AP Statistics is a college level course covering material traditionally taught in an introductory statistics class. It will probably be unlike any other math course you have taken. The major topics covered by this course include Exploring Data, Sampling and Experimentation, Anticipating Patterns, and Statistical Inference. However, we will be using communication skills (oral and written) in conjunction with the math to interpret our calculations. We will also be continually looking at how statistics is applied and the relationships between all parts of the statistical process.

**To get ready for the rigorous work required by this course, this packet consists of review topics studied in the prerequisite math courses with some introduction of written communication. There are many examples of each type of problem. You should try each type of problem and mark those problems you still have questions over. We will take some time during each class in the first week to go over some problems from this packet. Because we also must begin on the Statistics curriculum right away, this time to review will be limited. All students are encouraged to come in before and after school for any additional help on this packet. Even though this packet will not be collected, the first assessment for this course will be over these topics. (The test over these pre-requisite skills will be given by the end of the second week of classes.)

Graphing calculators and other technology will be used to better understand these processes and analyze results more effectively. This AP exam allows you to use a calculator for the entire exam. So, we will be learning how to use our graphing calculators with every lesson. You are strongly encouraged to get one of your own (preferably a TI-83 or TI-84). The school has a limited number of graphing calculators for check out.

All students will have the opportunity to take the AP Statistics AB Exam in May for a fee. Some colleges will award college credit depending on the AP Exam score. This class is also available for concurrent enrollment credit for a fee. You are encouraged to research different Advanced Placement and concurrent credit policies of universities and colleges for more information.

If you feel you need help with some of these topics, the best resource may be the internet. There are great sites where you can just type in the topics to get help. There are some sites listed below.

Some websites for reference:

https://www.khanacademy.org/

http://www.stattrek.com

https://www.purplemath.com/modules/index.htm

https://www.coolmath.com/algebra

Online Calculators:

desmos.com

Algebra 2 Book:

Algebra 2 Textbook

I. BACKGROUND ALGEBRA SKILLS

A. Solving Equations

a) Using the formula $z = \frac{x-\mu}{\sigma}$, solve for

i. $z \; when \; x = 200, \mu = 188, and \; \sigma = 24$

ii. $\mu \; when \; z = 2, x = 135, and \; \sigma = 16$

iii. $x \; when \; z = -3.35, \mu = 60, and \; \sigma = 4$

iv. $\sigma \; when \; z = 2.5, x = \; 100, and \; \mu = 102$

b) Solve for y in each equation shown below when $x = 2$.

i. $\sqrt{y} = -2x + 11$

ii. $log \; y = \; 1.2 + 0.8x$

iii. $\frac{1}{y} = 3x - 14$

c) Using the formula $b_0 = \bar{y} - b_1\bar{x}$, solve for

i. $b_0 \; when \; \bar{y} = 23, b_1 = 4 \; , and \; \bar{x} = 2.5$

ii. $b_1 \; when \; b_0 = 752, \bar{y} = 11, and \; \bar{x} = 114$

d) Using the formula $ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$, solve for

i. $ME \; when \; z^* = 2.56, \hat{p} = 0.75, \hat{q} = 0.25, and \; n = 400$

ii. $n \; when \; ME = 0.05, \; z^* = 1.96, \hat{p} = 0.5 \; and \; \hat{q} = 0.5$

B. Making Scatterplots

For each set of data, make a scatterplot. Make sure to keep the context of the data by labeling the axes and choosing an appropriate title for the graphs. You should have consistent scales for both axes.

1. Each student measured the length of his/her right foot and height. The results are recorded below.

Foot Length (cm)	Height (cm)
24	159
22	148
19	126
23	157
20	138
24	162
28	180
25	161
17	122
24	155
26	173
22	146

2. Tread depths of tires were measured for different distances. The results are recorded below.

Distance Driven (1000s of Miles)	Tread Depth (mm)
17	5.3
41	1.2
25	4.0
4	6.8
34	2.3
22	4.1
48	0.3
36	1.9
13	5.7
30	3.3
45	1.9
9	6.0

C. Writing Equations of Lines in Slope-Intercept Form

Find the slope of the lines that goes through the two given points.
1) (3,1) and (9,5)

2) (1,1) and (-2,7)

3) (-4,-3) and (8,0)

4) (-1,4) and (-4,-5)

5) (0,-4) and (-4,6)

6) (-6,3) and (6,-1)

Write the equation in slope-intercept form of the line that goes through the given points.
1) (-1,-6) and (-3,-8)

2) (-1,3) and (3,5)

3) (-2,5) and (1,-7)

4) (1/2, 2) and (-3/2, 4)

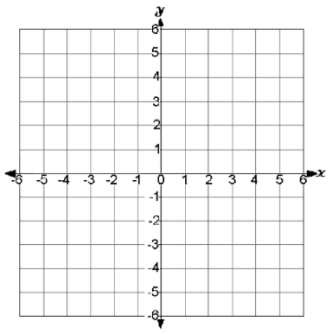
5) (1,3) and (-7,-3)

6) (-3,8) and (0,0)

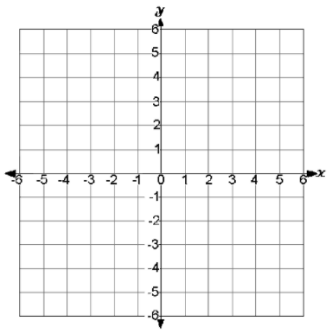
D. Graphing Lines

Sketch the graph of each line.

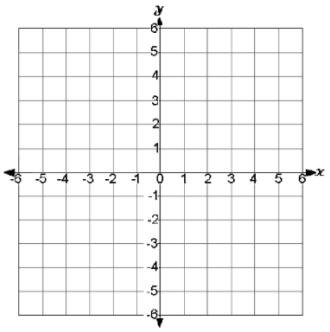
1). $y = \frac{1}{2}x + 4$



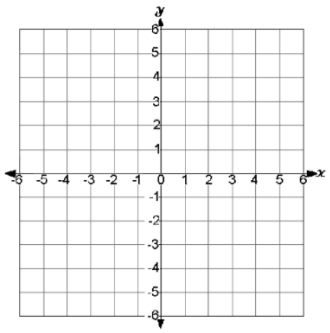
2) $y = -\frac{5}{3}x + 1$



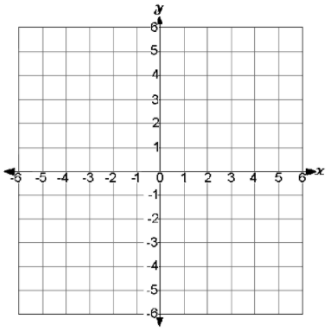
3) $y = -3x$



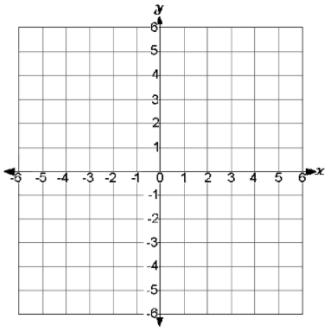
4) $y = 4$



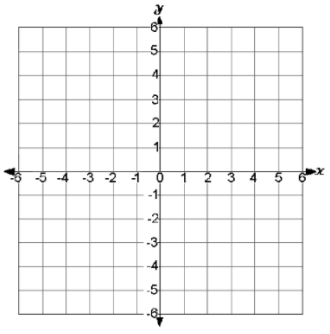
5) $y = \frac{9}{5}x + 2$



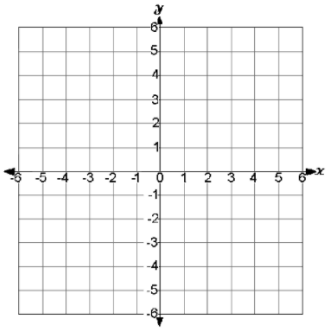
6) $y = -x + 5$



7) $y = -\frac{3}{2}x$



8) $y = 2x - 5$



E. Solving Systems of Equations

Using the Elimination Method, solve each system of equations. Make sure to write your answers as coordinates.

1) $\begin{matrix} 3x + 2y = 11 \\ 7x - y = 3 \end{matrix}$

2) $\begin{matrix} 3x - 4y = 18 \\ x + 3y = -7 \end{matrix}$

3) $\begin{matrix} 5x + 2y = -8 \\ 9x - 4y = -22 \end{matrix}$

4) $\begin{matrix} x - 5y = 15 \\ 4x - 3y = 26 \end{matrix}$

5) $\begin{matrix} 2x + 5y = 11 \\ -3x + 8y = -1 \end{matrix}$

6) $\begin{matrix} 7x - 3y = 2 \\ 5x + 4y = -17 \end{matrix}$

II. TYPES OF DATA, DISPLAYS, AND LISTS

A. Types of Data

There are two types of data: **qualitative** (or categorical) and **quantitative**.

Qualitative variables or categorical variables are variables that categorize individuals (place them in groups). These variables may take on values that are labels for categories. Examples are eye color (blue, hazel, etc.), gender (male or female), method of transportation to school (bike, car, bus, etc.), class rank (senior, junior, etc.).

Quantitative variables are numerical variables that represent an amount or quantity. There are two kinds of these: **discrete** and **continuous**.

Discrete variables are quantitative variables that assume only a countable number of values. Examples of these include shoe size (…, 6, 6 ½ , 7, 7 ½ , …), score on a test, class size, number of cans collected for MIFA.

Continuous variables are quantitative variables that can assume an infinite number of values. In the case of continuous variables, the values can generally assume any decimal quantity within a small range of values (even though we may round the answer like when we measure our height). These are typically values that result from some kind of measurement. The units of measurement are pounds/ inches/ Kelvin/ degrees/ feet/ etc. Examples are height, weight, surface area of oranges, era in baseball (3.23, 2.78, etc.), GPA.

Just because your variable's values are numbers, don't assume that it's quantitative. For example, 9, 10, 11, and 12 are labels for different class rankings at PWH. Class rank is a qualitative variable (even though it may be answered with a 9, 10, 11, or 12). Social security number is another example of a numerical output that is not a quantitative variable. SSN doesn't stand for any type of numerical quantity (you are not the 412,327,642nd person born in the US!). Phone number is not a quantitative variable either. The 719 area code is a designation for a geographic region; it is not a numerical quantity.

1) For the examples below, identify each as either categorical or numerical.

a) Time it takes to get to school
b) Number of people under 18 living in a household
c) Hair color
d) Temperature of a cup of coffee
e) Teacher salaries
f) Gender
g) Height
h) Amount of oil spilled
i) Age of Oscar winners
j) Jellybean flavors
k) Zip codes
l) Types of shoes owned
m) In which year did you take Algebra I?
n) Where did you eat your last meal? (1=home, 2=restaurant, 3=other)
o) How many AP classes will you be taking this year?
p) What grade did you earn in your last math class: A, B, C, D, or F?

B. Categorical Displays

There are several types of displays used for categorical data as described below:

- Frequency Table-a table that summarizes data by displaying counts
- Relative Frequency Table- a table that summarizes data by showing proportions or percents
- Bar Graph – one axis is the values of your variable (order does not matter) and the other axis is the frequency or relative frequency (percentages). The bars do not touch.
- Pie Chart – the whole group of values is shown as a circle. Each "piece of the pie" corresponds to the relative frequency of the values.
- Two-Way Tables (Contingency)-used for overlapped categories where one variable is contingent upon another

<u>More on these this fall</u>

- Pictographs
- Stacked Bar
- Mosaic Plots

All displays must include context and units.

A physical therapist wants to get a sense of the types of rehabilitation required by her patients. She gets a simple random sample of 20 patients and records the body part requiring rehabilitation.

Back	Back	Knee	Elbow	Knee
Knee	Back	Elbow	Hip	Knee
Elbow	Back	Back	Back	Elbow
Back	Hip	Back	Knee	Knee

Begin the Frequency Table by tallying the data and then summarize each part with a count.

Body Parts Needing Rehab	Tally	Frequency
Back	𝍸 III	8
Hip	II	2
Knee	𝍸 I	6
Elbow	IIII	4

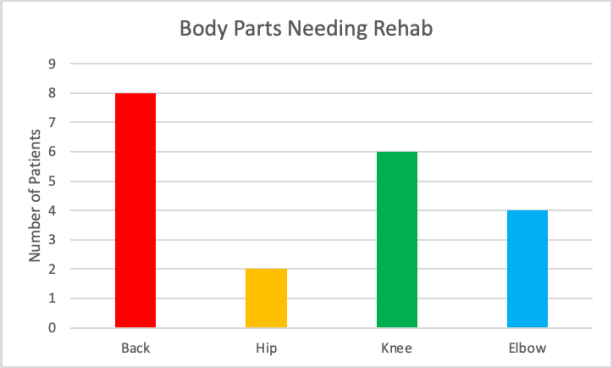
Relative Frequency Tables should compare each count to the total number.

Body Parts Needing Rehab	Relative Frequency
Back	**8/20**
Hip	**2/20**
Knee	**6/20**
Elbow	**4/20**

The relative frequencies should add to exactly 100%. If not, there could be overlapping categories on a category might be left out.

Frequency Bar Graph

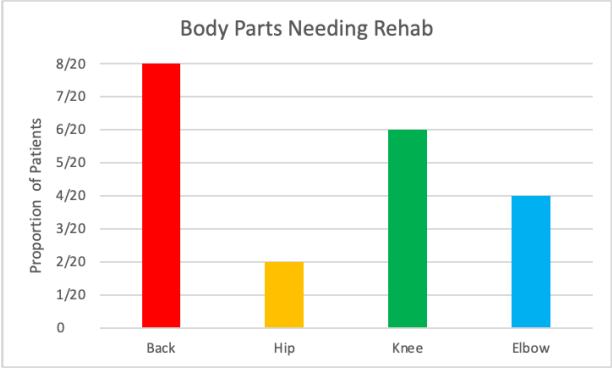
The context of this data is found in the title as well as the *x*- and *y*-axis labels.



Bars shouldn't touch on a bar graph. The categories are separate so the bars are separate.

Relative Frequency Bar Graph

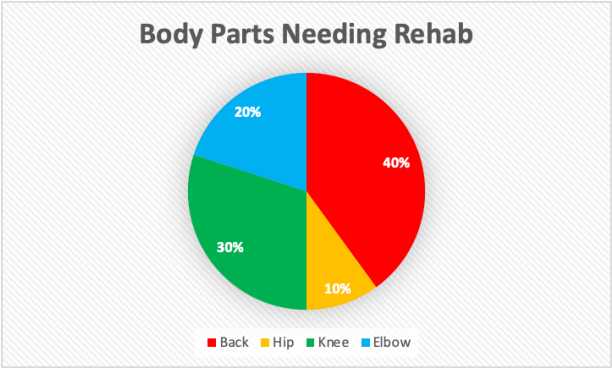
This might look like the same as the frequency bar graph but notice the scale and label on the *y*-axis.



Don't use pie charts if there a lot of categories. The display will be too busy and difficult to read.

To determine the central angle that forms each "piece" use the following:

central angle = (relative frequency) (360°)



Back
$^{8}/_{20} \times 360° = 144°$
Hip
$^{2}/_{20} \times 360° = 36°$
Knee
$^{6}/_{20} \times 360° = 108°$
Elbow
$^{4}/_{20} \times 360° = 72°$

Be careful not use 3-dimensional pie charts because they could be misleading.
Two-Way Tables (Contingency)
The table below is comparing a student's grade level and the type of lunch they eat at school.

	Grades 3-5	Grades 6-8	Total
Hot Lunch	18	22	40
Cold Lunch	17	33	50
Total	35	55	90

If you randomly select a student, what are the chances the student is from Grades 6-8 and eats hot lunch? $\frac{22}{90} \approx 24.4\%$

What percent of students eat cold lunch? $\frac{50}{90} \approx 55.6\%$

Of the hot lunch eaters, what proportion of students are from Grades 3-5? $\frac{18}{40}$

2) Twenty students were selected at random and asked what their favorite class is. The results are shown below.

English	History	History	Math	Science
Math	English	Spanish	History	Science
Spanish	Science	English	Math	Math
History	Math	English	History	Science
Science	History	Math	Science	Math

For this data make a frequency table, relative frequency table, frequency bar graph, and a relative frequency bar graph.

3) For the data shown below, create a pie chart and a bar chart. Make sure to add context using a title and labels.

In 1997, there were 92, 353 deaths from accidents in the United States. Among those were 42,340 deaths from motor vehicle accidents, 11,858 from falls, 10,163 from poisoning, 4051 from drowning, and 3601 from fires. The rest were listed as "other" causes.

4) People were asked whether they preferred apples or bananas. Each person's age and their response is recorded in the table below.

	Prefers apples	Prefers bananas	Total
Less than 13 years old	10		27
13 years old or older		10	23
Total		27	

Find the missing values of the table and answer the following questions:

a) What percent of respondents were less than 13 years old and liked bananas?

b) What proportion of respondents preferred apples?

c) Of those who were 13 years old or older, what percent preferred apples?

B. Data and Lists

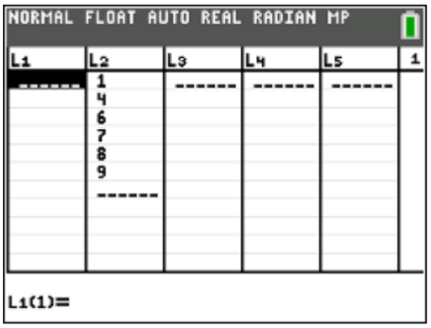
Quantitative data can be stored in lists on the TI-84 calculator. There are several ways to create a list. From the home screen, braces can be used to define a data set, which then can be stored in one of the list names L1 through L6 (see figure 1.1) It seems much easier to use the calculator's editing abilities.

From the STAT menu, choose EDIT. You may now enter the data into any of the lists (figure 1.2)

Figure 1.1

$\{1,4,6,7,8,9\} \rightarrow L_2$

Figure 1.2:



In either case, new lists can be created from existing lists, such as L_2+5 (figure 1.3 and 1.4) Make sure when you enter the command that you are on the L_3 name and not within the list of numbers. Press ENTER after entering this formula. (L_2 can be entered by pressing 2^{ND} and the number 2 on your calculator.)

Figure 1.3

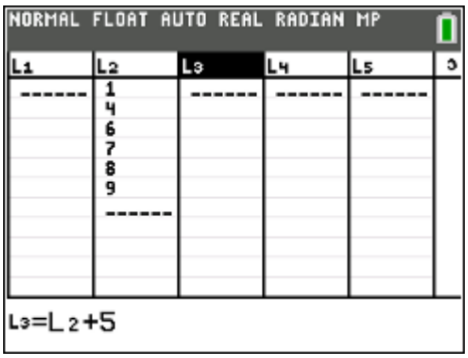
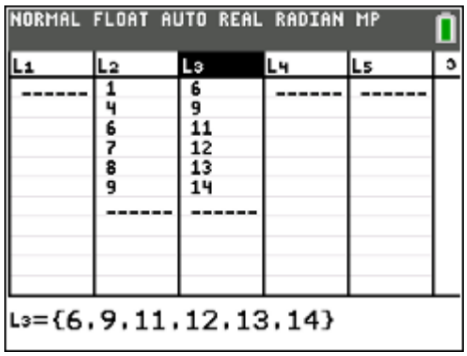


Figure 1.4



Lists may also be given their own names and will be retained in memory until deleted. This is particularly useful for data that will be used repeatedly.

Example 1: Create a named list for the following set of running speeds in mph for various animals:

Cheetah	70	Warthog	30
Lion	50	Cat	30
Coyote	43	Man	27.89
Hyena	40	Pig	11
Greyhound	39.35	Tortoise	.17
Rabbit	35	Snail	.03

To create a named list go to the list editor (STAT, EDIT) and move to the right past L_6. A "name" prompt will appear and the list name can be typed (figure 1.5). The alpha lock should be on. The values can be entered in the usual way (figure 1.6).

Figure 1.5

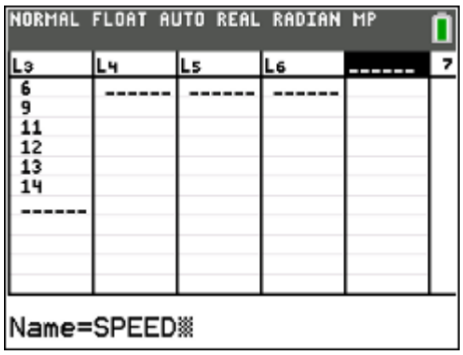
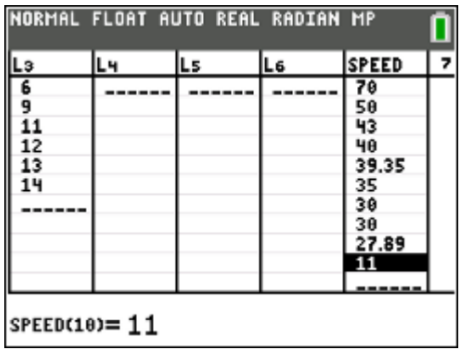


Figure 1.6



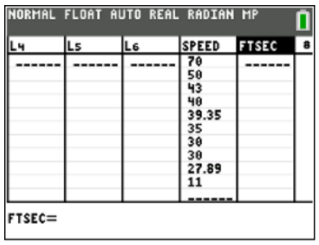
Example 2: Create a new list showing the speeds in feet per second.

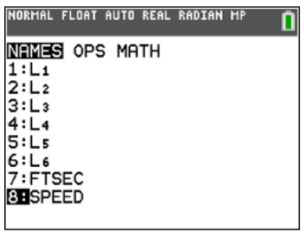
New lists can be created from named lists on either the home screen or in the list editor. On the home screen, the speeds in mph can all be converted to ft/sec and stored in a list named FTSEC by a single command (figure 1.7). In order for the TI-84 to distinguish a user defined list name from other symbols it is necessary to preface a list name with a special character L that is located in the LIST: OPS sub-menu. The L character may also be found in the CATALOG. First create a list named FTSEC using the steps outlined above.

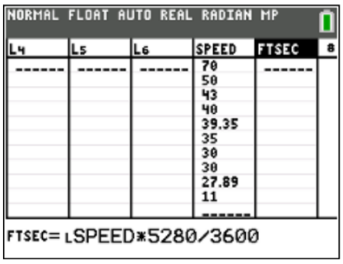
From the home screen:

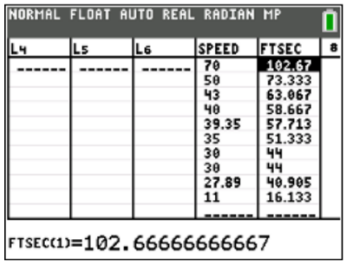
Figure 1.7: $_L$SPEED *5280 / 3600→ $_L$FTSEC

From the list editor: Highlight the list name icon FTSEC, then go to the LIST: Names (2nd STAT) , pick the list named SPEED, multiply by 5280 and divide by 3600, Press ENTER.









To delete a list, place your cursor over the list name in the List Editor. Press CLEAR and ENTER. If you accidentally press DEL and ENTER, the list will disappear but it isn't gone. You will have to go to the home screen, then STAT, SetUp Editor. Ex. SetUpEditor L_1, L_2, L_3, L_4, L_5, L_6 and then press ENTER. The home screen will say Done but if you go back into your List Editor, your list should be back.

To delete a value within a list, place your cursor over the value and then press DEL. If you need to insert a value (before), place your cursor within the list and press 2^{ND}, INS. The calculator will automatically put a zero until you change the value.

5) Create a list L_1 using {4,7,9,11,14,17,20}
From that, create new lists using any method.
a)$L_1 - 7$: _____
b) $2 * L_1$: _____
c) L_1^2: _____
d) $\sqrt{L_1}$: _____
e) $\ln L_1$: _____

6) Create a list PLANM showing the mean distance from the sun in millions of miles for each planet. Create a new list PLANK showing the mean distance in millions of kilometers. (How do you convert from miles to kilometers?)

Planet	Mean distance in millions of miles	Mean distance in millions of kilometers
Mercury	36.0	
Venus	67.24	
Earth	92.9	
Mars	141.71	
Jupiter	483.88	
Saturn	887.14	
Uranus	1783.98	
Neptune	2796.46	

C. Quantitative Displays

There are four different types of displays for quantitative data described below:

- Frequency Table-just like with categorical data, frequency tables summarize numerical data with counts
- Relative Frequency Table-just like with categorical data, relative frequency tables summarize numerical data by showing proportions or percents
- Histogram-one axis is the values or range of values of your variable and the other axis is the frequency or relative frequency. The bars do touch.
- Stem and Leaf Chart (Stemplot)-separates data by place values and displays the original values
- Dotplot-displays data on a number line, the number of dots on a value shows the frequency of that value
- Boxplot-based on a 5-number summary, a boxplot shows the distribution of a data set by 25% increments, could display outliers

<u>More on this in the fall</u>

- Ogives

We can use our calculators to show many of these displays. Try practicing the calculator steps as you go through this section.

Frequency Table

Old Faithful Eruptions (in seconds) in Lesser Known Calistoga, CA

108	108	99	105	103	103	94
102	99	106	90	104	110	110
103	109	109	111	101	101	
110	102	105	110	106	104	
104	100	103	102	120	90	
113	116	95	105	103	101	
100	101	107	110	92	108	

To make a frequency table for quantitative data, we need to create bins (widths) to help sort the data. When creating bins, the minimum number of bins should be 5. All bins should be the same.

Notice that the minimum is 90 seconds and the maximum is 120 seconds. There really is $120 - 90 + 1 = 31$ seconds represented here. We will make a frequency table with 5 bins. Since 31 is not divisible by 5, then we will round this up to 35 since 35 is divisible by 5.

$$35\ values \div 5\ bins = 7\ seconds\ in\ each\ bin$$

Starting with 90 and counting up 7 numbers, our bins will be 90-96, 97-103, 104-110, 111-117, 118-124. Notice the second bin starts with 97 not 96. If 96 was a value in our data set, we need one specific place to put this value, not two.

Bins	Tally	Frequency
90-96	𝍸	5
97-103	𝍸 𝍸 𝍸 I	16
104-110	𝍸 𝍸 𝍸 IIII	19
111-117	III	3
118-124	I	1

Just like we did with the Frequency Table for categorical data, we record a tally for each value in the data set and then summarize with a count.

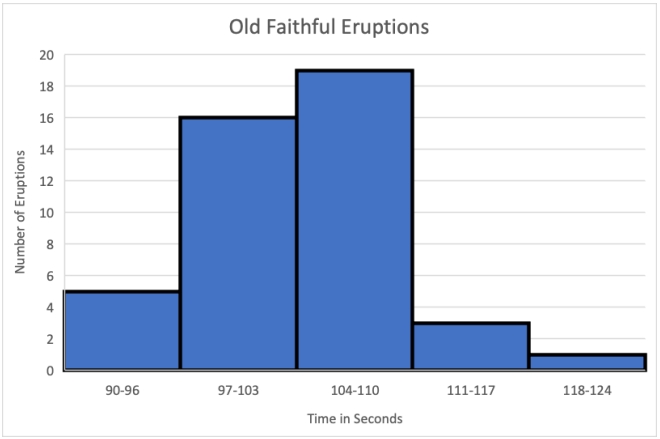
Relative Frequency Table

Bins	Relative Frequency
90-96	5/44
97-103	16/44
104-110	19/44
111-117	3/44
118-124	1/44

Just like we did with the Relative Frequency Table for categorical data, we record a tally for each value in the data set and then summarize with a proportion or a percent.

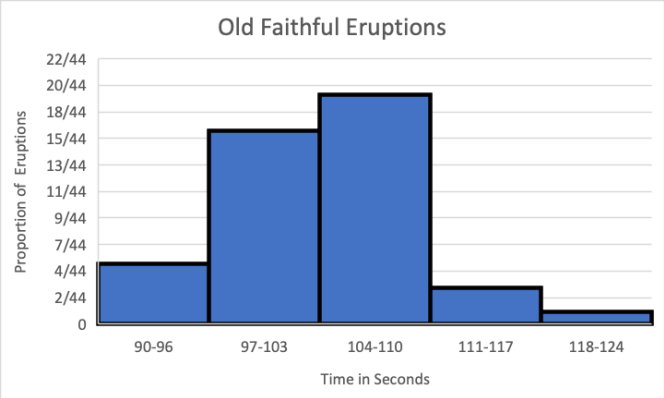
Frequency Histogram

Bars should touch on a histogram. The number line is continuous.



Relative Frequency Histogram

This might look like the same as the frequency histogram but notice the scale and label on the *y*-axis.

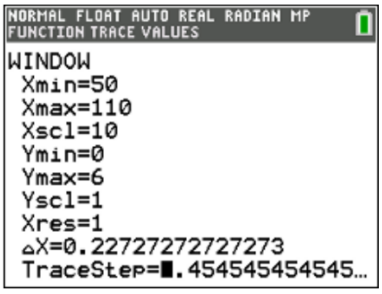


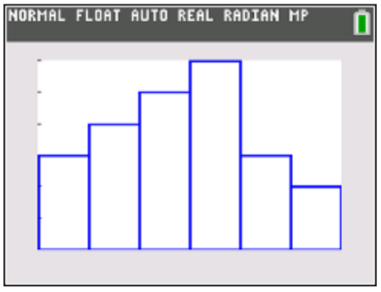
On the TI-84

Enter the following test scores into List 3 and create a histogram using intervals of width 10:

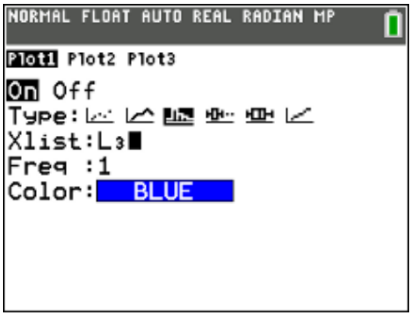
100, 100, 98, 95, 92, 88, 85, 84, 84, 81, 80, 77, 75, 72, 70, 70, 66, 65, 63, 60, 59, 55, 50

First enter the scores into L3. Then turn on your STATPLOT. Click the 2ND function of the Y= button to get to the menu. Choose one of the Plots by pressing 1, 2, or 3. Now activate the plot by selecting ON. At the TYPE prompt choose the third icon for histogram. At the X LIST, make sure it says L3 (2ND function on the numbers 1 to 6). Frequency should be 1. To control the width of the bars (called the bin width), set up the window manually. Xscl=10. Press GRAPH.





You may now TRACE the graph to see the intervals and their frequencies. When drawing a histogram, make sure both axes are labeled and have appropriate scales.



Stemplot

To make a stemplot, you will need to split the data values by place values. This could be tens and ones or ones and decimal places for example. These values do not necessarily need to be recorded from least to greatest. Make sure the spacing is consistent throughout the leaves. (Zeros should occupy as much space as a one.). If there are no values with a particular stem, do not skip stems. The leaves should be left blank. Depending on the size of the data, you can split the stems.

The data set below was recorded at a watering hole. The number of birds was recorded each hour.
13, 14, 16, 17,18, 12, 27, 28, 23, 25, 28, 32, 35, 39, 44, 46, 59, 67

Number of Birds at a Watering Hole Each Hour

Stems (tens)	Leaves (ones)
1	3 4 6 7 8 2
2	7 8 3 5 8
3	2 5 9
4	4 6
5	9
6	7

Key: 1|3 = 13 birds

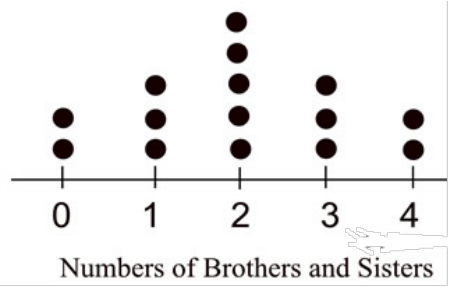
Notice the title and the key adds context for this stemplot. You only need to explain one value from the stemplot with the correct units.

Dotplot

Notice how a dot is plotted for each data value above the number line. Dots should be given consistent space. If you don't have a value on the number, don't skip it.

Numbers of Brothers and Sisters

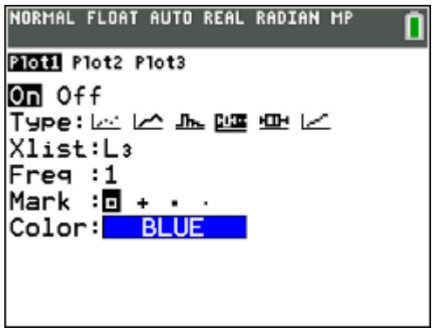
0, 0, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4



Boxplot
Using the same data from the histogram we made on page 17 enter the following test scores into List 3

100, 100, 98, 95, 92, 88, 85, 84, 84, 81, 80, 77, 75, 72, 70, 70, 66, 65, 63, 60, 59, 55, 50

and create a boxplot. Make sure your axis has a scale and label. In the STATPLOT menu, you will notice two boxplot options. The first one allows for possible outliers. I recommend you always use this one.



As you trace from left to right notice the 5-number summary (minimum, first quartile, median, third quartile, and the maximum). Any outliers will show up as squares or dots to the left or right of the boxplot.

7) For the data in the table shown below, create a stemplot (using tens as stems and ones as leaves). Make sure to label and include a key.

A marketing consultant observed 50 consecutive shoppers at a supermarket. One variable of interest was how much each shopper spent in the store. Here is the data (rounded to the nearest dollar), arranged in increasing order.

3	9	9	11	13	14	15	16	17	17
18	18	19	20	20	20	21	22	23	24
25	25	26	26	28	28	28	28	32	35
36	39	39	41	43	44	45	45	47	49
50	53	55	59	61	70	83	86	86	93

Stems	Leaves

Key (with context and units):

8) Try using your calculator to make a histogram of the same set of data as above. Copy your histogram below and make sure to give it context with a title and axes labels.

9) Use your calculator to make a boxplot with outliers. Don't forget to add a number line and context to your display.

10) A survey asked people "How long does it take you to eat breakfast?" The data below shows the responses in minutes. Make a dotplot for the data including a number line. Don't forget to add context.

Minutes	0	1	2	3	4	5	6	7	8	9	10	11	12
Number of People	6	2	3	5	2	5	0	0	2	3	7	4	1

III. CALCULATIONS FOR QUANTITATIVE DATA

There are two categories of numbers that are used to describe a set of data: measures of center and measures of spread.

A. Measures of Center:
1. The **mean** is a calculated "central" value of a set of numbers. It is the sum of all the data values divided by the number (n) of values.

Ex. {4, 36,10,22, 9} $mean = \bar{x} = \frac{\sum x_i}{n} = \frac{4+36+10+22+9}{5} = 16.2$

2. The **median** is the value that separates the bottom 50% of data from the top 50% of data. It is the middle element of an ordered set of data that is odd in number. It is the mean of the two middle elements of an ordered set of data that is even in number.

Ex. {4, 36, 10, 22, 9} = {4, 9, 10 22, 36} median = 10 because 10 in the middle value

{4, 36,10,22, 9, 43} = {4, 9, 10 | 22, 36, 43} median =16 because $\frac{10+22}{2} = 16$

B. Measures of Spread:
1. The **range** is a measure of the spread of the entire data. It is calculated by subtracting the minimum value from the maximum value. The minimum and maximum are considered to be measures of position.

Ex. {4, 36,10,22, 9} = {4, 9, 10, 22, 36} range = 36 – 4 =32

2. The **interquartile range (IQR)** is a measure of the spread of the middle 50% of the data. It is calculated by subtracting the 25th percentile (Q1) from the 75th percentile (Q3). Q1 is the median of the lower half of the data. It separates the bottom 25% of values from the top 75% of values. Q3 is the median of the upper half of the data. It separates the top 75% of values from the bottom 25% of values. In neither of these cases is the median considered in the top half or the bottom half of the data. (We omit the median when finding the quartiles.) The quartiles are considered to be measures of position.

Ex. {4, 36, 10, 22, 9} = {4, 9, 10, 22, 36} IQR = 29 – 6.5 = 22.5

Ex. {4, 36, 10, 22, 9, 43} = {4, 9, 10, 22, 36, 43} IQR = 36 – 9 = 27

3. The **standard deviation** is the measure of the spread around the mean. It is calculated using the following formula (for a sample):

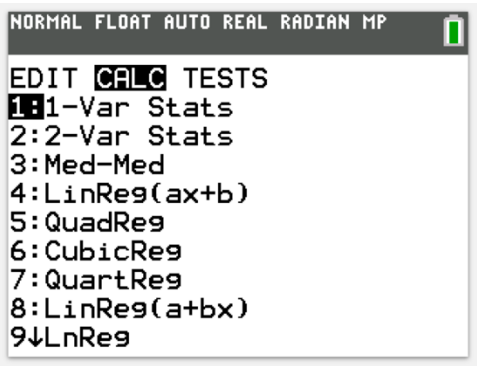
$$s = \sqrt{\frac{\sum(x_i - \underline{x})^2}{n-1}}$$

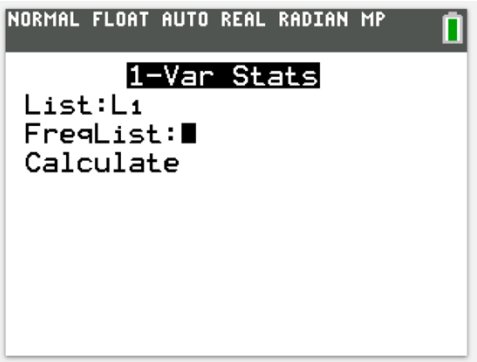
Ex. {4, 36, 10, 22, 9, 43}

$$s = \sqrt{\frac{\sum(x_i - \underline{x})^2}{n-1}} = \sqrt{\frac{(4-16.2)^2+(36-16.2)^2+\cdots+(43-16.2)^2}{5-1}} = 12.89$$

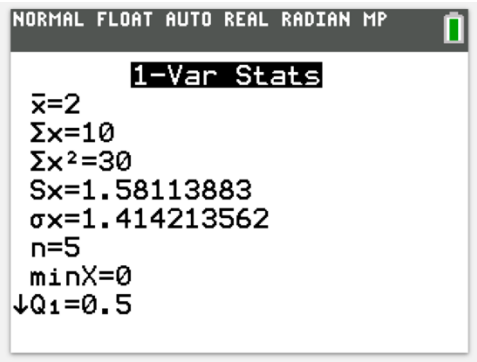
This means that the average number differs from the mean by about 12.89 units. The smaller the standard deviation the closer the data should be clustered around the mean. We will learn more about this measure of spread in the fall.

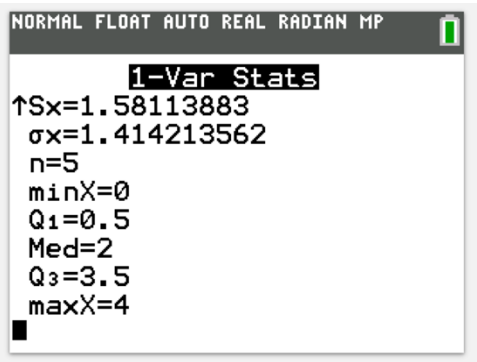
On the TI-84 calculator,
Go to Stat, Edit, and enter data into List 1 (or any open list).
Go to Stat, Calc, One-Var Stat, Calculate. The List should match where the data is. Usually, the FreqList is blank unless you have quantities in a second list.





You will find the mean ($\bar{x}$), population standard deviation (σ_x), sample standard deviation (S_x), The sample size (n), the minimum, first quartile (Q_1), median, third quartile (Q_3), and the maximum for the data set.





1) Calculate the statistics in the table below on the number of home runs Mark McGuire hit in each season from 1982-2001.

70	52	22	49	3	32	58	39
39	65	42	29	9	32	9	33

	Mark McGuire's Home Runs
Mean	
Minimum	
Maximum	
Median	
1st Quartile (Q_1)	
3rd Quartile (Q_3)	
Range	
Interquartile Range (IQR)	

2) Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

154	109	137	115	152	140	154	178	101
103	126	126	137	165	165	129	200	148

And for 20 first-year college men:

108	140	114	91	180	115	126	92	169	146
109	132	75	88	113	151	70	115	187	104

Calculate the statistics shown in the table below <u>for each</u> set of data.

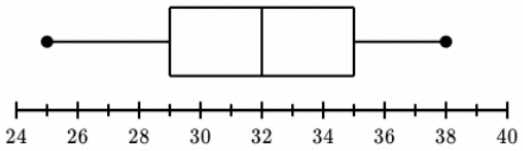
	Women's Scores	Men's Scores
Mean		
Minimum		
Maximum		
Median		
1st Quartile (Q_1)		
3rd Quartile (Q_3)		
Range		
Interquartile Range (IQR)		

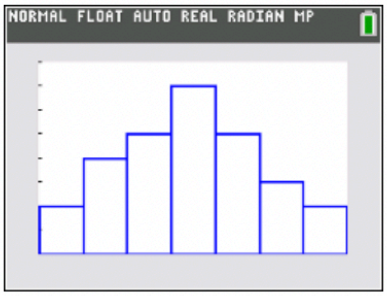
IV. ASSESSING THE SHAPE OF A GRAPH

When describing a set of data we look at the following features:
(1) Shape (2) center (3) spread (4) clusters, gaps, and outliers

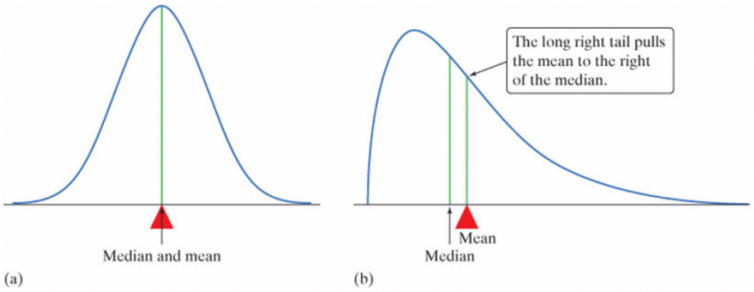
We have several terms that we use to describe the shape but this packet will concentrate on only two: symmetric and skewed.

You can tell if a graph is symmetric if a vertical line in the center divides the graph into two fairly congruent shapes. The following sets of data can be described as symmetric:

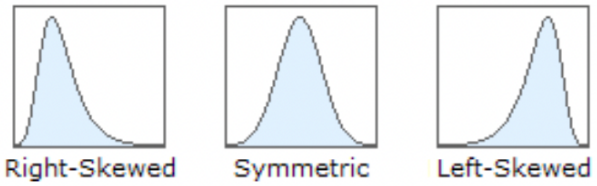




The mean and the median are approximately the same in a symmetric graph. Figure (a) below. You can tell if a graph is skewed if the graph has a big clump of data on either the left or the right with a tendency to get flatter on the opposite side.



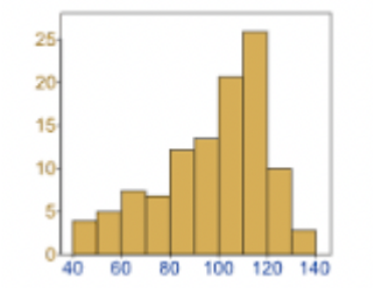
The distribution in figure (b) is skewed right. The mean gets pulled toward the tail, which is the direction of the skewness. So the mean > median.



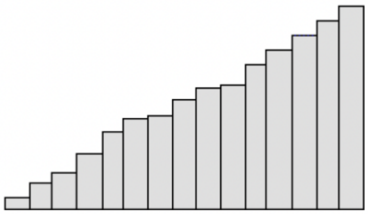
The mean is always pulled closer toward the tail.

Determine if the distribution is symmetric, skewed left or skewed right. Then determine if the mean or the median is the larger value. If they are equal, state so.

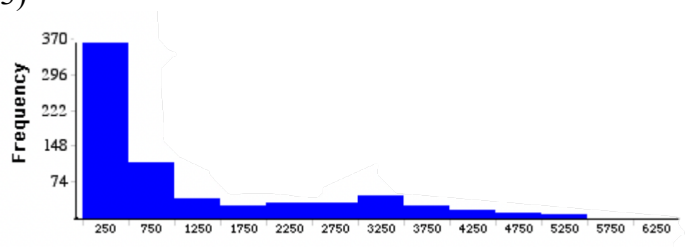
1)



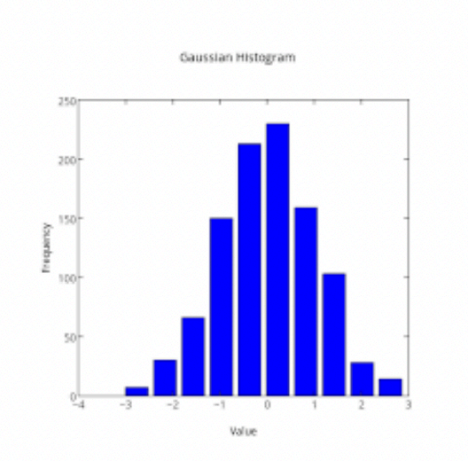
2)



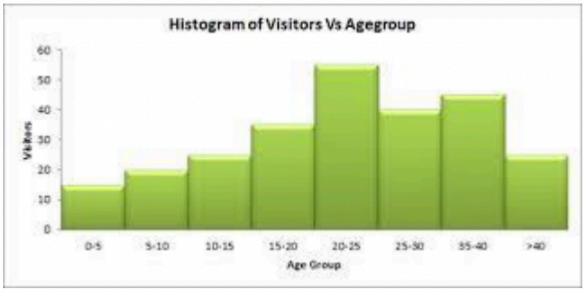
3)



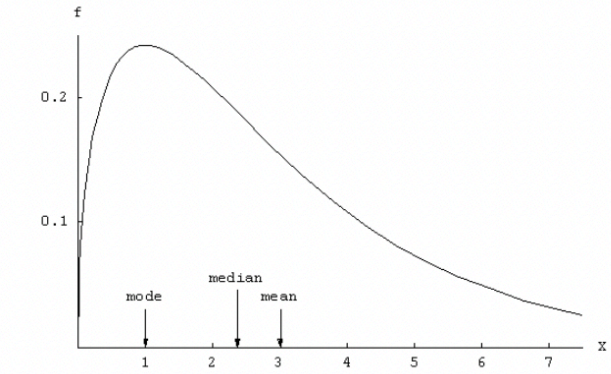
4)



5)


Histogram of Visitors Vs Agegroup
Visitors
60
50
40
30
20
10
0
0-5
5-10
10-15
15-20
20-25
25-30
35-40
>40
Age Group

6)


f
0.2
0.1
mode
median
mean
1
2
3
4
5
6
7
X

7)

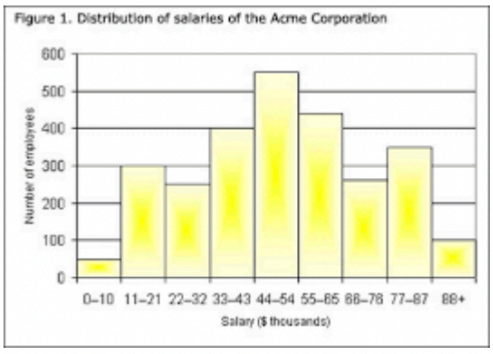
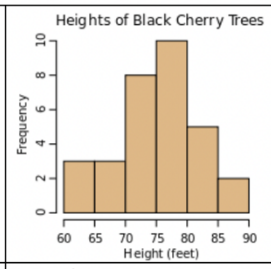


Figure 1. Distribution of salaries of the Acme Corporation
Number of employees
600
500
400
300
200
100
0
0–10
11–21
22–32
33–43
44–54
55–65
66–78
77–87
88+
Salary ($ thousands)

8. For the following graphs, find the shape, center (use the median), and spread (use the range). If there are any other notable features evident in the graph (clusters, gaps or outliers), then say where they are. Otherwise, do not comment on them. Note: To find the center of these graphs, use the frequencies found on the y-axis. Count how many are in each bar. If the bar has an interval, use the midpoint. Add these up and divide by 2. This tells you where the median is located. Which bar is this value located in?

1



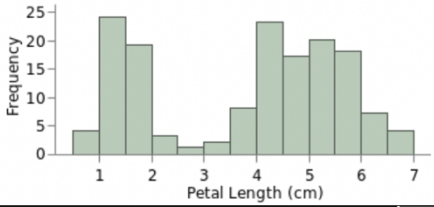
Shape _____

Center _____

Spread _____

Clusters? _____

2



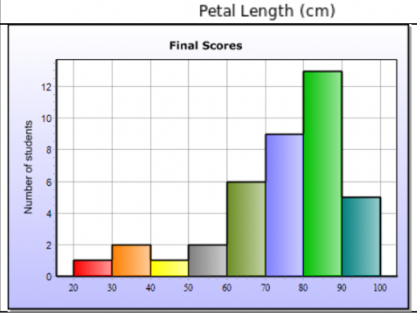
Shape _____

Center _____

Spread _____

Clusters? _____

3



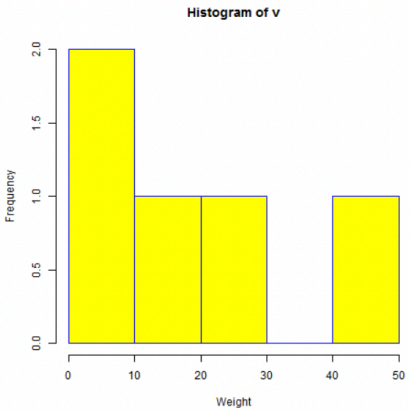
Shape _____

Center _____

Spread _____

Clusters? _____

4



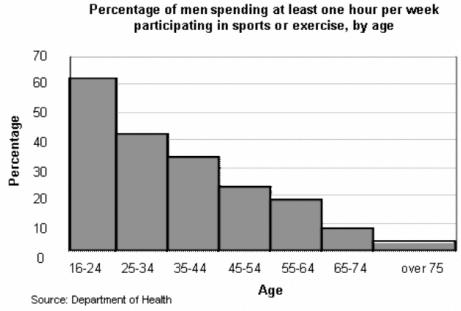
Shape _____

Center _____

Spread _____

Clusters? _____

5



Shape _____

Center _____

Spread _____

Clusters? _____

V. COUNTING PRINCIPLE AND PROBABILITY

A. Counting Principle

The **Basic Counting Principle** says that when there are $\boldsymbol{m}$ ways to do one thing and $\boldsymbol{n}$ ways to do another, then there are $\boldsymbol{m} \times \boldsymbol{n}$ ways of doing both.

1) Andy has asked his girlfriend to make all the decisions for their date on her birthday. She will pick a restaurant and an activity for the date. Andy will choose a gift for her. The local restaurants include Mexican, Chinese, Seafood, and Italian. The activities she can choose from are Putt-Putt, bowling, and movies. Andy will buy her either candy or flowers. How many outcomes are there for these three decisions?

2) A travel agent plans trips for tourists from Chicago to Miami. He gives them three ways to get from town to town: airplane, bus, train. Once the tourists arrive, there are two ways to get to the hotel: hotel van or taxi. How many different ways can the travel agent get tourists from Chicago to a hotel in Miami?

3) Suppose you can choose from 3 different shirts, 2 pairs of pants, and 5 pairs of shoes. Find out how many different outfit combinations you can make.

4) A sandwich comes with a choice of soda or tea and a side of fries, chips, slaw, or salad. Find the number of different meal combinations.

B. Permutations and Combinations

A **permutation** is an arrangement of all or part of a set of objects, *with* regard to the order of the arrangement or labels assigned to the objects.

Ex. Suppose we have a set of three letters: A, B, and C. We might ask how many ways we can arrange 2 letters from that set. Each possible arrangement would be an example of a permutation. The complete list of possible permutations would be: AB, AC, BA, BC, CA, and CB.

The formula $P(n,r) = \frac{n!}{(n-r)!}$ calculates the number of permutations of n things are taken r at a time.

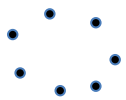
A **combination** focuses on the selection of objects *without* regard to the order in which they are selected. Thus, the arrangements of AB and BA represent two different permutations, because the order is different. However, they represent only 1 combination, because order is not important in a combination. Duplicates must be eliminated from the list of arrangements.

The formula $C(n,r) = \frac{n!}{(n-r)!r!}$ calculates the number of combinations n things are taken r at a time.

Decide whether each describes a permutation or a combination. Then find the different number of arrangements for each.

1) Kamisha wants to line up her 4 stuffed polar bears on a shelf over her bed. How many different arrangements are possible?

2) The dots on the left represent seven towns that are to be joined by telephone cables. If each town is connected directly to each of the other towns, how many cables will be required?



3) Each locker at the school has a dial with 50 numbers on it. A lock is opened by dialing 3 different numbers in a certain order. How many different sequences are possible?

4) Twenty people report for jury duty. How many different 12-person juries can be chosen?

5) How many different 5-card hands can be dealt from a standard deck of 52 cards?

6) In how many different ways can a president, vice-president, secretary, and treasurer be elected from a club with 32 members?

7) Regular license plates in California begin with a digit from 1-9, followed by 3 letters, followed by three digits, each from 0-9. How many different license plates are possible?

8) How many different committees of three members can be formed from a group of eight people?

9) Pizza Makers Co. offers ten different toppings for its pizzas. Suppose you order a pizza with three different toppings. How many different ways can a 3-topping pizza be made?

10) Ten horses are competing in a race. How many different 1^{st}-2^{nd}-3^{rd} place finishes are possible?

C. Probability

Probability is the likelihood or chance of an event occurring.

Probability = $\frac{\text{the number of ways of achieving success}}{\text{the total number of possible outcomes}}$

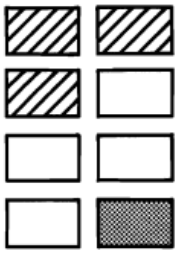
For example, the probability of flipping a coin and it being heads is ½, because there is 1 way of getting a head and the total number of possible outcomes is 2 (a head or tail). We write P(heads) = ½.

- The probability of something which is certain to happen is 1.
- The probability of something which is impossible to happen is 0.
- $0 \leq P(E) \leq 1$
- The probability of the complement of an event is 1 minus the probability of the event, $P(E^c) = 1 - P(E)$.

Multiple Events—Independent and Dependent Events

- Suppose now we consider the probability of 2 events happening. For example, we might throw 2 dice and consider the probability that both are 6's.
- We call two events **independent** if the outcome of one of the events doesn't affect the outcome of another. For example, if we throw two dice, the probability of getting a 6 on the second die is the same, no matter what we get with the first one- it's still 1/6. The P(rolling a 6 and rolling a 6)=(1/6)(1/6)=1/36
- On the other hand, suppose we have a bag containing 2 red and 2 blue balls. If we pick 2 balls out of the bag, the probability that the second is blue depends upon what the color of the first ball picked was. If the first ball was blue, there will be 1 blue and 2 red balls in the bag when we pick the second ball. So, the probability of getting a blue is 1/3. However, if the first ball was red, there will be 1 red and 2 blue balls left so the probability the second ball is blue is 2/3. When the probability of one event depends on another, the events are **dependent**. The P(blue and blue)=(2/4)(1/3)=2/12=1/6.

1) Using the cards on the right, find each probability if you choose one card at random.
a) P(striped)
b) P(shaded)
c) P(striped or white)
d) P(not white)
e) P(white or shaded)
f) P(not striped)
g) P(striped or white or shaded)



2) What is the probability of guessing the correct answer to a multiple-choice question if there are 5 choices?

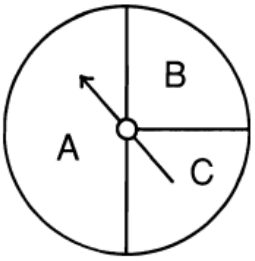
3) What is the probability that your birthday will fall on Saturday or Sunday?

4) A class of 25 students has 15 girls and 10 boys. If one student is chosen at random, what is the probability it is a girl?

5) What is the probability of guessing the correct answer to a true-false question?

6) What is the probability of winning a raffle if 500 tickets are sold and you buy 5 of them?

7) There are 26 letters in the alphabet. What is the probability that a letter chosen at random is in the word MATHEMATICS?

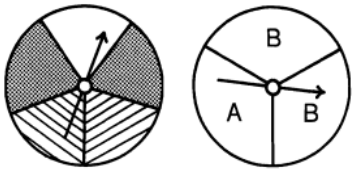


8) If you spin the spinner once,
a) If you spin the spinner once, what is the probability that it will stop on A?
b) If you spin the spinner once, what is the probability that it will stop on B?
c) If you spin the spinner 50 times, about how many times would you expect it to stop on A?
d) If you spin the spinner 80 times, about how many times would you expect it to stop on C?

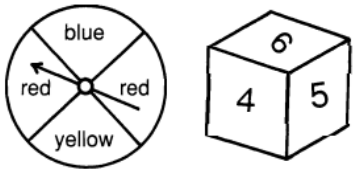
9) If you flip a coin 150 times, about how many times would you expect to get heads?

10) A magician asks you to pick a card from a standard deck of 52. What is the probability you choose an ace or a 10?

11) Find each probability if you spin both spinners.
a) P(white, A)
b) P(white, B)
c) P(striped, A)
d) P(not striped, A)
e) P(not white, B)

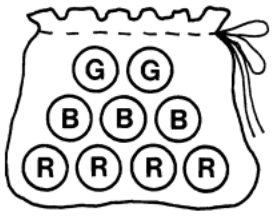


12)Find each probability if you spin the spinner and roll the number cube.
a) P(blue, 2)
b) P(blue, not 3)
c) P(yellow, even)
d) P(not blue, odd)



13) Each letter of the word BANANA is written on a card. Find each probability if you pick two cards without replacing the first.
a) P(B, then N)
b) P(B, then A)
c) P(A, then 6)
d) P(N, then N)
e) P(A, then A)
f)P(B, then B)

14) Find each probability if you pick a marble, do not replace it, then pick a second marble.
a) P(blue, then green)
b)P(green, then green)
c) P(green, then not green)
d) P(red, then not blue)



15) There were 6 purple socks and 4 orange socks in a drawer. Zucky picked one sock without looking and then another without looking (or replacing the first). What is the probability that he picked 2 purple socks?

16) There are 8 batteries in a box. 3 are good and 5 are bad. If 2 batteries are selected from the box, what is the probability that there is one bad battery selected followed by one good battery?

17) Two hundred seniors were surveyed about their genders and eye color. The results are recorded below in a 2-way table.

		Eye color			
		Brown	Blue	Green	Total
Gender	Male	50	40	20	110
	Female	40	40	10	90
	Total	90	80	30	200

a. Suppose that one of these seniors is randomly selected. What is the probability that the selected student is a male?

b. Suppose that one of these seniors is randomly selected. What is the probability that the selected student has blue eyes?

c. Suppose that one of these seniors is randomly selected. What is the probability that the selected student is a male and has blue eyes?

d. Suppose that one of these seniors is randomly selected. What is the probability that the selected student is a male or has blue eyes?

e. Suppose that one of these seniors is randomly selected. What is the probability that the selected student has blue eyes, given that the student is male?

f. Suppose that one of these seniors is randomly selected. What is the probability that the selected student is a male, given that the student has blue eyes?

VI. CONTEXT/RECENT STUDY

Find a recent article on a study in the news. Pick a topic that is of interest to you. This could be a career field you are going into or something you find weird/puzzling. In a paragraph or two, summarize the study including any findings. Make sure to include:

- *Who* is being studied?
- *What* about those individuals is being recorded/analyzed (i.e., what are the variables?) Do you think the variables are categorical or quantitative in nature?
- *When* was the data collected?
- *Where* was the data collected (geographical)?
- *Why* do you think this data was collected and analyzed?
- *How* was the data collected and analyzed? In other words, what methods were used?
- Give a conclusion for the study.
- Write down three questions you have about the study.

Make sure your summary does not include any personal opinions. (Bring a copy of your W's and a copy of the article with you.). You will be presenting your W's in the first couple of weeks of class.