{high} = invnorm(percentile, \mu, \sigma) = invnorm(0.75, 5274, 600) \doteq \boxed{\$5678.69}$$

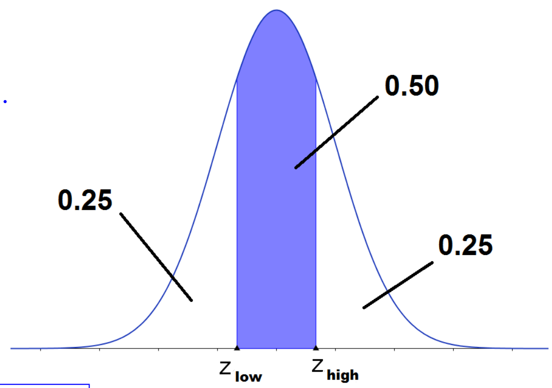


z-table solution:
Let z_{low} and z_{high} represent the z scores from the standard normal distribution corresponding to the 25th and 75th percentiles. Then we use the z score formula solved for x:

$$X = \mu + z \cdot \sigma$$

From the z table we get that $z_{low} = -0.67$, and by symmetry, $z_{high} = 0.67$, and

Standard Normal Distribution



$$x_{low} = \mu + z \cdot \sigma \doteq \$5274 + (-0.67) \cdot (\$600) \approx \boxed{\$4872}$$

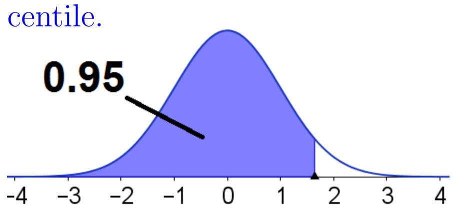
$$x_{high} = \mu + z \cdot \sigma \doteq \$5274 + (0.67) \cdot (\$600) \approx \boxed{\$5676}$$

23. A prestigious college decides to only take applications from students who have scored in the top 5% on the SAT test. The SAT scores are approximately normally distributed with a mean of 490 and a standard deviation of 70. Find the score that is necessary to obtain in order to qualify for applying to this college.

Calculator solution:
Let X be the continuous random variable (CRV) representing SAT scores. A student must score in the 95th (or higher) percentile in order to be admitted, so the area left of the unknown value is 0.95.

$$\begin{aligned} X &= invnorm(percentile, \mu, \sigma) \\ &= invnorm(0.95, 490, 70) \doteq \boxed{605} \end{aligned}$$

z-table solution:
First, we find the z value from the standard normal distribution that corresponds to the 95th percentile.



$$\begin{aligned} z &= invnorm(0.95) \\ &\doteq 1.644853626 \end{aligned}$$

Second, we solve $z = \dfrac{X - \mu}{\sigma}$ for X to get

$$X = \mu + z \cdot \sigma$$

Afterwards, replace μ with 490, z with 1.644853626 and σ with 70 so that

$$X = 490 + (1.644853626) \cdot (70) \doteq \boxed{605}$$

24. Americans ate an average of 25.7 pounds of Krusty-O Cereal each last year and spent an average of \$61.50 per person doing so. If the standard deviation for consumption is 3.75 pounds and the standard deviation for the amount spent is \$5.89, find the following:

(a) The probability that the sample mean Krusty-O cereal consumption for a random sample of 40 American consumers exceeded 27 pounds.

Let $\overline{X}$ = the continuous random variable (CRV) representing a randomly selected sample mean consumption amount. We find the area under the sampling distribution with

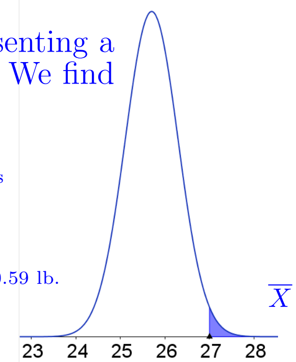
$$P(\overline{X} > 27) = normalCdf(\overline{X}, \infty, \mu_{\overline{X}}, \sigma_{\overline{X}})$$

$$= normalCdf(27, 10^9, 25.7, 3.75/\sqrt{40})$$

$$\doteq \boxed{0.0142}$$

Sampling Distn. of Sample Means

$\mu_{\overline{X}} = 25.7$ lb. and $\sigma_{\overline{X}} = 3.75/\sqrt{40} \doteq 0.59$ lb.



Std. Normal Distn.

z-table solution:

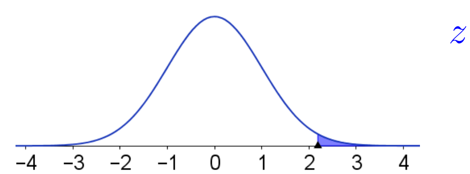
$$P(\overline{X} > 27) = P\left(z > \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(z > \frac{27 - 25.7}{3.75/\sqrt{40}}\right)$$

$$\doteq P(z > 2.19)$$

$$= normalCdf(2.19, 100)$$

$$\doteq \boxed{0.0143}$$



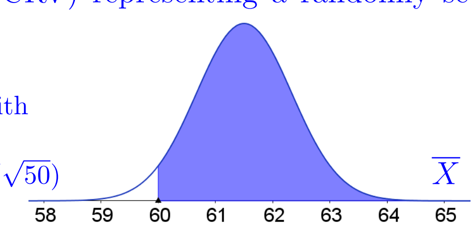
(b) The probability that for a random sample of 50, the the average yearly amount spent on Krusty-O Cereal was between \$60.00 and \$100.

Let $\overline{X}$ = the continuous random variable (CRV) representing a randomly selected sample mean yearly amount spent.

We find the area under the sampling distribution with

$$P(60 < \overline{X} < 100) = normalCdf(60, 100, 61.50, 5.89/\sqrt{50})$$

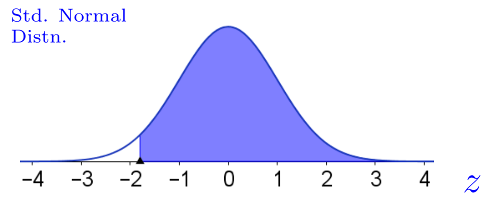
$$\doteq \boxed{0.9641}$$



Sampling Distn. of Sample Means

$\mu_{\overline{X}} = \$61.50$ and $\sigma_{\overline{X}} = 5.89/\sqrt{50} \doteq \$0.83$

z-table solution:



$P(60 < \overline{X} < 100)$

$= P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)$

$= P\left(\frac{60-61.50}{5.89/\sqrt{50}} < z < \frac{100-61.50}{5.89/\sqrt{50}}\right)$

$\doteq P(-1.8 < z < 46.22)$

$= P(z < 46.22) - P(z < -1.8) \doteq \boxed{0.9641}$

25. Use the information in the table below to answer parts a through f. The ages (in years) of 10 men and their systolic blood pressures (in millimeters of mercury) are listed in the table.

Age, x	16	25	39	45	49	64	70	29	57	22
Systolic Blood Pressure, y	109	122	143	132	199	185	199	130	175	118

(a) What is the sample correlation coefficient, r? $\boxed{0.908}$

(b) Describe the type of correlation and interpret the correlation in the context of the data.

There is a strong positive correlation. As age increases, systolic blood pressure tends to increase.

(c) Find the equation of the regression line for the data. $\hat{y} = 1.705x + 80.255$

(d) Use the regression equation to predict the value of y for $x = 42$. $\boxed{152}$

(e) Use the regression equation to predict the blood pressure for a man aged 67. $\boxed{194}$

Statistics   Practice Test 3   Chapters 6 and 7

1. Find the critical value z_c that is needed to set up a 92% confidence interval estimate for the population mean.

2. First-semester GPAs for a random selection of freshmen at a large university are shown below. Estimate the value of the population mean GPA of the freshman class with 99% confidence. Assume $\sigma = 0.62$ and that the distribution of first-semester GPAs is normal.

1.9	3.2	2.0	2.9	2.7	3.3
2.8	3.0	3.8	2.7	2.0	1.9
2.5	2.7	2.8	3.2	3.0	3.8
3.1	2.7	3.5	3.8	3.9	2.7

3. Find the critical value t_c that corresponds to a 90% interval, assuming $n = 10$.

4. The approximate costs for a 30-second spot for various cable networks in a random selection of cities are shown below. Estimate the population mean cost for a 30-second advertisement on cable network with 90% confidence. Assume the population of costs is approximately normal.

14	55	165	9	15	66	23	30	150
22	12	13	54	73	55	41	78	

5. A university dean of students wishes to estimate the average number of hours students spend doing homework per week. The standard deviation from a previous study is 6.2 hours. How large a sample must be selected if he wants to be 99% confident of finding whether the population mean differs from the sample mean by 1.5 hours?

6. Thirty randomly selected students took the calculus final. If the sample mean was 95 and the sample standard deviation was 6.6, construct a 99% confidence interval for the mean score of all students.

7. A study of 35 golfers showed that their average score on a particular course was 92. The standard deviation of the population is 5. Find the 95% confidence interval of the mean score for all golfers.

8. A recent study of 75 workers found that 53 people rode the bus to work each day. Find the 95% confidence interval of the proportion of all workers who rode the bus to work

9. It is believed that 25% of U.S. homes have a direct satellite television receiver. How large a sample is necessary to estimate the population proportion of homes that have a direct satellite television receiver within 3 percentage points? How large a sample is necessary if nothing is known about the proportion? Use 95% level of confidence.

10. A recent poll showed results from 2000 professionals who interview job applicants. 26% of them said the biggest interview turnoff is that the applicant did not make an effort to learn about the job or the company. A 99% confidence interval estimate was used and the margin of error was ± 3 percentage points. Describe what is meant by the statement "the margin of error was ± 3 percentage points."

11. Find the critical values χ^2_L and χ^2_R needed to set up a 95% confidence interval for the population standard deviation when the sample size is 18. Assume the population is normally distributed.

12. You randomly select and weigh a sample of 30 allergy medicine pills. The sample standard deviation is 1.2 milligrams. Assuming the weights are normally distributed, construct 99% confidence intervals for the population variance and standard deviation.

For each hypothesis test question listed below, answer parts a through i.

(a) Which test is appropriate here?
(b) Write the symbolic form of the claim.
(c) Write the null and alternative hypotheses.
(d) What number is the test statistic equal to?
(e) What formula should be used for the the standardized test statistic?
(f) What number is the standardized test statistic equal to? Round to the hundredths.
(g) What p-value do you obtain? Round to the ten-thousandths.
(h) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.
(i) State a full sentence conclusion stating the decision you made.

---

13. A poll of 1000 adult Americans reveals that 48% of the voters surveyed prefer the Democratic candidate for the presidency. At the 0.05 level of significance, test the claim that at least half of all voters prefer the Democrat.

14. A manufacturer claims that its televisions have an average lifetime of five years (60 months) with a population standard deviation of seven months. Eighty-one televisions were selected at random, and the average lifetime was found to be 58 months. With $\alpha = 0.025$, is the manufacturers claim supported?

15. In tests of a computer component, it is found that the mean time between failures is 520 hours. A modification is made which is supposed to increase the time between failures. Tests on a random sample of 10 modified components resulted in the following times (in hours) between failures.

518	548	561	523	536
499	538	557	528	563

At the 0.05 significance level, test the claim that for the modified components, the mean time between failures is greater than 520 hours.

16. A simple random sample of 15-year old boys from one city is obtained and their weights (in pounds) are listed below. Use $\alpha = 0.01$ to test the claim that these sample weights come from a population with a mean equal to 150 lb. Assume that the population standard deviation of the weights of all 15-year old boys in the city is known to be 16.6 lb.

150	139	158	151	134	189	157	144	175	127
158	125	167	166	142	145	162	143	172	145
160	164	133	144	156	184	122	188	155	157

17. An airline claims that the no-show rate for passengers is less than 5%. In a sample of 420 randomly selected reservations, 19 were no-shows. At $\alpha = 0.01$, test the airlines claim.

18. A large software company gives job applicants a test of programming ability and the mean for that test has been 160 in the past. Twenty-five job applicants are randomly selected from one large university and they produce a mean score and standard deviation of 183 and 12, respectively. Use a 0.05 level of significance to test the claim that this sample comes from a population with a mean score greater than 160.

19. Suppose a sample of 21 measurements from a normal population had a variance of 5 inches squared. Use $\alpha = 0.05$ to test the claim that the population variance is at least 8 inches squared.

20. Suppose a sample of 16 measurements from a normal population had a sample standard deviation of 4 cm. Use $\alpha = 0.05$ to test the claim that the population standard deviation is fewer than 5 cm.

21. A precision instrument is guaranteed to read accurately to within 2 units. A sample of four instrument readings on the same object yielded the measurements 352, 351, 351, 354. Assuming the sample is taken from a population distribution that is normal, test the claim that σ is more than 0.7 units. Use a 1% level of significance.

22. A cigarette manufacturer wishes to test the claim that the variance of nicotine content of its cigarettes is not more than 0.644 milligrams squared. A sample of 20 cigarettes has a variance of 0.760 miligrams squared. Use a 5% level of significance to test the claim.

23. A nationwide study of American homeowners revealed that 66% have one or more lawn mowers. A lawn equipment manufacturer, located in Omaha, feels the estimate is too low for households in Omaha. Test the claim that the proportion with lawn mowers in Omaha is higher than 66%. Among 500 randomly selected homes in Omaha, 68% had one or more lawn mowers.

24. A fast food outlet claims that the mean waiting time in line is less than 3.8 minutes. A random sample of 20 customers has a mean of 3.7 minutes and standard deviation of 0.6 minute. If $\alpha = 0.05$, test the fast food outlet's claim.

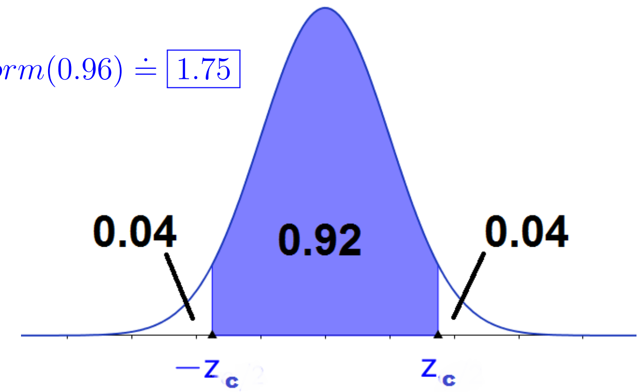
25. A consumer group claims that the mean annual consumption of coffee by a person in the United States is 23.2 gallons. A random sample of 90 people in the U.S. has a mean annual coffee consumption of 21.6 gallons. Assume the population standard deviation is 4.8 gallons. Test the group's claim with $\alpha = 0.05$.

Statistics **Test 3** **Name:** KEY

1. Find the critical value z_c that is needed to set up a 92% confidence interval estimate for the population mean.

We are given $c = 0.92$. Then

$$z_c = invnorm\left(\frac{1+c}{2}, 0, 1\right) = invnorm(0.96) \doteq \boxed{1.75}$$

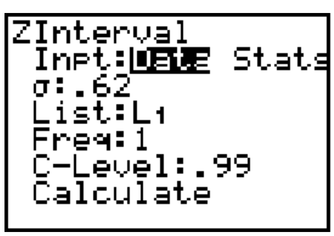


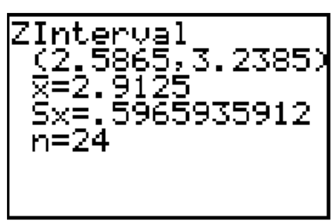
2. First-semester GPAs for a random selection of freshmen at a large university are shown below. Estimate the value of the population mean GPA of the freshman class with 99% confidence. Assume $\sigma = 0.62$ and that the distribution of first-semester GPAs is normal.

1.9	3.2	2.0	2.9	2.7	3.3
2.8	3.0	3.8	2.7	2.0	1.9
2.5	2.7	2.8	3.2	3.0	3.8
3.1	2.7	3.5	3.8	3.9	2.7

Calculator Solution:

Put the data in your calculator's list one then run a z-interval on the calculator. The result from running the z-interval is shown below. The answer is $\boxed{2.5865 < \mu < 3.2385}$.





Formula Solution: We find the sample mean is $\overline{X} \doteq 2.9125$. Using $c = 0.99$ we determine that

$z_c = invnorm\left(\frac{1+c}{2}, 0, 1\right) = invnorm\left(\frac{1.99}{2}, 0, 1\right) \doteq 2.58$. The z-interval formula is

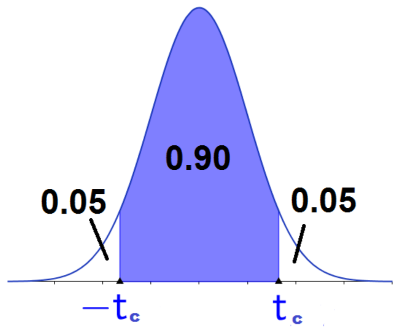
the inequality: $\overline{X} - z_c \cdot \left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{X} + z_c \cdot \left(\frac{\sigma}{\sqrt{n}}\right)$, or equivalently,

$2.9125 - 2.58\left(\frac{0.62}{\sqrt{24}}\right) < \mu < 2.9125 + 2.58\left(\frac{0.62}{\sqrt{24}}\right)$, or equivalently, $\boxed{2.58 < \mu < 3.24}$.

3. Find the critical value t_c that corresponds to a 90% interval, assuming $n = 10$.

We are given $c = 0.90$. Then

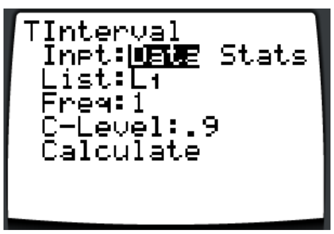
$$\begin{aligned} t_c =& invT\left(\frac{1+c}{2},\ (n-1)\right) \\ =& invT(0.95, 9) \\ \doteq& \ 1.833112 \\ \approx& \ \boxed{1.833} \end{aligned}$$

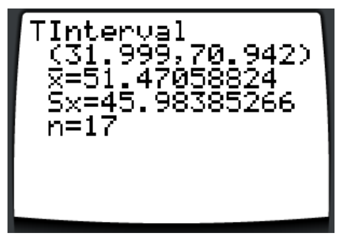


4. The approximate costs (in thousands) for a 30-second spot for various cable networks in a random selection of cities are shown below. Estimate the population mean cost for a 30-second advertisement on cable network with 90% confidence.Assume the population of costs is approximately normal.

14	55	165	9	15	66	23	30	150
22	12	13	54	73	55	41	78	

Calculator Solution: Enter the data in the calculator's list one and then run a t-interval (since σ is not given). The answer is $\boxed{32.0 < \mu < 70.9}$





Formula Solution: We find the sample mean is $\overline{X} \doteq 51.4705$ and sample standard deviation is $s \doteq 45.9839$. We use $c = 0.90$ to find that $t_c = invT\left(\frac{1+c}{2},\ (n-1)\right) = invT(0.95, 16) \doteq 1.746.$

Then, the formula for the t-interval is the inequality:

$$\overline{X} - t_c \cdot \left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{X} + t_c \cdot \left(\frac{s}{\sqrt{n}}\right),$$

or, equivalently

$$51.4705 - 1.746\left(\frac{45.9839}{\sqrt{17}}\right) < \mu < 51.4705 + 1.746\left(\frac{45.9839}{\sqrt{17}}\right),$$

or, equivalently

$$\boxed{32.0 < \mu < 70.9}$$

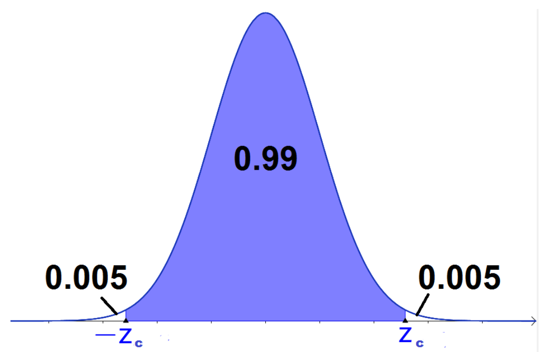
5. A university dean of students wishes to estimate the average number of hours students spend doing homework per week. The standard deviation from a previous study is 6.2 hours. How large a sample must be selected if he wants to be 99% confident of finding whether the population mean differs from the sample mean by 1.5 hours?

We are asked to determine the size, n, of a sample necessary for an interval estimate of the average weekly study amount. The formula is

$$n = \left(\frac{z_c \cdot \sigma}{E}\right)^2$$

We are told to assume $\sigma = 6.2$, $c = 0.99$ and $E = 1.5$. Then

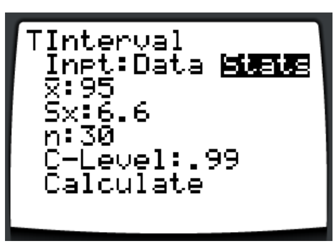
$$\begin{aligned} z_c &= invnorm\left(\frac{1+c}{2}, 0, 1\right) \\ &= invnorm(0.995) \\ &= 2.58, \text{ and} \end{aligned}$$

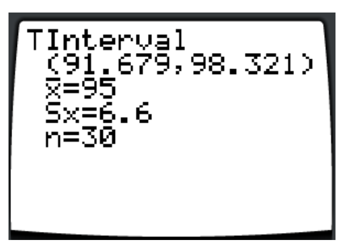


$$n = \left(\frac{z_c \cdot \sigma}{E}\right)^2 = \left(\frac{(2.58) \cdot (6.2)}{1.5}\right)^2 \doteq \boxed{114}$$

6. Thirty randomly selected students took the calculus final. If the sample mean was 95 and the standard deviation was 6.6, construct a 99% confidence interval for the mean score of all students.

Calculator Solution: We are not given a value for σ, so we the t-interval command on the calculator to set up the confidence interval (however since $n \geq 30$, we could use a z-interval). We are given the values $n = 30$, $c = 0.99$, $\overline{X} = 95$ and $s = 6.6$ in the statement of the problem. The input and outputs screens from running the t-interval are below. The answer is $\boxed{92 < \mu < 98}$





Formula Solution: We are given $n = 30$, $c = 0.99$, $\overline{X} = 95$ and $s = 6.6$. Then $t_c = invT\left(\frac{1+c}{2},\ (n-1)\right) = invT\left(\frac{1.99}{2}, 29\right) = invT(0.995, 29) \doteq 2.756$. The formula for the t-interval is the inequality:

$$\overline{X} - t_c \cdot \left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{X} + t_c \cdot \left(\frac{s}{\sqrt{n}}\right),$$

or, equivalently

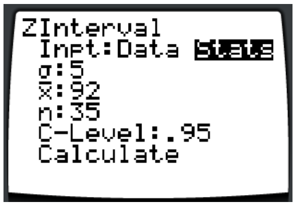
$$95 - 2.756\left(\frac{6.6}{\sqrt{30}}\right) < \mu < 95 + 2.756\left(\frac{6.6}{\sqrt{30}}\right),$$

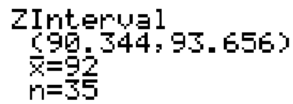
or, equivalently

$$\boxed{92 < \mu < 98}$$

7. A study of 35 golfers showed that their average score on a particular course was 92. The standard deviation of the population is 5. Find the 95% confidence interval of the mean score for all golfers.

Calculator Solution: We are given $\sigma = 5$, so we use the z-interval formula. We are given $n = 35$, $c = 0.95$ and $\overline{X} = 92$. The input and outputs screens from running the z-interval are below. The answer is $\boxed{90.3 < \mu < 93.7}$





Formula Solution: We are given the values $\sigma = 5$, $n = 35$, $c = 0.95$ and $\overline{X} = 92$. We use $c = 0.95$ to find $z_c = 1.96$. Then the formula for the confidence interval is the inequality

$$\overline{X} - z_c \cdot \left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{X} + z_c \cdot \left(\frac{\sigma}{\sqrt{n}}\right),$$

or, equivalently

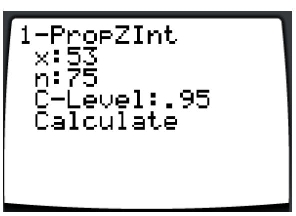
$$92 - 1.96\left(\frac{5}{\sqrt{35}}\right) < \mu < 92 + 1.96\left(\frac{5}{\sqrt{35}}\right),$$

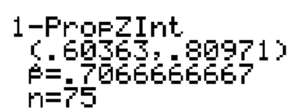
or, equivalently

$$\boxed{90.3 < \mu < 93.7}$$

8. A recent study of 75 workers found that 53 people rode the bus to work each day. Find the 95% confidence interval of the proportion of all workers who rode the bus to work.

Calculator Solution: We are given $n = 75$, $X = 53$ and $c = 0.95$. Use a 1-prop-z-interval on the calculator. The answer is $\boxed{0.60 < p < 0.81}$





We are given $n = 75$ and $X = 53$, so the sample proportion is $\hat{p} = \dfrac{53}{75}$.
95% confidence implies that $z_c = 1.96$ Formula Solution:

$$\hat{p} - z_c \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} < p < \hat{p} + z_c \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}},$$

or, equivalently

$$\frac{53}{75} - 1.96 \cdot \sqrt{\frac{\frac{53}{75} \cdot \left(1 - \frac{53}{75}\right)}{75}} < p < \frac{53}{75} + 1.96 \cdot \sqrt{\frac{\frac{53}{75} \cdot \left(1 - \frac{53}{75}\right)}{75}},$$

or, equivalently

$$\boxed{0.60 < p < 0.81}$$

9. It is believed that 25% of U.S. homes have a direct satellite television receiver. How large a sample is necessary to estimate the population proportion of homes that have a direct satellite television receiver within 3 percentage points? How large a sample is necessary if nothing is known about the proportion? Use 95% level of confidence.

$$n = \hat{p} \cdot \hat{q} \cdot \left(\frac{z_c}{E} \right)^2 = 0.25 \cdot 0.75 \cdot \left(\frac{1.96}{0.03} \right)^2 \doteq 801$$

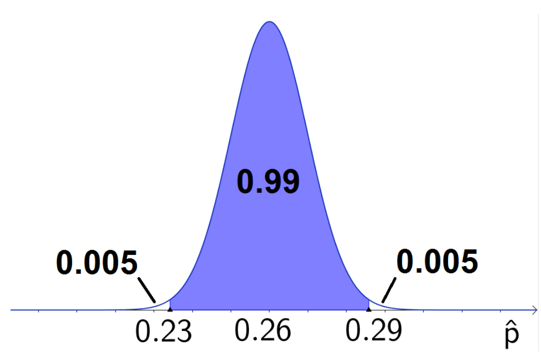
Assuming nothing is known about the proportion,

$$n = \hat{p} \cdot \hat{q} \cdot \left(\frac{z_c}{E} \right)^2 = 0.5 \cdot 0.5 \cdot \left(\frac{1.96}{0.03} \right)^2 \doteq 1068$$

10. A recent poll showed results from 2000 professionals who interview job applicants. 26% of them said the biggest interview turnoff is that the applicant did not make an effort to learn about the job or the company. A 99% confidence interval estimate was used and the margin of error was ±3 percentage points. Describe what is meant by the statement "the margin of error was ±3 percentage points."

When using 26% to estimate the value of the population percentage, the maximum likely difference between 26% and the population percentage is three percentage points, so the interval from 23% to 29% is likely to contain the population percentage.

The Sampling Distribution of Sample Proportions, $\hat{p}$.



11. Find the critical values χ_L^2 and χ_R^2 needed to set up a 95% confidence interval for the population standard deviation when the sample size is 18. Assume the population is normally distributed.

Because the sample size is 18, $df = n - 1 = 17$ The areas to the left of χ_L^2 and χ_R^2 are

(Area to the left of χ_L^2) $= \dfrac{1-c}{2} = \dfrac{1-.95}{2} = 0.025 = 2.5\%$

(Area to the left of χ_R^2) $= \dfrac{1+c}{2} = \dfrac{1+.95}{2} = 0.975 = 97.5\%$

Using $df = 17$ and the areas 2.5% and 97.5% you can find the critical values off of the chi-square table. We find that $\chi_L^2 = 7.564$ and $\chi_R^2 = 30.191$ So, for a chi-square distribution curve with 17 degrees of freedom, 95% of the area under the curve lies between 7.564 and 30.191.

12. You randomly select and weigh a sample of 30 allergy medicine pills. The sample standard deviation is 1.2 milligrams. Assuming the weights are normally distributed, construct 99% confidence intervals for the population variance and standard deviation.

Because the sample size is 30, $df = n - 1 = 29$ The areas to the left of χ_L^2 and χ_R^2 are

(Area to the left of χ_L^2) $= \dfrac{1-c}{2} = \dfrac{1-.99}{2} = 0.005 = 0.5\%$

(Area to the left of χ_R^2) $= \dfrac{1+c}{2} = \dfrac{1+.99}{2} = 0.995 = 99.5\%$

Using $df = 29$ and the areas 0.5% and 99.5% you can find the critical values off of the chi-square table. We find that $\chi_L^2 = 13.121$ and $\chi_R^2 = 52.336$

The left endpoint of the confidence interval is

$$\frac{(n-1)s^2}{\chi_R^2} = \frac{(30-1)1.2^2}{52.336} = 0.80$$

The right endpoint of the confidence interval is

$$\frac{(n-1)s^2}{\chi_L^2} = \frac{(30-1)1.2^2}{13.121} = 3.18$$

The confidence interval formula for σ^2 is the inequality

$$\frac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$$

or

$$0.80 < \sigma^2 < 3.18$$

The confidence interval formula for σ is

$$\sqrt{\frac{(n-1)s^2}{\chi_R^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_L^2}}$$

or equivalently,

12 (continued)

$$\sqrt{\frac{(30-1)1.2^2}{52.336}} < \sigma < \sqrt{\frac{(30-1)1.2^2}{13.121}}$$

or

$$0.89 < \sigma < 1.78$$

The population variance is between 0.80 milligrams squared and 3.18 milligrams squared, and the population standard deviation is between 0.89 milligrams and 1.78 milligrams. There is a 99% chance that these intervals contain the actual values of σ^2 and σ

THE SOLUTIONS TO PROBLEMS 19 THROUGH 22 CAN BE FOUND ON THIS WEBPAGE (see examples 1 through 4):

http://timbusken.com/hypothesis_testing1.html

(13) A poll of 1000 adult Americans reveals that 48% of the voters surveyed prefer the Democratic candidate for the presidency. At the 0.05 level of significance, test the claim that at least half of all voters prefer the Democrat.

$n = 1000$

$\hat{p} = 0.48$

$\alpha = 0.05$

$x = n \cdot \hat{p} = 480$

(a) (1 point) Which test (z-test, t-test or 1-prop-z-test) is appropriate here?

(b) (1 point) Write the symbolic form of the claim. $p \geq 0.50$ or $p_0 \geq 0.50$

(c) (2 points) Write the null and alternative hypotheses.

$H_0: p \geq 0.50$ (claim)

$H_A: p < 0.50$

(d) (1 point) What number is the test statistic equal to?

$\hat{p} = 0.48$

(e) (1 point) What formula should be used for the the standardized test statistic?

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

(f) (1 point) What number is the standardized test statistic equal to? Round to the hundredths.

$z \approx -1.26$ standard deviations

(g) (1 point) What p-value do you obtain? Round to the ten-thousandths.

p-value $= P(\hat{p} \leq 0.48) \approx 0.1030$

(h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

Since the p-value $> \alpha$, fail to reject H_0

(i) (1 point) State a full sentence conclusion stating the decision you made.

There is not sufficient evidence to warrant rejection of the claim.

(14) A manufacturer claims that its televisions have an average lifetime of five years (60 months) with a population standard deviation of seven months. Eighty-one televisions were selected at random, and the average lifetime was found to be 58 months. With $\alpha = 0.025$, is the manufacturer's claim supported?

Claim
$\mu = 60$
$\sigma = 7$
$n = 81$
$\bar{x} = 58$
$\alpha = 0.025$

(a) (1 point) Which test (z-test, t-test or 1-prop-z-test) is appropriate here?

(b) (1 point) Write the symbolic form of the claim.

$\mu = 60$ months or $\mu_0 = 60$ months

(c) (2 points) Write the null and alternative hypotheses.

$H_0: \mu = 60$ (claim)

$H_A: \mu \neq 60$

(d) (1 point) What number is the test statistic equal to?

$\bar{X} = 58$ months

(e) (1 point) What formula should be used for the the standardized test statistic?

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

(f) (1 point) What number is the standardized test statistic equal to? Round to the hundredths.

$Z \approx -2.57$ standard deviations

(g) (1 point) What p-value do you obtain? Round to the ten-thousandths.

p-value $= 2 \cdot P(\bar{x} \leq 58) \approx 0.0101$

(h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

Since p-value $\leq \alpha$, reject H_0

(i) (1 point) State a full sentence conclusion stating the decision you made.

There is sufficient evidence to warrant rejection of the claim.

(15) In tests of a computer component, it is found that the mean time between failures is 520 hours. A modification is made which is supposed to increase the time between failures. Tests on a random sample of 10 modified components resulted in the following times (in hours) between failures.

$n = 10$

$\alpha = 0.05$

518	548	561	523	536
499	538	557	528	563

At the 0.05 significance level, test the claim that for the modified components, the mean time between failures is greater than 520 hours.

(a) (1 point) Which test (z-test, t-test or 1-prop-z-test) is appropriate here?

(b) (1 point) Write the symbolic form of the claim.

$\mu > 520$ hours

(c) (2 points) Write the null and alternative hypotheses.

$H_0: \mu \leq 520$

$H_A: \mu > 520$ (claim)

(d) (1 point) What number is the test statistic equal to?

$\bar{x} = 537.1$ hours

(e) (1 point) What formula should be used for the the standardized test statistic?

$$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

(f) (1 point) What number is the standardized test statistic equal to? Round to the hundredths.

$t \approx 2.61$ standard deviations

(g) (1 point) What p-value do you obtain? Round to the ten-thousandths.

p-value $= P(\bar{x} \geq 537.1 \text{ hours}) \approx 0.0141$

(h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

Since the p-value $\leq \alpha$, reject Ho

(i) (1 point) State a full sentence conclusion stating the decision you made.

The sample data support the claim.

(16) $\alpha = 0.01$ $\sigma = 16.6$

A simple random sample of 15-year old boys from one city is obtained and their weights (in pounds) are listed below. Use a 0.01 significance level to test the claim that these sample weights come from a population with a mean equal to 150 lb. Assume that the population standard deviation of the weights of all 15-year old boys in the city is known to be 16.6 lb.

150	139	158	151	134	189	157	144	175	127
158	125	167	166	142	145	162	143	172	145
160	164	133	144	156	184	122	188	155	157

(a) (1 point) Which test (z-test, t-test or 1-prop-z-test) is appropriate here?

(b) (1 point) Write the symbolic form of the claim. $\mu = 150$ lbs

(c) (2 points) Write the null and alternative hypotheses.

$H_0: \mu = 150$ (claim)

$H_A: \mu \neq 150$

(d) (1 point) What number is the test statistic equal to?

$\bar{x} = 153.7\bar{3}$ lbs

(e) (1 point) What formula should be used for the the standardized test statistic?

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

(f) (1 point) What number is the standardized test statistic equal to? Round to the hundredths.

$z \approx 1.23$ standard deviations

(g) (1 point) What p-value do you obtain? Round to the ten-thousandths.

p-value $= 2 \cdot P(\bar{x} \geq 153.7\bar{3}$ lbs$) \approx 0.2180$

(h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

Since p-value > α, fail to reject H_0.

(i) (1 point) State a full sentence conclusion stating the decision you made.

There is not sufficient evidence to warrant rejection of the claim.

(17) An airline claims that the no-show rate for passengers is less than 5%. In a sample of 420 randomly selected reservations, 19 were no-shows. At $\alpha = 0.01$, test the airline's claim.

Claim
$p < 0.05$
$n = 420$
$x = 19$
$\hat{p} = \frac{x}{n}$
$\alpha = 0.01$

(a) (1 point) Which test (z-test, t-test or 1-prop-z-test) is appropriate here?

(b) (1 point) Write the symbolic form of the claim. $p < 0.05$

(c) (2 points) Write the null and alternative hypotheses.

$H_0: p \geq 0.05$

$H_A: p < 0.05$ (claim)

(d) (1 point) What number is the test statistic equal to?

$$\hat{p} = \frac{x}{n} = \frac{19}{420} \approx 4.5\%$$

(e) (1 point) What formula should be used for the the standardized test statistic?

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

(f) (1 point) What number is the standardized test statistic equal to? Round to the hundredths.

$Z \approx -0.45$ standard deviations

(g) (1 point) What p-value do you obtain? Round to the ten-thousandths.

p-value $= P(\hat{p} < 4.5\%) = 0.3272$

(h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

Since p-value > α, do not Reject Ho.

(i) (1 point) State a full sentence conclusion stating the decision you made.

There is not suff. sample evidence to support the claim

(18) A large software company gives job applicants a test of programming ability and the mean for that test has been 160 in the past. Twenty-five job applicants are randomly selected from one large university and they produce a mean score and standard deviation of 183 and 12, respectively. Use a 0.05 level of significance to test the claim that this sample comes from a population with a mean score greater than 160.

$n = 25$

$\bar{x} = 183$

$s = 12$

$\alpha = 0.05$

(a) (1 point) Which test (z-test, t-test or 1-prop-z-test) is appropriate here?

(b) (1 point) Write the symbolic form of the claim.

$\mu > 160$

(c) (2 points) Write the null and alternative hypotheses.

$H_0: \mu \leq 160$

$H_A: \mu > 160$ (claim)

(d) (1 point) What number is the test statistic equal to?

$\bar{x} = 183$

(e) (1 point) What formula should be used for the the standardized test statistic?

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

(f) (1 point) What number is the standardized test statistic equal to? Round to the hundredths.

$t \approx 9.58$

(g) (1 point) What p-value do you obtain? Round to the ten-thousandths.

p-value $= P(\bar{x} > 9.58) \approx 0.0000$

(h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

Since p-value $\leq \alpha$, reject H_0

(i) (1 point) State a full sentence conclusion stating the decision you made.

The sample data supports the claim.

23. A nationwide study of American homeowners revealed that 66% have one or more lawn mowers. A lawn equipment manufacturer, located in Omaha, feels the estimate is too low for households in Omaha. Test the claim that the proportion with lawn mowers in Omaha is higher than 66%. Among 500 randomly selected homes in Omaha, 68% had one or more lawn mowers.

(a) (1 point) Which test (z-test, t-test or 1-prop-z-test) is appropriate here?

$x = n \cdot p = 500(0.68) = 340$

(b) (1 point) Write the symbolic form of the claim.

$p > 0.66$

(c) (2 points) Write the null and alternative hypotheses.

$H_0: p \leq 0.66$

$H_A: p > 0.66$

(d) (1 point) What number is the test statistic equal to?

$\hat{p} = 0.68$

(e) (1 point) What formula should be used for the the standardized test statistic?

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

(f) (1 point) What number is the standardized test statistic equal to? Round to the hundredths.

$Z = 0.94$ (0.94406...)

(g) (1 point) What p-value do you obtain? Round to the ten-thousandths.

p-value = 0.1726

(h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

Since p-value > α, do not reject H_0

(i) (1 point) State a full sentence conclusion stating the decision you made.

There is not suff. sample evidence to support the claim

24. A fast food outlet claims that the mean waiting time in line is less than 3.8 minutes. A random sample of 20 customers has a mean of 3.7 minutes and standard deviation of 0.6 minute. If $\alpha = 0.05$, test the fast food outlet's claim.

$\mu < 3.8$
$n = 20$
$\bar{x} = 3.7$
$s = 0.6$
$\alpha = 0.05$

Key

(a) (1 point) Which test (z-test, t-test or 1-prop-z-test) is appropriate here?

(b) (1 point) Write the symbolic form of the claim.

$\mu < 3.8$

(c) (2 points) Write the null and alternative hypotheses.

$H_0: \mu \geq 3.8$
$H_A: \mu < 3.8$

(d) (1 point) What number is the test statistic equal to?

$\bar{x} = 3.7$

(e) (1 point) What formula should be used for the the standardized test statistic?

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

(f) (1 point) What number is the standardized test statistic equal to? Round to the hundredths.

$t = -0.75$ $(-0.74535\ldots)$

(g) (1 point) What p-value do you obtain? Round to the ten-thousandths.

p-value $= 0.2326$

(h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

Since p-val > α, do not reject H_0.

(i) (1 point) State a full sentence conclusion stating the decision you made.

There is not suff. sample evidence to support the claim.

Key

25.

3. A consumer group claims that the mean annual consumption of coffee by a person in the United States is 23.2 gallons. A random sample of 90 people in the U.S. has a mean annual coffee consumption of 21.6 gallons. Assume the population standard deviation is 4.8 gallons. Test the group's claim with $\alpha = 0.05$.

$n = 90$

$\bar{x} = 21.6$

$\sigma = 4.8$

$\alpha = 0.05$

(a) (1 point) Which test (z-test, t-test or 1-prop-z-test) is appropriate here?

(b) (1 point) Write the symbolic form of the claim.

$\mu = 23.2$

(c) (2 points) Write the null and alternative hypotheses.

$H_0: \mu = 23.2$

$H_A: \mu \neq 23.2$

(d) (1 point) What number is the test statistic equal to?

$\bar{x} = 21.6$

(e) (1 point) What formula should be used for the the standardized test statistic?

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

(f) (1 point) What number is the standardized test statistic equal to? Round to the hundredths.

$Z = -3.16$

(g) (1 point) What p-value do you obtain? Round to the ten-thousandths.

p-value $= 0.0016$

(h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

Since p-val $\leq \alpha$, reject Ho.

(i) (1 point) State a full sentence conclusion stating the decision you made.

There is suff. evidence to warrant rejection of the claim.

Appendix D — Labs

Lab 1

NAME: ______________________

Directions: Worth 70 total points. The lab is due at the beginning of class on Monday, otherwise you may be subject to a 15% late penalty for turning it in after I start class. Please staple your work to this cover sheet before you turn it in.

Use this data set below to complete exercises 1 through 8.

The following are the scores of 30 college students on a statistics test.

75	52	80	96	65	79	71	87	93	95
69	72	81	61	76	86	79	68	50	92
83	84	77	64	71	87	72	92	57	98

---

1. (5 points) Make a frequency distribution table of the test scores above ***using five classes***. Make sure to round your class width up to the next whole number.
2. (10 points) Make an Expanded Frequency Distribution (table) that includes midpoints, relative frequencies, cumulative frequencies and percentage cumulative frequencies.
3. (5 points) Make a relative frequency distribution table using the information in your Expanded Frequency Distribution .
4. (5 points) Make a cumulative frequency distribution table using the information in your Expanded Frequency Distribution .
5. (5 points) Graph the frequency histogram of the test scores by hand.
6. (5 points) Graph the relative frequency histogram of the test scores by hand.
7. (5 points) Graph the frequency polygon of the test scores by hand.
8. (5 points) Graph the ogive of the test scores by hand.

Use this data set below to complete exercises 9 through 13.

Data set: Triglyceride levels (in milligrams per deciliter of blood) of 26 patients

209	140	155	170	265	138	180	295	250
320	270	225	215	390	420	462	150	200
400	295	240	200	190	145	160	175	

Lab 1

9. (10 points) Construct an expanded frequency distribution table for the triglyceride levels of the 26 patients in the sample using five classes. Your table needs to include midpoints, relative frequencies, cumulative frequencies and percentage cumulative frequencies.

10. (5 points) Graph (by hand) a frequency histogram for the triglyceride levels data, using the information in your Expanded Frequency Distribution .

11. (5 points) Graph (by hand) a frequency polygon for the triglyceride levels data, using the information in your Expanded Frequency Distribution .

12. (5 points) Graph (by hand) an ogive for the triglyceride levels data, using the information in your Expanded Frequency Distribution .

VIDEO RESOURCES

Here is the web address that links to the video topics list below:
http://timbusken.com/labs.html

Video Topics List

- How to Enter Data and Sort data in the TI83/84+ Calculator
- What if I accidentally deleted L1, or some other list?
- How to Construct a Frequency Distribution Table with Interval Classes
- How to make the expanded Freq Distn
- How to Make a Frequency Distribution Table with a Data Set (by hand)
- How to sketch the graph of a Frequency Histogram
- How to sketch the graph of a Frequency polygon
- How to sketch the graph of a Relative Frequency Histogram
- How to Graph a Histogram with the TI-83+ and TI84+ Calculator

Lab 2

Lab2 — Calculator Lab NAME: ____________________

You may want to watch these two videos before you start this lab.

- How to run '1-VAR-STATS'
  youtube.com/watch?v=KKIijkkXoNM
- How to find the mean, stD and variance of a frequency distribution
  youtube.com/watch?v=E9_By3PZZ7M

Use this sample to complete this lab. The sample represents final exam scores from a geography class.

78 72 83 79 85 64 69 87 72 63 74

77 52 38 91 66 97 90 74 63 94 68 42

1. (2 points) Learn how to enter the data into your calculator's list environment. Learn how to have the calculator **sort the data** set in an ascending fashion. Write the sorted list of scores here. Identify the value representing the median.

1. ____________

Use the sample given above to find the value of each statistic below.

2. (2 points) mean — 2. ____________

3. (2 points) median — 3. ____________

4. (2 points) mode — 4. ____________

5. (2 points) standard deviation — 5. ____________

6. (2 points) variance — 6. ____________

7. (2 points) range — 7. ____________

8. (2 points) minimum — 8. ____________

Lab2 — Calculator Lab

9. (2 points) quartile 1

9. ____________

10. (2 points) quartile 2

10. ____________

11. (2 points) quartile 3

11. ____________

12. (2 points) maximum

12. ____________

13. (7 points) Sample data is summarized in the frequency distribution below. Find the mean of the distribution with the calculator.

13. ____________

class	frequency
0.0—0.9	23
1.0—1.9	54
2.0—2.9	78
3.0—3.9	42

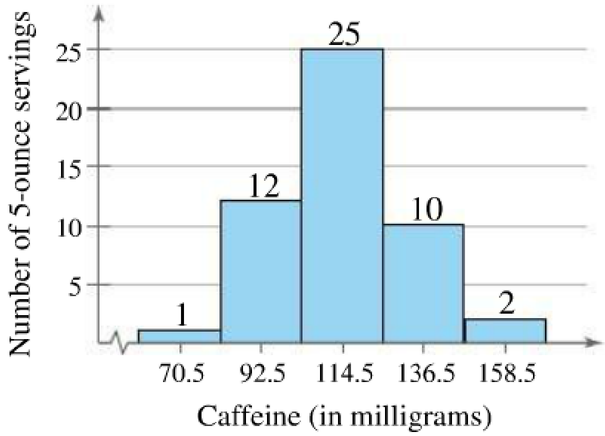
14. (7 points) Sample data is summarized in the frequency distribution below. Find the standard deviation of the distribution with the calculator.

14. ____________

class	frequency
0.0—0.9	23
1.0—1.9	54
2.0—2.9	78
3.0—3.9	42

Lab2 — Calculator Lab

15. The amounts of caffeine in a sample of five-ounce servings of brewed coffee are shown in the histogram. Use the calculator to find the mean, standard deviation and variance. ***Include the correct units in your answers!***.

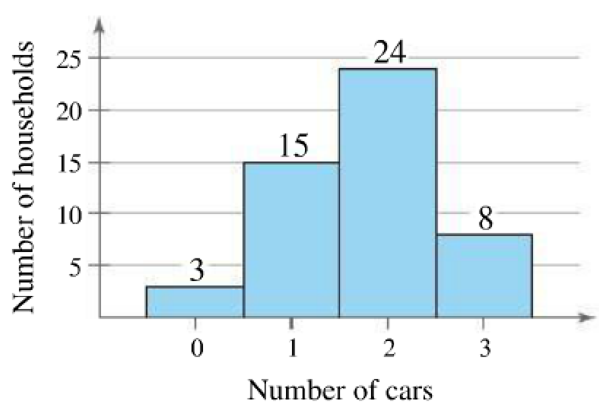


(a) (6 points) Find the mean (a) ______________

(b) (6 points) Find the standard deviation (b) ______________

(c) (6 points) Find the variance (c) ______________

16. The results from a random sample of the number of cars per household in a region are shown in the histogram. Use the calculator to find the mean, standard deviation and variance. ***Include the correct units in your answers!***.



(a) (6 points) Find the mean (a) ______________

(b) (6 points) Find the standard deviation (b) ______________

(c) (6 points) Find the variance (c) ______________

Lab2 — Calculator Lab

The data below is a sample of retirement ages of 24 randomly selected doctors in San Diego. (Section 2.1 Exercise 39)

70	54	55	71	57	58	63	65	60	66	57	62
63	60	63	60	66	60	67	69	69	52	61	73

Use the sample given above to find the value of each statistic below.

17. (2 points) mean — 17. ____________

18. (2 points) median — 18. ____________

19. (2 points) mode — 19. ____________

20. (2 points) standard deviation — 20. ____________

21. (2 points) variance — 21. ____________

22. (2 points) range — 22. ____________

23. (2 points) minimum — 23. ____________

24. (2 points) quartile 1 — 24. ____________

25. (2 points) quartile 2 — 25. ____________

26. (2 points) quartile 3 — 26. ____________

27. (2 points) maximum — 27. ____________

Lab 3

NAME: ____________________

1. (9 points) 100 students were surveyed this morning. One survey question was, "How would you rate your level of difficulty finding parking on campus this morning?" 8 students said 'nearly impossible,' 11 students said 'very difficult,' 26 students said 'difficult,' 25 students said they 'didn't park,' 20 students said 'easy,' and 10 students said 'very easy.' ***Sketch a hand-drawn pareto chart in the space below this question***.

Use the data to complete the rest of this lab***. The population data represents final exam scores from a geography class.***

$$31, 49, 19, 62, 50, 24, 45, 23, 51, 32, 48, 55, 60, 40, 35, 54, 26, 57, 37, 43,$$

$$65, 50, 55, 18, 53, 41, 50, 34, 67, 56, 44, 4, 54, 57, 39, 52, 45, 35, 51, 63, 42$$

2. (10 points) Sketch a hand-drawn stem-and-leaf plot of geography final exam scores ***in the space below this question***. Use one row per stem.

Lab 3

3. (1 point) Find the minimum

3. ____________

4. (2 points) Find Q_1

4. ____________

5. (2 points) Find Q_2

5. ____________

6. (2 points) Find Q_3

6. ____________

7. (1 point) Find the maximum

7. ____________

8. (2 points) Find the lower fence.

8. ____________

9. (2 points) Find the upper fence

9. ____________

10. (2 points) Are there any outliers? If so, list them in the answer blank for this question.

10. ____________

11. (9 points) ***Sketch a hand-drawn graph of a modified boxplot in the space below this question*** for the geography final exam scores data. Label the x axis numbers and the variable you're graphing on your graph.

12. (2 points) Find the mean. Round to the tenths decimal place value column.

12. ____________

13. (2 points) Find the standard deviation. Round to the tenths decimal place value column.

13. ____________

Lab 3

14. (2 points) What formula should be used to find z-scores for the data?

14. ____________

15. (2 points) What is the z-score of the data value: 60? Use your rounded values of the mean and standard deviation from above. Round the z-score to the hundredths.

15. ____________

16. (2 points) Is 60 an ordinary test score, unusual test score or very unusual test score?

16. ____________

17. (2 points) What is the z-score of the data value: 4? Use your rounded values of the mean and standard deviation from above. Round the z-score to the hundredths.

17. ____________

18. (2 points) Is 4 an ordinary test score, unusual test score or very unusual test score?

18. ____________

19. (20 points) Make a handwritten, two-column table with the data. List the sorted data in column one of your table. List the percentile rank of each data value in column two. *Round each percentile to the nearest whole number percent.*

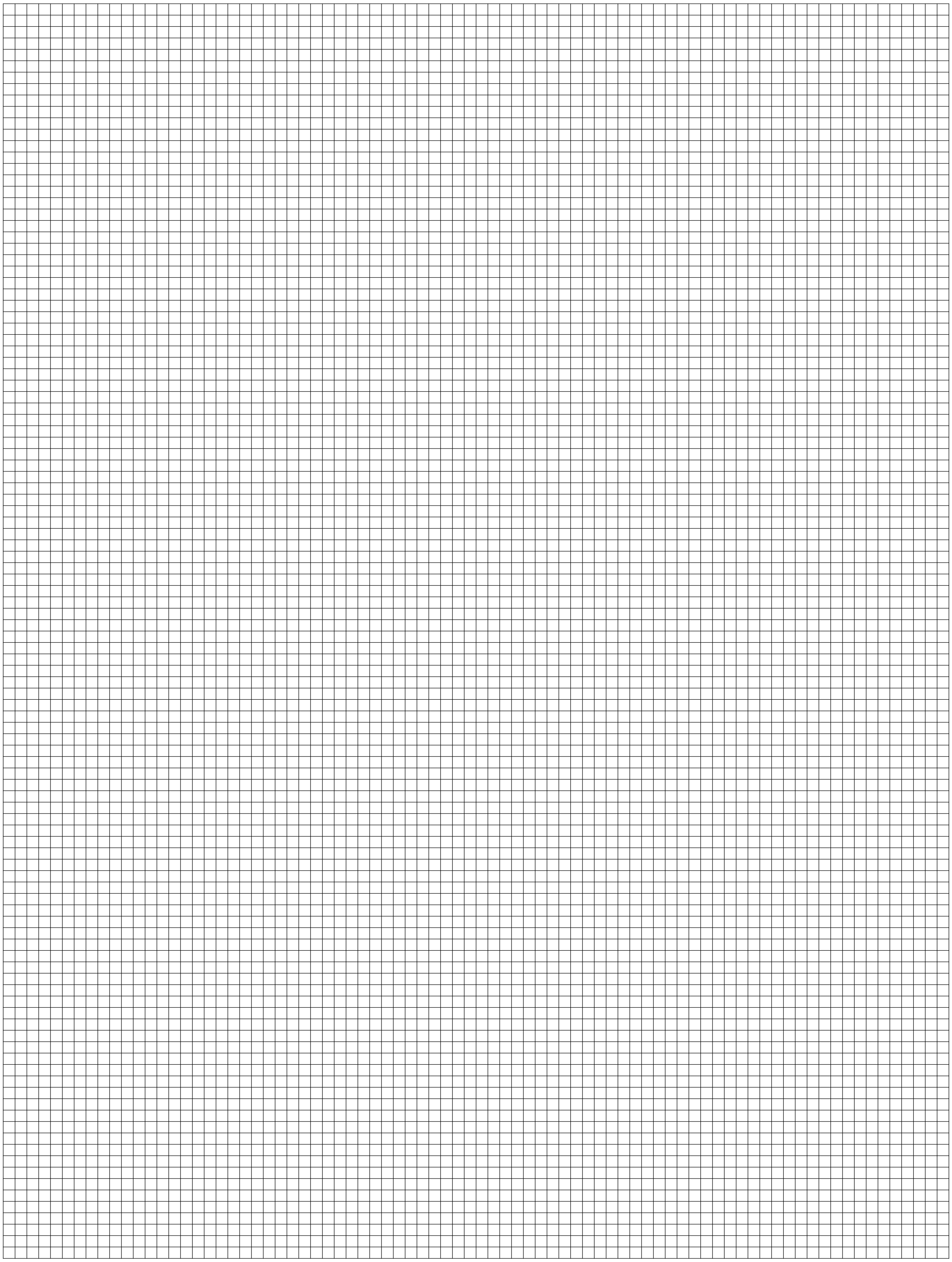
20. (20 points) Use your table of values to construct a hand-drawn percentile graph. Please use the graph paper I provided on the next page. Label the y-axis with the word 'PERCENTILE' and the x-axis with the phrase 'Geography Final Exam Scores.'

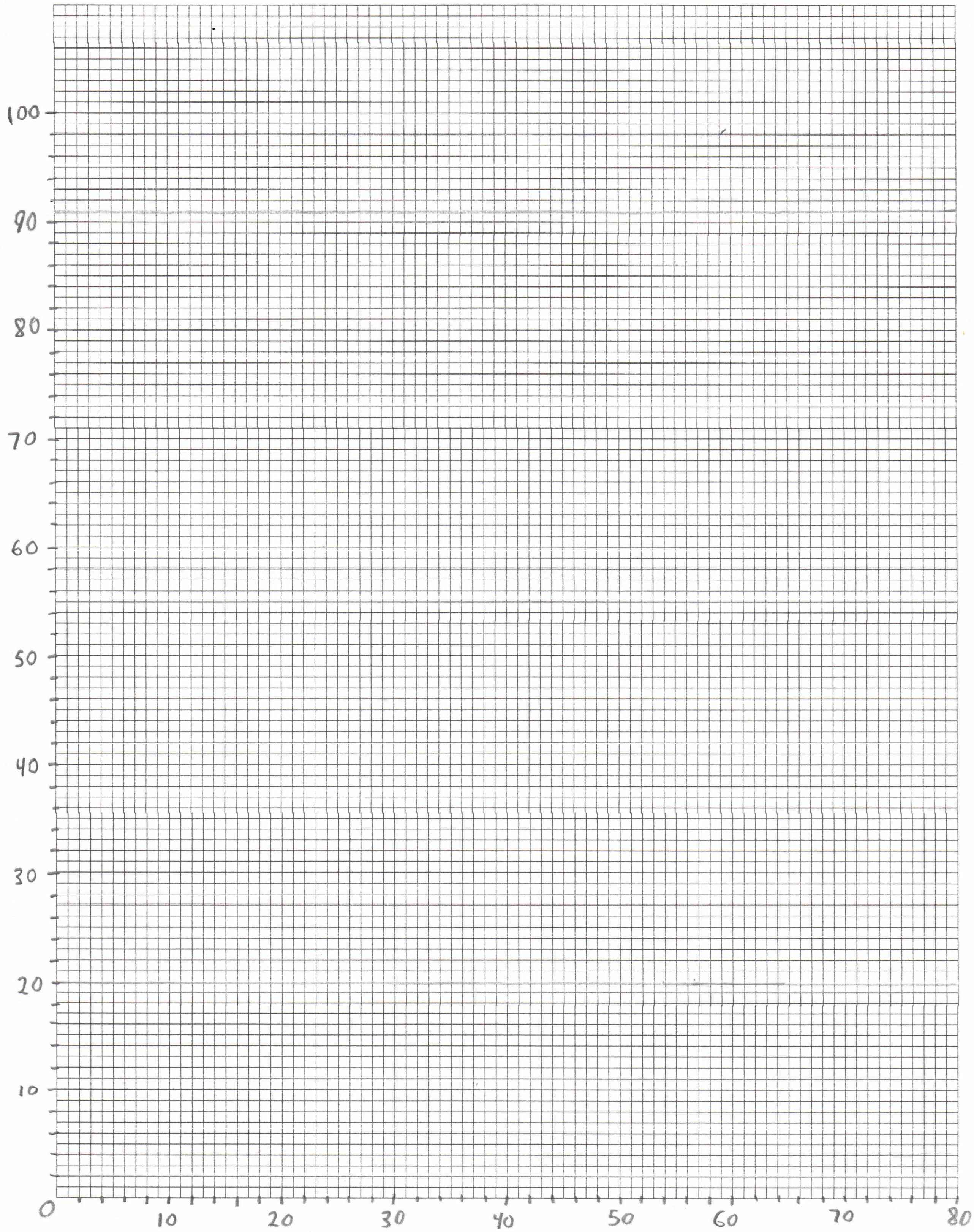
21. (1 point) What is the percentile rank of 60?

21. ____________

22. (3 points) What is the meaning/interpretation of the percentile rank of 60?

22. ____________




100
90
80
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70
80

Lab 4 — StatCrunch Lab NAME: ____________________

You will need to access these two excel files for this lab

- http://www.timbusken.com/2-1-39.xls
- http://www.timbusken.com/Survey1.xlsx

You should watch these two videos before you start this lab.

- How to get 'StatCrunch' software
  http://www.youtube.com/watch?v=dF-9BOx18WQ
- How to Load an Excel File into StatCrunch
  http://www.youtube.com/watch?v=FNw_9mDiCls

LOAD THE FIRST EXCEL FILE (named 2-1-39.xls) INTO STATCRUNCH. This video (link below) will show you exactly how to complete questions 1 through 9. After you complete questions 1 through 9, completing the rest of the lab should be self-explanatory. Start the video at 5:00 minutes to skip to 'how to do question 1.'

- https://www.youtube.com/watch?v=3FfF6kpRSqE&feature=youtu.be&t=4m59s

Lab 4: How to Make Tables and Sketch Graphs with StatCrunch

Name: ____________________

Directions:

1. Graphs and tables cannot be drawn by hand. Use StatCrunch or another software. Turn in these lab directions sheets along with your graphs and tables.
2. Graphs must have your first and last name on the title or you will receive no credit for the graph.
3. I will not accept a digital copy of the lab. You must print and turn in a physical copy if you want credit
4. The lab is due BEFORE class starts on the due date. Check the course calendar for the due date.

Lab Questions

1. Construct a **relative frequency histogram** for the data below. The data is a sample of retirement ages of 24 randomly selected doctors in San Diego. (Section 2.1 Exercise 39)

70	54	55	71	57	58	63	65	60	66	57	62
63	60	63	60	66	60	67	69	69	52	61	73

2. Use a **stem-and-leaf plot** that has two rows for each stem to display the data. The data represent the winning scores from Super Bowl I to Super Bowl XLVII. (Section 2.2 Exercise 22) You are not required to include your name on this graph.

35	33	16	23	16	24	14	24	16	21	32	27
35	31	27	26	27	38	38	46	39	42	20	55
20	37	52	30	49	27	35	31	34	23	34	20
48	32	24	21	29	17	27	31	31	21	34	

3. Use a **pie chart** to display the data. The data represents the results of an online survey that asked adults how they will invest their money in 2016. (Section 2.2 Exercise 25)

Invest more in stocks	562
Invest more in bonds	144
Hold on to more cash	288
Invest the same as last year	461

4. The numbers of tornadoes by state in 2012 are listed. Use statcrunch to make an **expanded frequency distribution** table (stats→tables→frequency) that includes frequencies, relative frequencies, cumulative frequencies and cumulative relative frequencies. Use 20 for the class width and use 0 for the lower class limit of the first class in your table.

87	0	0	29	19	26	0	1	40	25
0	2	39	33	20	145	65	53	1	17
0	7	39	75	32	4	48	1	0	1
3	8	17	8	18	41	0	15	0	10
10	37	114	1	1	16	0	2	3	6

5. Graph a **histogram** for the tornadoes data set. **Don't forget to label your graph's axes and include your name in the title of the graph.**

6. Is the data skewed left, skewed right, symmetric or neither symmetric nor skewed?

6. ____________

7. Which is larger, the median or the mode?

7. ____________

8. Graph a **dotplot** for the tornadoes data set. **Don't forget to label your graph's axes and include your name in the title of the graph.**

9. The table below shows the years of experience of 14 registered nurses and their annual salaries (in thousands of dollars). Construct a **scatter plot** with the years of experience graphed along the x-axis. Label your graph's axes and title.

years exp.	0.5	2	4	5	7	9	10	12.5	13	16	18	20	22	25
salary	40.2	42.9	45.1	46.7	50.2	53.6	54	58.4	61.8	63.9	67.5	64.3	60.1	59.9

LAB 4

NOW LOAD THE SECOND EXCEL FILE INTO STATCRUNCH

10. Make a dotplot with the 'hours of sleep' variable. 48 students were asked how many hours per night they slept.

11. What is the five number summary for the 'hours of sleep' variable?

    11. ____________

12. What is the average number of the 'hours of sleep' variable?

    12. ____________

13. What is the standard deviation of the 'hours of sleep' variable?

    13. ____________

14. What are the units on the standard deviation of the 'hours of sleep' variable?

    14. ____________

15. Make a scatter plot. Use the student 'height' variable for your x values and use the student 'shoe size' variable for your y values. Label your graph's axes and title with your name.

16. Make an expanded frequency table for the student 'height' variable that includes columns for frequency, percent of total, and cumulative percent of total.

17. Using the information in your table, what height (in inches) is in the 90th percentile?

    17. ____________

18. Using the information in your table, what percentage of the students are 66 inches tall?

    18. ____________

19. Make a stemplot for the variable 'Tim's Age'

20. Make a histogram for the variable 'Tim's Age'

END OF LAB 4

Lab 5 NAME: ____________________

Directions: In order to get full credit for lab, your lab report should be only four pages long (each page is worth 25% of the lab) and include the following:

1. This page is the lab cover sheet. Print this page and answer the questions on it using your summary statistics and graph.
2. Page 2: This should be a statcrunch graph of all five boxplots on a single graph. ***Your graph must include a title with your name on it, unless you don't want credit for the graph. Your graph must use fences and clearly indicate the location of outliers.***
3. Page 3: This page should be a printout of the summary statistics (***generated from statcrunch***) for each of the five areas of the factory
4. Page 4: This page be a typed, short paragraph explaining your recommendations about which factory areas workers must be provided with protective ear wear. Use NIOSH's recommendation that all worker exposures to noise be controlled below a level equivalent to 85 dBA for eight hours to minimize occupational noise induced hearing loss.
5. You will want to reference this webpage (http://www.timbusken.com/lab-BOXPLOTS.html) for background information and instruction on this lab.

Lab Questions

1. (5 points) Is the distribution of noise levels in area 2 positively skewed, negatively skewed, symmetric, or not skewed or symmetric? Use Pearson's index of skewness.

   1. ____________

2. (5 points) Is the distribution of noise levels in area 5 positively skewed, negatively skewed, symmetric, or not skewed or symmetric? Use Pearson's index of skewness.

   2. ____________

3. (5 points) Explain the purpose of the factory study.

4. (5 points) About 75% of noise level readings in Area 5 are above what amount of decibels?

   4. ____________

5. (5 points) What percent of noise level readings in Area 3 are above 89 decibels?

   5. ____________

Lab 6

Lab – Calculator Lab Name: ____________________

Learning Objectives: Find the Mean, Variance and Standard Deviation of a Discrete Probability Distribution using the Calculator. Find probabilities with a table.

VIDEO LINK: https://www.youtube.com/watch?v=cI8lCZoJr1c&t=0s

Include the units in each of your answers, wherever appropriate.

Baier's Electronics manufactures computer parts that are supplied to many computer companies. Despite the fact that two quality control inspectors at Baier's Electronics check every part for defects before it is shipped to another company, a few defective parts do pass through these inspections undetected. Let x denote the number of defective computer parts in a shipment of 400. The following table gives the probability distribution of x.

x	0	1	2	3	4	5
$p(x)$	.02	.20	.30	.30	.10	.08

1. Find the mean of the distribution given above. 1. ____________

2. Find the standard deviation of the probability distribution. 2. ____________

3. Find the variance of the probability distribution. 3. ____________

The maximum patent life for a new drug is 17 years. Subtracting the length of time required by the FDA for testing and approval of the drug provides the actual patent life of the drug—that is, the length of time that a company has to recover research and development costs and make a profit. Suppose the distribution of the lengths of patent life for new drugs is as shown here:

Years, x	3	4	5	6	7	8	9	10	11	12	13
$p(x)$	.03	.05	.07	.10	.14	.20	.18	.12	.07	.03	.01

4. Find the expected number of years of patent life for a new drug. 4. ____________

5. Find the standard deviation of the probability distribution. 5. ____________

6. Find the variance of the probability distribution. 6. ____________

LAB 6

The H2 Hummer limousine has eight tires on it. A fleet of 1300 H2 limos was fit with a batch of tires that mistakenly passed quality testing. The following table lists the frequency distribution of the number of defective tires on the 1300 H2 limos.

Number of defective tires	0	1	2	3	4	5	6	7	8
Number of H2 limos	59	224	369	347	204	76	18	2	1

7. ***Construct a probability distribution table*** for the numbers of defective tires on these limos. Let x denote the number of defective tires on a randomly selected H2 limo. List your table in the space below this question.

8. If you randomly select a limo from this fleet, how many defective tires would you ***expect*** it to have?

8. ____________

9. Find the mean of the probability distribution. 9. ____________

10. Find the standard deviation of the probability distribution. 10. ____________

11. Find the variance of the probability distribution. 11. ____________

Find the following probabilities.

12. $P(x = 3)$ 12. ____________

13. $P(2 \leq x \leq 4)$ 13. ____________

14. $P(x \geq 3)$ 14. ____________

15. $P(x > 4)$ 15. ____________

16. Would any of the probabilities from the previous four questions be considered unusual? Explain your reasoning.

LAB 6

The maximum patent life for a new drug is 17 years. Subtracting the length of time required by the FDA for testing and approval of the drug provides the actual patent life of the drug—that is, the length of time that a company has to recover research and development costs and make a profit. Suppose the distribution of the lengths of patent life for new drugs is as shown here:

Years, x	3	4	5	6	7	8	9	10	11	12	13
$p(x)$	.03	.05	.07	.10	.14	.20	.18	.12	.07	.03	.01

Find the following probabilities.

17. $P(5)$ 17. ____________

18. $P(x = 10)$ 18. ____________

19. $P(6 \leq x \leq 9)$ 19. ____________

20. $P(x \geq 5)$ 20. ____________

21. $P(x > 4)$ 21. ____________

22. $P(x > 10)$ 22. ____________

LAB 6

Years, x	3	4	5	6	7	8	9	10	11	12	13
$p(x)$	.03	.05	.07	.10	.14	.20	.18	.12	.07	.03	.01

Find the following probabilities.

23. $P(x \geq 6)$ 23. ____________

24. $P(x < 11)$ 24. ____________

25. $P(x < 7)$ 25. ____________

26. $P(x \leq 12)$ 26. ____________

Lab 7

Lab – Discrete Distributions   Name: ____________________

Learning Objectives: Find probabilities with a table. Find binomial probabilities. Find the Mean, Variance and Standard Deviation of a Binomial Distribution. Expected value problems.

Defective Parts Baier's Electronics manufactures computer parts that are supplied to many computer companies. Despite the fact that two quality control inspectors at Baier's Electronics check every part for defects before it is shipped to another company, a few defective parts do pass through these inspections undetected. Let x denote the number of defective computer parts in a shipment of 400. The following table gives the probability distribution of x.

x	0	1	2	3	4	5
$p(x)$	.02	.20	.30	.30	.10	.08

1. How many defective computer parts would you expect to receive in a shipment of 400?

1. ____________

Find each probability below.

2. What is the probability that a shipment of 400 computer parts has no defective parts?

2. ____________

3. What is the probability that at least one computer part in a shipment of 400 is defective?

3. ____________

4. What is the probability that less than four computer parts in a shipment of 400 are defective?

4. ____________

LAB 7

Hospital ER A review of emergency room records at a rural hospital was performed to determine the probability distribution of the number of patients entering the emergency room during a 1-hour period. The following table lists the distribution.:

Patients per hour, x	0	1	2	3	4	5	6	7	8	9	10
Probability, $p(x)$	.03	.05	.07	.10	.14	.20	.18	.12	.07	.03	.01

5. How many patients are expected to enter the emergency room during a 1-hour period?

5. ____________

6. What is the standard deviation of the number of patients entering the emergency room during a 1-hour period?

6. ____________

Find the following probabilities.

7. What is the probability that the number of patients entering the emergency room during a 1-hour period is at least two?

7. ____________

8. What is the probability that the number of patients entering the emergency room during a 1-hour period is not more than eight?

8. ____________

9. What is the probability that the number of patients entering the emergency room during a 1-hour period exceeds eight?

9. ____________

LAB 7

Hospital ER A review of emergency room records at a rural hospital was performed to determine the probability distribution of the number of patients entering the emergency room during a 1-hour period. The following table lists the distribution.:

Patients per hour, x	0	1	2	3	4	5	6	7	8	9	10
Probability, $p(x)$	.03	.05	.07	.10	.14	.20	.18	.12	.07	.03	.01

Find the following probabilities.

10. What is the probability that the number of patients entering the emergency room during a 1-hour period is less than nine?

10. ____________

11. Would it be unusual to have more than eight patients enter the emergency room during a 1-hour period? (Hint: use the 5% rule introduced in Section 3.1)

11. ____________

12. $P(x > 7)$

12. ____________

13. $P(x < 3)$

13. ____________

14. $P(5)$

14. ____________

15. Graph the probability distribution in the space below.

LAB 7

16. ***Binomial Tables*** Let X be a binomial random variable with $n = 5$ and $p = 0.30$. *Use the Binomial Tables* (Link Below) to obtain the correct probability distribution needed to answer probability questions.
http://timbusken.com/assets/statistics/binomPDF.pdf

(a) ***Write your distribution table in the space below.***

Find each probability.

(b) $P(X = 5)$ (b) ____________

(c) $P(X \geq 1)$ (c) ____________

17. ***College Degrees*** Nearly 40 percent of working-aged Americans now hold a college degree, according to a new report from the Lumina Foundation. Suppose a group of 32 Americans are randomly selected and asked if they hold a college degree or not. Let x denote the number of Americans in this sample of 32 who say that they hold a college degree. Obtain the binomial probability distribution table necessary to answer the questions given below. ***You do NOT have to write your probability distribution table on this paper.***

(a) How many college graduates would you expect to be in the group of 32?

(a) ____________

(b) What is the standard deviation of the number of college graduates in the group of 32?

(b) ____________

(c) What is the probability there are 2 college graduates in the group of 32?

(c) ____________

(d) What is the probability there are 3 or fewer college graduates in the group?

(d) ____________

(e) What is the probability that there are at least 6 college graduates are in the group?

(e) ____________

LAB 7

18. ***Marketing*** A fast food chain store conducted a taste survey before marketing a new hamburger. The results of the survey showed that 70% of the people who tried this hamburger liked it. Encouraged by this result, the company decided to market the new hamburger. Assume that 70% of all people like this hamburger. On a certain day, eight customers bought it for the first time.

    (a) Let x denote the number of customers in this sample of eight who will like this hamburger. Using the binomial probabilities table or your calculator, obtain and write down in the space below the probability distribution of x, and draw a graph of the probability distribution.

    (b) Determine the mean and standard deviation of x. (b) ____________

    (c) How many of the eight people who bought the burger would you expect to like the burger?

    (c) ____________

    (d) Using the probability distribution of part a, find the probability that exactly three of the eight customers will like this hamburger.

19. ***Basketball*** Michael Jordon is a retired American professional basketball player. Jordan played 15 seasons in the National Basketball Association for the Chicago Bulls and Washington Wizards. Jordon's career stats indicate that he makes 83.5% of the free throws he tries. Assuming this percentage will hold true for future attempts, find the probability that in the next ten tries, the number of free throws he will make is

    (a) exactly 9 (a) ____________

    (b) exactly 7 (b) ____________

http://timbusken.com/labs.html

Lab 9 NAME: ____________________

1. (3 points) Find the critical value z_c that is needed to set up a 93% confidence interval estimate for the population mean.

1. ____________

2. (3 points) Find the critical value t_c that corresponds to a 99% interval, assuming $n = 15$.

2. ____________

3. (5 points) The approximate costs for a 30-second spot for various cable networks in a random selection of cities are shown below. Estimate the population mean cost for a 30-second advertisement on cable network with 90% confidence. Assume the population of costs is approximately normal.

14	55	165	9	15	66	23	30	150
22	12	13	54	73	55	41	78	

3. ____________

4. (5 points) An alumni association wants to estimate the mean debt of this year's college graduates. It is known that the population standard deviation of the debts of this year's college graduates is \$11,800. How large a sample should be selected so that the estimate with a 99% confidence level is within \$800 of the population mean?

4. ____________

5. (5 points) A study of 35 golfers showed that their average score on a particular course was 92. The standard deviation of the population is 5. Find the 95% confidence interval of the mean score for all golfers.

5. ____________

6. (5 points) A recent study of 100 workers found that 63 took their lunch to work each day. Find the 90% confidence interval of the proportion of all workers who take their lunch to work each day.

6. ____________

7. (5 points) It is believed that 35% of U.S. homes have a direct satellite television receiver. How large a sample is necessary to estimate the population proportion of homes that have a direct satellite television receiver within 3 percentage points? Use 90% level of confidence.

7. ____________

8. (4 points) Find the critical values χ^2_L and χ^2_R needed to set up a 95% confidence interval for the population standard deviation when the sample size is 18. Assume the population is normally distributed.

8. ____________

9. (5 points) You randomly select and weigh a sample of 30 allergy medicine pills. The sample standard deviation is 1.2 milligrams. Assuming the weights are normally distributed, construct 99% confidence intervals for the population variance and standard deviation.

9. ____________

Lab 10

NAME: ________________________

1. When working properly, a machine that is used to make chips for calculators does not produce more than 4% defective chips. Whenever the machine produces more than 4% defective chips, it needs an adjustment. A factory worker who works next to the machine all day claims that the machine needs adjusting. To check if the machine is working properly, the quality control department at the company often takes samples of chips and inspects them to determine if they are good or defective. One such random sample of 200 chips taken recently from the production line contained 12 defective chips. Test the factory worker's claim that the machine is producing more than 4% defective chips. Use a level of significance equal to 1%.

   (a) (1 point) Which test is appropriate here?

   (b) (1 point) Write the symbolic form of the claim.

   (c) (2 points) Write the null and alternative hypotheses.

   (d) (1 point) What number is the test statistic equal to?

   (e) (1 point) What formula should be used for the standardized test statistic?

   (f) (1 point) What number is the standardized test statistic equal to? <u>Round to the hundredths.</u>

   (g) (1 point) What p-value do you obtain? <u>Round to the ten-thousandths.</u>

   (h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) (1 point) State a full sentence conclusion stating the decision you made.

2. The management of Priority Health Club claims that its members lose an average of 10 pounds or more within the first month after joining the club. A consumer agency that wanted to check this claim took a random sample of 36 members of this health club and found that they lost an average of 9.2 pounds within the first month of membership. The population standard deviation is known to be 2.4 pounds. Use $\alpha = 0.05$ to test the company's claim.

(a) (1 point) Which test is appropriate here?

(b) (1 point) Write the symbolic form of the claim.

(c) (2 points) Write the null and alternative hypotheses.

(d) (1 point) What number is the test statistic equal to?

(e) (1 point) What formula should be used for the standardized test statistic?

(f) (1 point) What number is the standardized test statistic equal to? <u>Round to the hundredths.</u>

(g) (1 point) What p-value do you obtain? <u>Round to the ten-thousandths.</u>

(h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

(i) (1 point) State a full sentence conclusion stating the decision you made.

3. Grand Auto Corporation produces auto batteries. The company claims that its top-of-the-line Never Die batteries are good, on average, for at least 65 months. A consumer protection agency tested 45 such batteries to check this claim. It found that the mean life of these 45 batteries is 63.4 months, and the standard deviation is 3 months. Find the p-value for the test that the mean life of all such batteries is less than 65 months. Test the company's claim using $\alpha = 0.025$.

   (a) (1 point) Which test is appropriate here?

   (b) (1 point) Write the symbolic form of the claim.

   (c) (2 points) Write the null and alternative hypotheses.

   (d) (1 point) What number is the test statistic equal to?

   (e) (1 point) What formula should be used for the standardized test statistic?

   (f) (1 point) What number is the standardized test statistic equal to? Round to the hundredths.

   (g) (1 point) What p-value do you obtain? Round to the ten-thousandths.

   (h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) (1 point) State a full sentence conclusion stating the decision you made.

4. The makers of Flippin' Out Pancake Mix claim that one cup of their mix contains 11 grams of sugar. However, the mix is not uniform, so the amount of sugar varies from cup to cup. One cup of mix was taken from each of 24 randomly selected boxes. The sample variance of the sugar measurements from these 24 cups was 1.47 grams squared. Assume that the distribution of sugar content is approximately normal. Use $\alpha = 0.05$ to test the claim that the population variance is greater than 1 gram squared.

   (a) (1 point) Which test is appropriate here?

   (b) (1 point) Write the symbolic form of the claim.

   (c) (2 points) Write the null and alternative hypotheses.

   (d) (1 point) What number is the test statistic equal to?

   (e) (1 point) What formula should be used for the standardized test statistic?

   (f) (1 point) What number is the standardized test statistic equal to? <u>Round to the hundredths.</u>

   (g) (1 point) What p-value do you obtain? <u>Round to the ten-thousandths.</u>

   (h) (1 point) Do you reject the null hypothesis or fail to reject the null hypothesis? Explain.

   (i) (1 point) State a full sentence conclusion stating the decision you made.

Appendix F — Statistics Jeopardy

http://timbusken.com/jeopardy/stats1.html