AP STATISTICS SUMMER WORK

Dear AP Statistics Student,

Welcome to the beginning of your study of statistics. I look forward to working with you over the course of the next year and preparing you for the AP Statistics test in May. Enjoy your summer break!

We will hit the ground running on day 1 and I want to give you the opportunity to get a head start. This is your first assignment for the class which contains basic course content that you will need to understand. These foundational skills are essential in the study of statistics. Some of the material is review and other parts contain new concepts and terminology. You will need to know all of this information extremely well. This packet should take you roughly 1 hour to complete. There will be a quiz on the material in the first week of school. Please complete this packet on a separate sheet of paper. Label each part and the question number. This will be turned in the first week of school.

A few reminders

- Statistics is a very different "math" course. It does not follow the sequence of Alg1, Alg2, Pre-Calc, etc. This class is more about thinking and context than solving for "x". It is said that doing calculus is hard, but you know what you need to do. Doing Statistics is easy, but you first need to figure out what to do. You also need to be able to explain and justify your process, so come ready for a lot of exploration, hands on activities, and discussion/writing, while doing a lot of high quality work.

- **Calculator**: As with other math courses a <u>graphing calculator is required</u>. If you do not currently have one I would suggest that you purchase a TI-Nspire graphing calculator or you will need to rent one from the library within the first week of school

- **AP grading**: This is an AP course where students have the opportunity to <u>earn</u> college credit. Grades will reflect your ability to meet the course expectations which include content mastery and deadlines. Students pass the AP exam with a score of a 3 which will earn credit at most universities. With that in mind a 3 is the equivalent of a C, likewise a 4 is equivalent to a B and a 5 equivalent to an A. The grading in this class will follow the same structure. The AP exam is "curved" and the assessments will match that to prepare for the exam. Students will earn As for work that would earn a 5 on the AP exam, Bs for work that will earn 4s and so on. If you are taking this course solely for the grade and not learning the material, this is not the class for you.

- **Workload:** As an AP course, students are expected to be prepared for class and complete their some assignments outside of class. Students should expect to spend roughly 3-5 hours per week on class material and studying outside of class time in preparation to pass the AP exam.

- **Recommendation:** There is a new book out on learning and studying that is fantastic. I will try to teach you some concepts throughout the year, but I recommend you look at "Outsmart You Brain," by Daniel T Willingham. It is an easy read with a lot of great tips that will help you with all classes and learning in general. He also has a website with links to articles as well.

Mrs. Snell

Laura_snell@charleston.k12.sc.us

AP Statistics Name ____________________________________

HW #1: APS Primer

Part 1: Rounding

How many different ways are there to round? Is any single method more correct than another? In some courses, students are expected to round to a certain number of decimal places, some use significant figures, some round to zero, some even round up.

There are no hard and fast rules for rounding in AP Statistics!
The general rule of thumb is *round appropriately for the given situation.*

Rounding is an art that requires good intuition and number sense. If you don't round enough, your numbers will be hard to interpret. If you round too much, you are sacrificing the accuracy of your information. Your job as a statistician is to round in such a way that creates a good balance between the two. It is best practice to follow the standard rounding rules; that is half-round up.

Just remember: 5 and above round up, 4 and below round down!

EXAMPLES:

1. Round 29.4319 to the nearest hundredth
*"Hundredths" is the second decimal place, which gives 29.43. The next number is 1, so we round down; the final answer stays **29.43***

2. Round 4.39986 to the nearest ten thousandth
*"Thousandths" is the third decimal place, which gives 4.399. However, since next number is 8, we round up; the final answer becomes **4.400** (the next number up from 399 is 400)*

3. Round .037292319586
*This one is a judgment call. In AP Stat, we usually round to 2, 3, or 4 decimal places, depending on the situation. You should round at least to the nearest hundredth (**.04**), but you can do to the nearest thousandth (**.037**), or nearest ten-thousandth (**.0373**). All of these are technically correct! If you went further than this (such as .037293), you aren't wrong, but too many digits makes the number hard to contextualize and interpret – I would advise against it.*

Practice Problems – Check the answers to the odd-numbered ones in the back of the packet!

Round the following to the given decimal point.

1. Round 12.842 to the nearest tenth	2. Round 0.4892745 to the nearest hundredth	3. Round 0.0342119 to the nearest thousandth	4. Round 0.06049822 to the nearest ten thousandth

Round the following. There is no rounding rule in AP Stats. Use your best judgment.

5. Round 25.6895234	6. Round 0.033231532	7. Round 0.00279625	8. Round 0.63636363…

Part 2: Fractions, Decimals, and Percentages

In Statistics, we often deal with ***proportions*** – an amount out of a total. Proportions can be expressed as fractions, decimals, or percentages (which are just proportions out of 100).

Converting between forms of proportions

Fraction → Decimal: This is easy: just *divide* your numerator and denominator! Be aware that you may have to *round* the answer

Examples:

1. Convert $\frac{11}{14}$ to a decimal

   *11÷14 = 0.7857142857…you must round this! It rounds nicely at 4 decimal places, or **0.7857.***
   *You could also do **0.79** or **0.786** if you would like*

Percentage → Decimal: In order to use a percentage in an equation or a calculation, you MUST convert it into a decimal! Since percentages are always out of 100, just divide the percentage by 100!
**A shortcut to this is moving the decimal two places to the left!*

Examples:

1. Convert 35% to a decimal

   *35 ÷ 100 = **0.35.** Likewise, moving the decimal point 2 places to the left means **35.0 →.350,** or .35*

2. Convert 6.85% to a decimal

   *6.85 ÷ 100 = **0.0685.** Notice a ZERO between the decimal point and the 6 (you need to insert a zero in order to move 2 decimal places to the left)*

Decimal → Percentage: To turn a decimal into a percentage, simply do the opposite of the percentage-to-decimal procedure. You can *multiply* the decimal by 100, or move the decimal two places to the *right*.

Examples:

1. Convert 0.02 to a percentage

   *0.02 • 100 = **2%.** Likewise, moving the decimal point 2 places to the right means **0.02 →2.0**, or 2*

2. Convert $\frac{5}{12}$ to a percentage

   *5÷12 = 0.4166666…let's round to 0.4167. Then, 0.4167 • 100 (or going 2 decimal places to the right) = **41.67%.***

Using Percentages

*To find a percentage of a number (such as 25% of 84), **multiply** the percentage (*as a decimal*) by the number

Examples:

1. Find 25% of 72

   *25% is 0.25, and 0.25•72 = **18***

2. Find 3.9% of 749

   *3.9% is 0.039, and 0.039•749 = **29.211***
   NOTE: If you need a whole number, round to **29**

Practice Problems – Check the answers to the odd-numbered ones in the back of the packet!

1. Convert $\frac{13}{3}$ to a decimal	2. Convert $\frac{41}{563}$ to a decimal	3. Convert 70% to a decimal	4. Convert 8% to a decimal
5. Give 22.45% as a decimal	6. Give 100% as a decimal	7. Convert 0.672 to a percentage	8. Convert 0.0052 to a percentage
9. Convert $\frac{4}{25}$ to a percentage	10. Convert $\frac{11}{285}$ to a %	11. What is 17.2% of 89?	12. What is 3% of 446?

Part 3: Summary Statistics – Center and Spread

A *statistic* is a number that gives information about a set of data. Common examples include mean, median, mode (which we won't worry about in AP Stat), range, standard deviation, and more!

SYMBOLOGY

In statistics, we use a variety of *symbols* to represent statistics. Sometimes, the symbol used depends on whether we are talking about a **population** or a **sample** (select members of a given population)

	Mean	Standard Deviation	Median	Number of data points
POPULATION	**μ** ("mu")	**σ** ("sigma")	*No symbol* (Often abbreviated "**Med**.")	**N**
SAMPLE	$\bar{x}$ ("x-bar")	**s**		**n**

Measures of CENTER

The *center* of a data set lets us understand the "average" or "typical" value of a number in that data set. There are two main measures of center: **mean** and **median**.

MEAN

Add up all data points, then divide by the number of data points.

- **μ (or $\bar{x}$) =** $\frac{\sum x}{n}$
- "Sum of data points over number of data points"

Example 1: *Science grades of a* <u>*sample*</u> *of 15 juniors:*
91, 87, 66, 74, 85, 98, 43, 88, 77, 62, 83, 91, 89, 52, 100

This is a <u>SAMPLE</u>, so $\bar{x} = \frac{\sum x}{n} = \frac{1186}{15}$ **= 79.07**

Example 2: *Heights of* <u>*all*</u> *6 people in a family (inches):*
47, 58, 61, 65, 68, 70

This is the <u>POPULATION,</u> so $\mu = \frac{\sum x}{n} = \frac{369}{6}$ **= 61.5**

MEDIAN

The *middle number* of the data set, assuming that the data points are **in order** (smallest to largest)

- If there are 2 numbers in the middle, find the *mean* of those two numbers!
- A nice **trick** for finding the <u>*position*</u> of the median is to use $\frac{n+1}{2}$

Example 1: *Science grades of a* <u>*sample*</u> *of 15 juniors:*
91, 87, 66, 74, 85, 98, 43, 88, 77, 62, 83, 91, 89, 52, 100

$\frac{n+1}{2} = \frac{15+1}{2}$ **= 8.** Median is the **8th** number (IN ORDER)

43, 52, 62, 66, 74, 77, 83, **85,** 87, 88, 89, 91, 91, 98, 100

Example 2: *Heights of* <u>*all*</u> *6 people in a family (inches):*
47, 58, 61, 65, 68, 70

$\frac{n+1}{2} = \frac{6+1}{2}$ **= 3.5.** Median is **between the 3rd & 4th** number

47, 58, 61, 65, 68, 70; Average = $\frac{61+65}{2}$ **= 63**

Measures of SPREAD

The *spread* of a data set tells us whether the data points are far apart or clustered together. The most important measure of spread is **standard deviation,** which is the <u>**typical distance of the data points from the mean**</u>. Other measures of spread, such as Range and IQR, will be discussed in Part 5.

The formulas for **Standard Deviation** are as follows. Note that they are slightly different for a population and a sample (the sample one will be slightly larger to account for the fact that the sample doesn't include all members of a population)

Population: $\sigma = \sqrt{\frac{\Sigma(x_i-\mu)^2}{n}}$ Sample: $s = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}}$

You will NOT have to calculate Standard Deviation by hand in this course!

Be able to *interpret* and *compare* the Standard Deviations of different data sets:

- ***Larger*** **Standard Deviation:** The data is more spread out (points are typically further from the mean)
- ***Smaller*** **Standard Deviation:** The data is closer together (points are typically closer to the mean)

<u>**Example:**</u>

Data Set 1: 1, 2, 3, 17, 18, 19; $\mu = 10$, $\sigma = 8.04$ **Data Set 2:** 7, 8, 9, 11, 12, 13; $\mu = 10$, $\sigma = 2.16$

Notice how Data Set 1 is more spread out, while Data Set 2 is closer together. This is reflected in the fact that Set 1's Standard Deviation (8.04) is higher than Set 2's Standard Deviation (2.16)

Practice Problems – Check the answers to the odd-numbered ones in the back of the packet!

1. Find the mean and median of the following data set. **Show work** when appropriate!

Teaching experience of <u>all</u> teachers at a certain school (n = 15): 1, 3, 3, 3, 4, 4, 5, 5, 5, 6, 7, 7, 18, 23, 26

Symbol for mean: ________ *Value of mean:* ________ *Position of Median:* ________ *Value of Median:* _______

2. Find the mean and median of the following data set.

Weights of 8 randomly-selected chickens on a farm (in pounds): 5.4, 5.7, 6.2, 6.9, 7.2, 7.2, 8.1, 9.0

Symbol for mean: ________ *Value of mean:* ________ *Position of Median:* ________ *Value of Median:* _______

3. Find the mean and median of the following data set.

Temperature readings on <u>all</u> thermostats in an office building: 71, 72, 72, 74, 68, 74, 71, 72, 69, 76

Symbol for mean: ________ *Value of mean:* ________ *Position of Median:* ________ *Value of Median:* _______

4. List the following data sets in order from *least spread out* to *most spread out*. Then, **write** 1-2 sentences explaining how you could tell.

Teachers: $\sigma = 7.02$ **Chickens:** $s = 1.21$ **Thermostats:** $\sigma = 2.26$

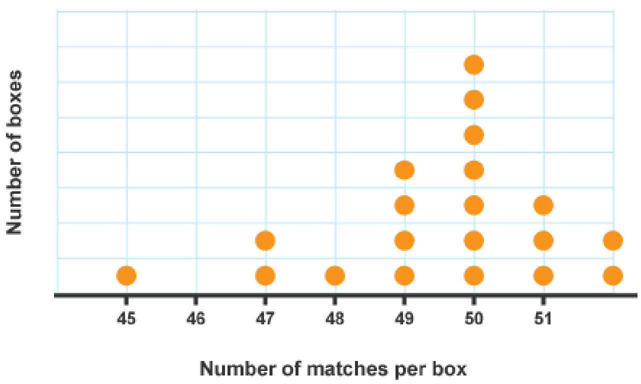
Part 4: Graphing Data: Dotplots and Stem-and-Leaf Plots

Statistics such as mean, median, and standard deviation are very useful in summarizing data and giving overall trends. But they don't tell the full story. By making a *graph* of the data, we can go *beyond* the numbers and see *shapes* and *patterns* in the data. Shown below are two common ways in which to graph data

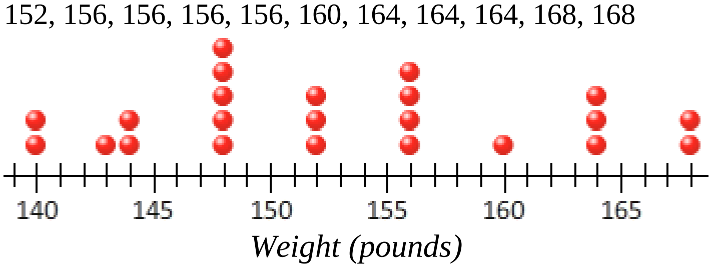
Dotplots

- Make an AXIS on the bottom (you can go by 1s, 2s, 5s, 10s…whatever makes sense for the data!)
- Put one dot for each data point on the axis. If there is more than one data point for a given value, *stack* the dots!

Example: *Number of matches in 20 randomly-selected boxes.*
45, 47, 47, 48, 49, 49, 49, 49, 50, 50, 50, 50, 50, 50, 50, 51, 51, 51, 52, 52



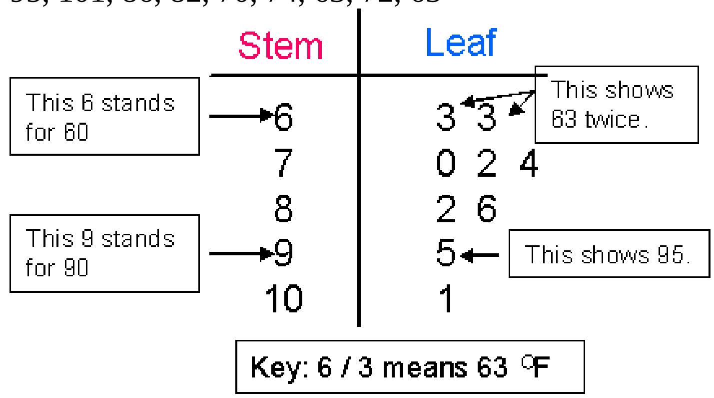
Example: *Weights of players on a high school baseball team*
140, 140, 143, 144, 144, 148, 148, 148, 148, 148, 152, 152, 152, 156, 156, 156, 156, 160, 164, 164, 164, 168, 168



Stem-and-Leaf Plots (also called Stemplots)

- Use a KEY to determine what the stems and leaves are worth
- DO NOT SKIP STEMS. If there are no data points for that stem, just keep the stem there and put no leaves after it. Skipping the stem will alter what the stemplot looks like.

Example: *Temperatures at OU football games, 2009*
95, 101, 86, 82, 70, 74, 63, 72, 63



Example: *Gross National Product (per capita) of West African countries*
180, 240, 260, 270, 310, 330, 360, 370, 390, 410, 480, 500, 710, 730, 890

Stem	Leaf
1	80
2	40 60 70
3	10 30 60 70 90
4	10 80
5	00
6	
7	10 30
8	90

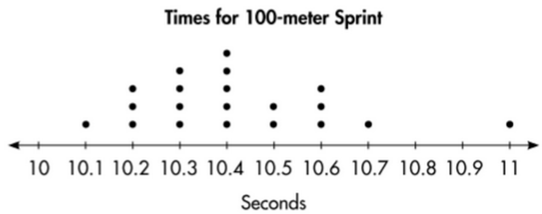
Notice how the 6 is still here even though there are no data points in this interval

Key: 4 | 80 means an income of $480.

Practice Problems – *Check the answers to the odd-numbered ones in the back of the packet!*

1. Construct a well-labeled (i.e., *label the units on the axis*!) **dot plot** for the fuel efficiency of a random sample of 2015 model year vehicles: 16, 23, 24, 27, 29, 30, 30, 32, 32, 31, 31, 31, 31, 31, 31, 31, 40.

2. Using the dot plot shown, find the **mean** and **median** 100-meter sprint time



3. The following stem plot shows the final exam scores of a class with 10 students. List the scores of each student:

stem	leaf
6	9
7	
8	7 8 8 9
9	0 6 7 7
10	0

Key: 6|8 means 68

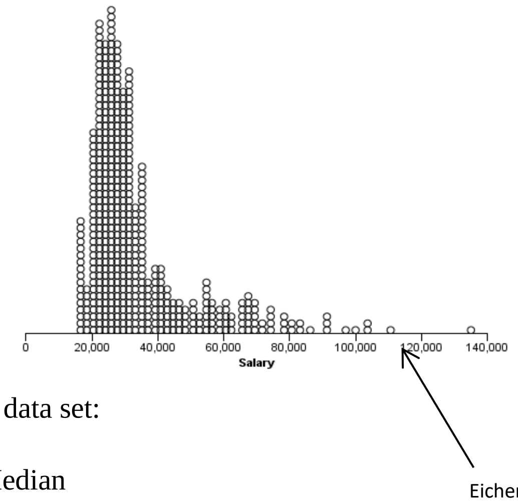
4. Find the **mean** and **median** of the data shown in the stem plot. ***NOTE:*** *Look at the key carefully!*

Stem	Leaf
2	0 2 3 6
3	2 3 5 6 7
4	6 8 9
5	4 7
6	2
7	3

KEY: 4|6 = 4.6

Part 5: Graphing Data: Box and Whisker Plots

We can see that dot plots and stem plots are effective ways of graphing and analyzing data sets. But what if a data set had hundreds, or even *thousands*, of data points? Take the dot plot shown at the right, for instance – analyzing each and every dot would take *forever!*

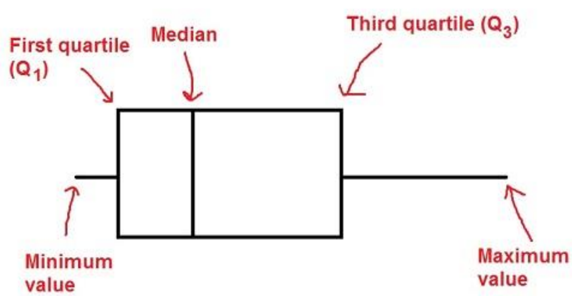


This is why it is important to be able to use *summary statistics* (such as mean, median, and standard deviation) to analyze our data. But in doing so, we lose our ability to *graph* the data and *see* what's going on.

Or do we?

This is the beauty of **Box and Whisker Plots** (or "Boxplots" for short).
A box and whisker plot is a graphical representation of just 5 numbers in a data set:

- **Minimum (Min):** Smallest value in the data set
- **Lower Quartile (Q1):** Midpoint between the Minimum and the Median
- **Median (Med):** Midpoint of the entire data set (see Part 3)
- **Upper Quartile (Q3):** Midpoint between the Median and the Maximum
- **Maximum (Max):** Largest value in the data set



NOTE: The Mean and Standard Deviation of a data set are **NOT** included in a boxplot. **DO NOT** try to include them or talk about them!

It is important to note that **no matter how far apart or close together these 5 numbers are, ¼ (25%) of the data is in each of the 4 sections of the boxplot** (lower whisker, lower half of the box, upper half of the box, and upper whisker)

Boxplots also allow us to measure **Spread** in 2 ways:

- **Range:** Difference between the extremes of the data (Maximum minus Minimum)
- **Interquartile Range (IQR):** Difference between the two *quartiles* (Q3 – Q1)

Example 1: Points scored by Russell Westbrook, 2016 NBA Playoffs

14, 14, 16, 19, 19, 24, 25, 26, 27, 28, 28, 29, 30, 31, 31, 35, 36, 36

**Finding MEDIAN:* (18 games. $\frac{18+1}{2}$ = 9.5, so average the 9th and 10th numbers)

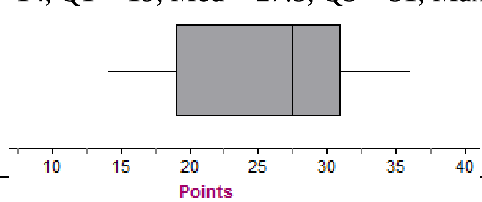
14, 14, 16, 19, 19, 24, 25, 26, **27, 28,** 28, 29, 30, 31, 31, 35, 36, 36

Median = 27.5

**Finding QUARTILES:* There are 18 data points, so *split* the data set in half! *Find the middle of <u>each half!</u>*

Q1 14, 14, 16, 19, **19,** 24, 25, 26, 27 *Q3* 28, 28, 29, 30, **31,** 31, 35, 36, 36

***BOXPLOT:** Min = 14, Q1 = 19, Med = 27.5, Q3 = 31, Max = 36



RANGE: 36 – 14 = **22**

IQR: 31 – 19 = **12**

Example 2: Ages of 9 employees in an office

37, 24, 51, 46, 62, 28, 35, 49, 55

Finding MEDIAN:* (9 people. $\frac{9+1}{2}$ = 5, it's the 5th number. **Remember that numbers must be <u>in order</u>!)

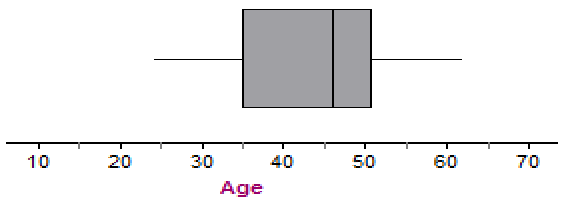
24, 28, 35, 37, **46,** 49 51, 55, 62

Median = 46

Finding QUARTILES:* There are 9 data points, so *split* the data set in half! ***NOTE:* *If there is one number that serves as the median, as with this data set,* ***it is not included in <u>either</u> half!***

Q1		*Q3*
24, **28, 35,** 37	46 (not included)	49, **51, 55,** 62
Average: 31.5		*Average: 53*

***BOXPLOT:** Min = 24, Q1 = 31.5, Med = 46, Q3 = 53, Max = 62



RANGE: 62 – 24 = **<u>38</u>**

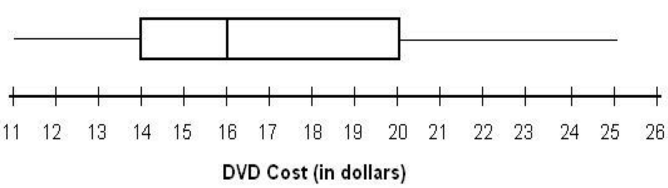
IQR: 53 – 31.5 = **<u>21.5</u>**

Practice Problems – *Check the answers to the odd-numbered ones in the back of the packet!*

1. Construct a boxplot for the following data set. **Be sure to include a labeled <u>axis</u>** like the examples above!

17, 21, 24, 26, 31, 33, 36, 37, 41, 48

2. Analyze the following boxplot:



Minimum: Range:

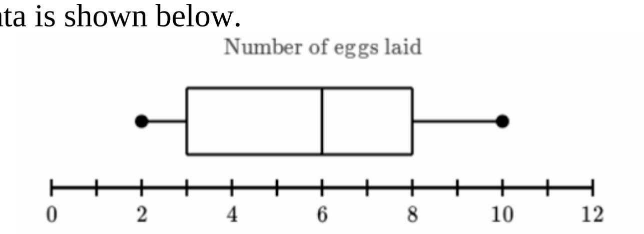
Q1: IQR:

Median: Mean:

Q3:

Maximum:

3. A farmer has 168 laying hens. He recorded how many eggs each hen laid in one week. A boxplot of the data is shown below.



HINT: *Remember that each of the 4 sections of the boxplot has ¼, or 25%, of the data points.*

Find the number of chickens that laid…

A. More than 8 eggs _____ B. Fewer than 6 eggs ______

C. Between 3 & 8 eggs ______ D. Between 2 & 8 ______

Part 6: Ways of Counting

Fundamental Counting Principle
A tree-diagram can be a very useful way to organize/visualize the variety and number of options in a given situation. Many times, however, we are just interested in the number of different arrangements rather than the arrangements themselves. This is where the Fundamental Counting Principle is handy.

If we have two or more groups of items to choose from, the number of ways they can be arranged is the product of the number in each group.

$$a_1 \cdot a_2 \cdot \ldots \cdot a_n \quad \text{where } n \text{ is the number of groups}$$

Example: Suppose we flip a fair coin, choose a card from a standard (52 card) deck, and roll a 6 sided die
Some possible arrangements:

H, Ace ♥, 1	H, Ace ♥, 2	H, Ace ♥, 3	H, Ace ♥, 4	H, Ace ♥, 5	H, Ace ♥, 6
T, Ace ♥, 1	T, Ace ♥, 2	T, Ace ♥, 3	T, Ace ♥, 4	T, Ace ♥, 5	T, Ace ♥, 6…

Listing every arrangement and counting them would be time-consuming, at best. Using the Fundamental Principle of Counting there are

$$2 * 52 * 6 = 624 \text{ different arrangements.}$$

Replacement
When selecting a certain number of items from ONE group, whether we replace each item after selection can drastically change the number of arrangements, and therefore the probability of a given event. In these cases, the order in which items are selected matters. That is… *1,2 is distinct from 2,1*.

Consider drawing four cards from a standard (52 card) deck:

Counting WITH Replacement

There are 52 options for the first card. We then put the card back and draw again from all 52 cards. In this case, <u>with replacement</u>, each card could be drawn more than once and the probability of drawing each card does not change.

$$52 * 52 * 52 * 52 = 52^4 = 7{,}311{,}616$$

There are 7,311,616 ways to select 4 cards with replacement (where the order in which they are drawn matters.)

$$\boldsymbol{n^r}$$
where n is a certain number of options taken r at a time

Example:
Given four cards (♠ ♣ ♥ ♦), how many ways are there to draw two with replacement?

$$4 * 4 = 4^2 = 16$$

♠ ♠	♠ ♣	♠ ♥	♠ ♦
♣ ♠	♣ ♣	♣ ♥	♣ ♦
♥ ♠	♥ ♣	♥ ♥	♥ ♦
♦ ♠	♦ ♣	♦ ♥	♦ ♦

Counting WITHOUT Replacement

There are 52 options for the first card. Since we do NOT put the card back there are now only 51 left to choose from. Then 50, then 49. The probability of selecting each card *changes* each time we draw a card.

$$52 * 51 * 50 * 49 = 6{,}497{,}400$$

There are 6,497,400 ways to select 4 cards without replacement (where the order in which they are drawn matters. This is called a ***permutation***.)

$$\boldsymbol{n * (n-1) * (n-2) * \ldots * (n-(r-1))}$$

or

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

where n is a certain number of options taken r at a time

Example:
Given four cards (♠ ♣ ♥ ♦), how many ways are there to draw two without replacement?

$$4 * 3 = 12$$

	♠ ♣	♠ ♥	♠ ♦
♣ ♠		♣ ♥	♣ ♦
♥ ♠	♥ ♣		♥ ♦
♦ ♠	♦ ♣	♦ ♥	

Combinations

When the order does NOT matter (1,2 is no different from 2,1 and is only counted once), we need to account for the duplicate arrangements and remove them from our counting.

Using the same card example, we can choose any 2 cards out of the 4. Since ♥ ♦ is the same as ♦ ♥ and duplicates aren't counted, that leaves us with:

	♠ ♣	♠ ♥	♠ ♦
~~♣ ♠~~		♣ ♥	♣ ♦
♥ ♠	♥ ♣		♥ ♦
♦ ♠	♦ ♣	♦ ♥	

there are now only **<u>6</u>** ***combinations*** of 4 cards taken 2 at a time.

We can modify the formula for permutations to account for the extra results.

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

Example: How many ways are there to choose 2 cards from 4?

$$ {}_4C_2 = \binom{4}{2} = \frac{4!}{2!\,(4-2)!} = \frac{4!}{2! * 2!} = \frac{4 * 3 * \cancel{2 * 1}}{2 * 1 * \cancel{2 * 1}} = \frac{12}{2} = 6 $$

*note your calculator will do this as well. Go to MENU> 5:Probability>3:Combinations

Practice Problems – *Check the answers to the odd-numbered ones in the back of the packet!*

1. The school lockers have 3-digit combinations using the numbers 0 through 49. How many different combinations are there?

2. How many 3-digit even numbers are there?

3. There are 8 runners in a race. How many ways could the runners come in first and second place?

4. $25 prizes are awarded to the two fastest runners in the race. How many different ways can the prizes be awarded if there are 8 runners?

Calculate the following.

5. $\binom{5}{5}$

6. ${}_5C_0$

7. $\binom{10}{3}$

8. ${}_{10}C_7$

9. $\binom{1}{2}$

Part 7: Probability: Theoretic and Experimental Probability

In Statistics, one of the primary questions we seek to answer is *How likely is X to occur?* This **probability** is the basis for much of our interpretation of the studies we encounter.

For equally likely outcomes:

$$P(event) = \frac{\#\ of\ favorable\ outcomes}{\#\ of\ outcomes\ in\ the\ sample\ space}$$

The ***sample space*** **S** of a random experiment is the set of all possible outcomes.

There are two basic types of probabilities with which we will interact. A ***theoretical probability*** describes the expected likelihood of a certain event, whereas an ***experimental probability*** describes the likelihood of a certain event based on observation.

Theoretical Probability

This tells us what we would expect to have happen in the long run if we did many trials.

Example:
Consider a fair six-sided die; what is the probability of rolling a 2 or a 4?

There are six possible outcomes for one roll of the die S={1, 2, 3, 4, 5, 6}. Since the die is fair, each number is equally likely to be rolled. The probability of rolling each number is $\frac{1}{6}$

The rolls we are interested in are {2, 4}. Since either a 2 or a 4 would work, this means we have two favorable outcomes out of 6.

$$P(2\ or\ 4) = \frac{2}{6} = \frac{1}{3} \approx 0.33$$

If we rolled a fair six-sided die many times, approximately $\frac{1}{3}$ of the rolls should result in either a 2 or a 4.

Experimental Probability

Here we are estimating the probability of a certain event based on trials.

Example:
A fair six-sided die was rolled 20 times, the results are shown in the table.

4	1	3	3	5
5	3	5	2	2
6	2	6	2	4
4	5	4	6	4

$$P(2\ or\ 4) = \frac{9}{20} \approx 0.45$$

In 20 rolls of a fair six-sided die, 0.45 of the rolls were either a 2 or a 4.

Notice that the experimental probability is not the same as the theoretical probability. *Why is this?* The experimental probability *couldn't* have been 0.33. One third of 20 is 6.67. Since we cannot have 0.67 of a roll, there was no way we could have gotten the same result! This doesn't mean the die isn't actually fair.

In the long run, after many trials, the experimental probability should approach the theoretical probability. We often use an experimental probability to *predict* theoretical probability.

Practice Problems – *Check the answers to the odd-numbered ones in the back of the packet!*

A quarter, a nickel, and a penny are all flipped.

1. What is the sample space?

Calculate the following:

2. P(quarter is heads)	3. P(penny H and nickel H)	4. P(one coin is heads)	5. P(at least one heads)

Thirty trials of the above coin flips were run, the results are shown in the table.

Trial	Quarter	Nickel	Penny
1	T	H	T
2	H	H	H
3	T	H	T
4	H	H	H
5	T	T	T
6	H	H	T
7	T	T	T
8	H	H	H
9	H	T	T
10	H	T	T
11	T	T	H
12	H	T	H
13	T	H	T
14	T	H	T
15	T	H	T
16	H	H	H
17	T	H	H
18	H	H	H
19	T	H	H
20	H	T	H
21	T	T	H
22	T	T	T
23	T	T	H
24	H	T	H
25	H	H	H
26	T	T	T
27	H	T	H
28	H	T	H
29	T	T	T
30	H	T	H

6. P(quarter is heads)	7. P(penny H and nickel H)
8. P(one coin is heads)	9. P(at least one heads)

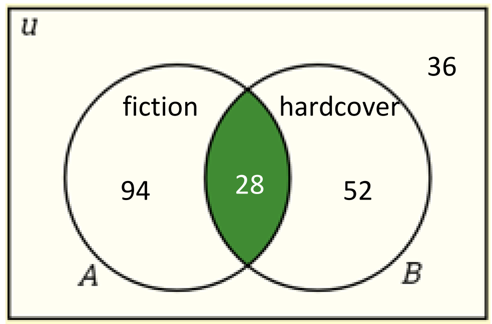
10. Should we be concerned that the theoretical and experimental probabilities differ? Explain.

Part 8: Probability: Two-way Tables and Venn Diagrams

Organizing the outcomes of random events in Venn diagrams and two-way tables can help us visualize the sample space and find the probabilities of new events based on given events.

$P(A \cap B)$

The ***intersection*** of events A and B. (A *and* B)



	Fiction	Non-fiction	
Hardcover	28	52	80
Paperback	94	36	130
	122	88	210

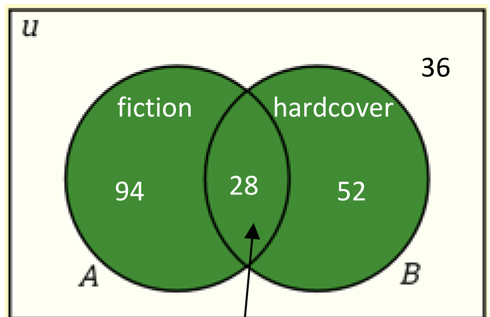
$$P(\text{fiction } \cap \text{hardcover}) = \frac{28}{210}$$

$P(A \cup B)$

The ***union*** of events A and B. (A *or* B)

28 books are counted in both totals! Make sure to take it out of one of set so they aren't counted twice.

$$P(A) + P(B) - P(A \cap B) = \frac{80 + 122 - 28}{210}$$



The overlap is counted in BOTH set A and set B.

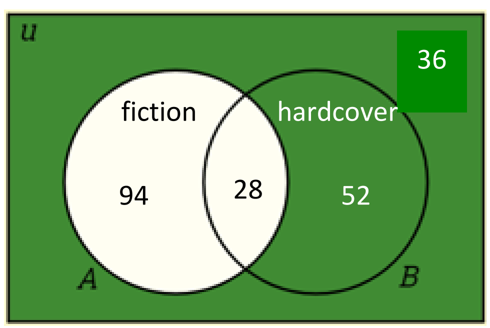
	Fiction	Non-fiction	
Hardcover	28	52	80
Paperback	94	36	130
	122	88	210

$$P(\text{fiction } \cup \text{hardcover}) = \frac{94 + 28 + 52}{210}$$

$P(A^C)$

The ***complement*** of event A. (Not A)

We can think of the complement of A as $1 - P(A)$



	Fiction	Non-fiction	
Hardcover	28	52	80
Paperback	94	36	130
	122	88	210

$$P(\text{Not Fiction}) = \frac{88}{210} = 0.42$$

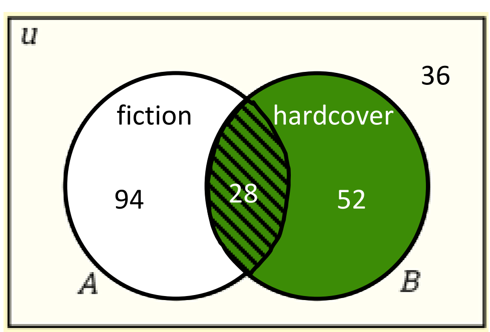
$$1 - P(\text{Fiction}) = 1 - \frac{122}{210} = 1 - 0.58 = 0.42$$

$P(A|B)$

The ***conditional probability*** of event A given that event B has occurred.

This is the probability of an event only out of a certain event.

How many are fiction out of those that are hardcover?



	Fiction	Non-fiction	
Hardcover	28	52	80
Paperback	94	36	130
	122	88	210

$$P(\text{Fiction}|\text{Hardcover}) = \frac{28}{80}$$

Practice Problems – *Check the answers to the odd-numbered ones in the back of the packet!*

There are 434 juniors and seniors at Haslett HS. Of those, 38 are enrolled in AP Stats, 79 in AP Psych, and 15 are taking both courses.

	AP Stats	No AP Stats	
AP Psych	15	64	79
No AP Psych	23	332	355
	38	396	434

Find the following probabilities.

1. $P(\text{APP} \cap \text{APS})$

2. $P(\text{APP} \cup \text{APS})$

3. $P(\text{APP}^{C})$

4. $P(\text{APP} \mid \text{APS})$

Is taking APP independent of taking APS? In other words, is a student just as likely to take AP Psych without taking AP Stats as they are if they are taking AP Stats?

NO! Of course not! Why? If a student is taking one AP, they are likely to take more. There is a connection. Mathematically speaking, you have learned the independence rule:

$$P(A \cap B) = P(A) * P(B)$$

5. Use the independence rule to show that taking AP Stats and AP Psych are not independent.

Now the final question… is doing well in AP Statistics independent of doing well in other AP courses?

ANSWER KEY

Part 1: Rounding

1) 12.8	3) 0.034	5) 25.7 *or* 25.69 *or* 25. 690 *or* 25.6895	7) 0.0028

Part 2: Fractions, Decimals, and Percentages

NOTE: *Your answers may be slightly different due to rounding. That's fine, as long as you rounded correctly*

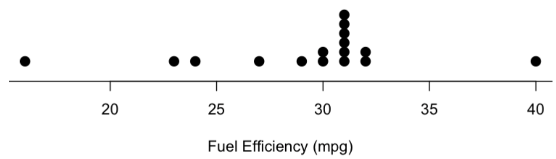
1) 4.33	3) 0.7	5) 0.2245	7) 67.2%	9) 16%	11) 15.308

Part 3: Summary Statistics – Center and Spread

1) Symbol of Mean: μ Value of Mean: 8 years Position of Median: 8th Value of Median: 5 years	3) Symbol of Mean: μ Value of Mean: 71.9 degrees Position of Median: 5th & 6th Value of Median: 72 degrees

Part 4: Graphing Data – Dot plots

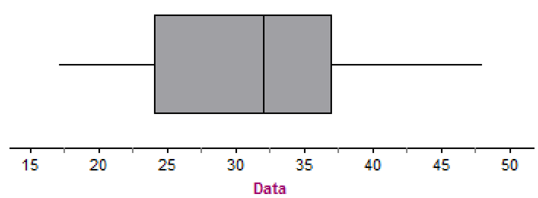
1)



3) 69, 87, 88, 88, 89, 90, 96, 98, 97, 100

Part 5: Graphing Data – Boxplots

1. Min = 17, Q1 = 24, Med = 32, Q3 = 37, Max = 48



3. A. 42 chickens   C. 84 chickens

Part 6: Ways of Counting

1) 125,000	3) 56	5) 1	7) 120	9) 0

Part 7: Probability: Theoretical and Experimental Probability

1) S={HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}	3) 0.25	5) 0.875	7) 0.267	9) 0.867

Part 8: Probability: Two-Way Tables and Venn Diagrams

1) 0.035	3) 0.818