Math Fundamentals for Statistics II (Math 95)

Unit 4: Connections

Scott Fallstrom and Brent Pickett
“The ‘How’ and ‘Whys’ Guys”

2nd Edition (Summer 2016)

Table of Contents

4.1: Solving Equations

One key skill that is needed in statistics is the ability to solve equations. In Math 52, we used the box method as well as an algebraic method. Many of the problems relate to equations like this:

$y = mx + b$ or $z = \dfrac{x - \mu}{\sigma}$

The first equation, $y = mx + b$, is used with lines. Remember that m represents the slope or the rate of change, and b represents the y-intercept.

The second equation, $z = \dfrac{x - \mu}{\sigma}$, is used when normalizing data in statistics. While we may not even know what that means at this stage, the skill we need is not understanding the equation but knowing how to solve it.

Example (1): Solve $y = 2x + 7$ when…

A) $x = 7$

B) $y = 4$

Solution:

A) Since $y = 2x + 7$, just substitute in the $x = 7$.

$y = 2x + 7 \Rightarrow y = 2(7) + 7 \Rightarrow y = 14 + 7 \Rightarrow y = 21$

B) Since $y = 2x + 7$, just substitute in the $y = 4$.

$y = 2x + 7 \Rightarrow 4 = 2x + 7$. We want to get x by itself, so next we can subtract 7 on each side and then divide by 2 on both sides.

$4 = 2x + 7 \Rightarrow 4 - 7 = 2x \Rightarrow -3 = 2x \Rightarrow x = -\dfrac{3}{2} = -1.5$

EXPLORE (1)! Solve for the missing piece in the equation $y = 3x - 5$ if we know…

A) $x = -6$

C) $y = 10$

B) $x = 11$

D) $y = -11$

Example (2): Solve $z = \frac{x-11}{2}$ when…

A) $x = 7$

B) $z = 4$

Solution:

A) Since $z = \frac{x-11}{2}$, just substitute in the $x = 7$.

$$z = \frac{x-11}{2} \Rightarrow z = \frac{7-11}{2} \Rightarrow z = \frac{-4}{2} = -2$$

B) Since $z = \frac{x-11}{2}$, just substitute in the $z = 4$.

$4 = \frac{x-11}{2}$. We want to get x by itself, so next we can multiply by 2 on each side and then add 11 on both sides.

$$4 = \frac{x-11}{2} \Rightarrow 4 \cdot 2 = \left(\frac{x-11}{2}\right) \cdot 2 \Rightarrow 8 = x - 11 \Rightarrow 8 + 11 = x - 11 + 11 \Rightarrow 19 = x$$

EXPLORE (2)! Solve the equation $z = \frac{x-7.5}{1.5}$ if we know…

A) $x = 6$

C) $z = 10$

B) $x = 0$

D) $z = -3$

EXPLORE (3)! Solve the equation $z = \frac{11-\mu}{1.5}$ if we know…

A) $\mu = 6$

B) $z = 4$

In statistics, it's imperative to be able to read charts and tables like what we've seen in this class.

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	0.7704	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

In this table you can look up a *z*-number on the outside and find a value in the body of the table.

Example (3): What is the value associated with a **z-score** of 1.47?

Solution: Look up 1.4 on the left column, then move in until you get to 0.07 and the result is 0.9292.

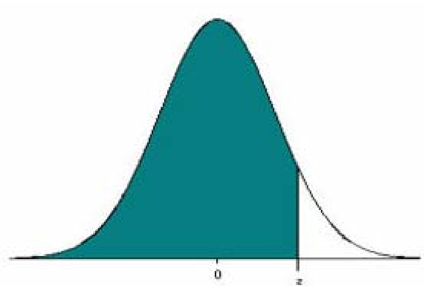
EXPLORE (4)! Find the values from the table:

A) What is the value associated with a *z*-score of:

i. 2.15 ii. 0.43 iii. 1.87

B) What *z*-score links with a body-value of…

i. 0.9830 ii. 0.9656 iii. 0.6141



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	0.7704	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

Another equation that comes up in statistics helps to explain the error. $E = z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$

In the formula, *n* represents the number of people in the sample while *E* is the error, and the $z_{\frac{\alpha}{2}}$ portion is found using the table above. The **<u>level of confidence</u>** is called *α*, and is often 90%, 95%, or 99%. Here's how you use it: take the level of confidence and divide it by 2. Then add 50% to that result. Look up that number in the body of the table. If the number is between two values, then choose the value that is halfway between.

<u>Example (4):</u> Find the $z_{\frac{\alpha}{2}}$ corresponding to a confidence level of 90%.

<u>Solution:</u> Take 90% and divide by 2: 45%. Add 50% to get 95%. Look up 95% = 0.9500 in the table. It's between 0.9495 (1.64) and 0.9505 (1.65), so we pick the value halfway between: 1.645.

EXPLORE (5)! Find the $z_{\frac{\alpha}{2}}$ corresponding to:

A) 95%

B) 99%

Example (5): Solve $E = z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$ if we know the level of confidence is 95%, $\sigma = 12.4$, and n = 850, find the value of E.

Solution: For 95%, the $z_{\frac{\alpha}{2}} = 1.96$. Plug in all the values and grab that calculator!

$$E = z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow E = (1.96)\left(\frac{12.4}{\sqrt{850}}\right) \approx 0.83362.$$

We could also solve this formula for n if we know the values of everything else.

Example (6): Solve $E = z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$ if we know the level of confidence is 99%, $\sigma = 17.7$, and E = 0.48, find the value of n.

Solution: For 99%, the $z_{\frac{\alpha}{2}} = 2.575$. Plug in all the values and grab that calculator!

$E = z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow 0.48 = (2.575)\left(\frac{17.7}{\sqrt{n}}\right) \Rightarrow \frac{0.48}{2.575} = \left(\frac{17.7}{\sqrt{n}}\right)$ Since the variable we want is on the bottom, we can use our proportion skills from Math 52.

$$\frac{0.48}{2.575} = \left(\frac{17.7}{\sqrt{n}}\right) \Rightarrow 0.48\sqrt{n} = 2.575(17.7) \Rightarrow \sqrt{n} = \frac{2.575(17.7)}{0.48} \Rightarrow n = \left(\frac{2.575(17.7)}{0.48}\right)^2 \approx 9016.0959$$

So there were about 9,016 people in the survey.

EXPLORE (6)! Solve these equations for the missing piece in $E = z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$.

A) the level of confidence is 95%, $\sigma = 9.68$, and $n = 611$

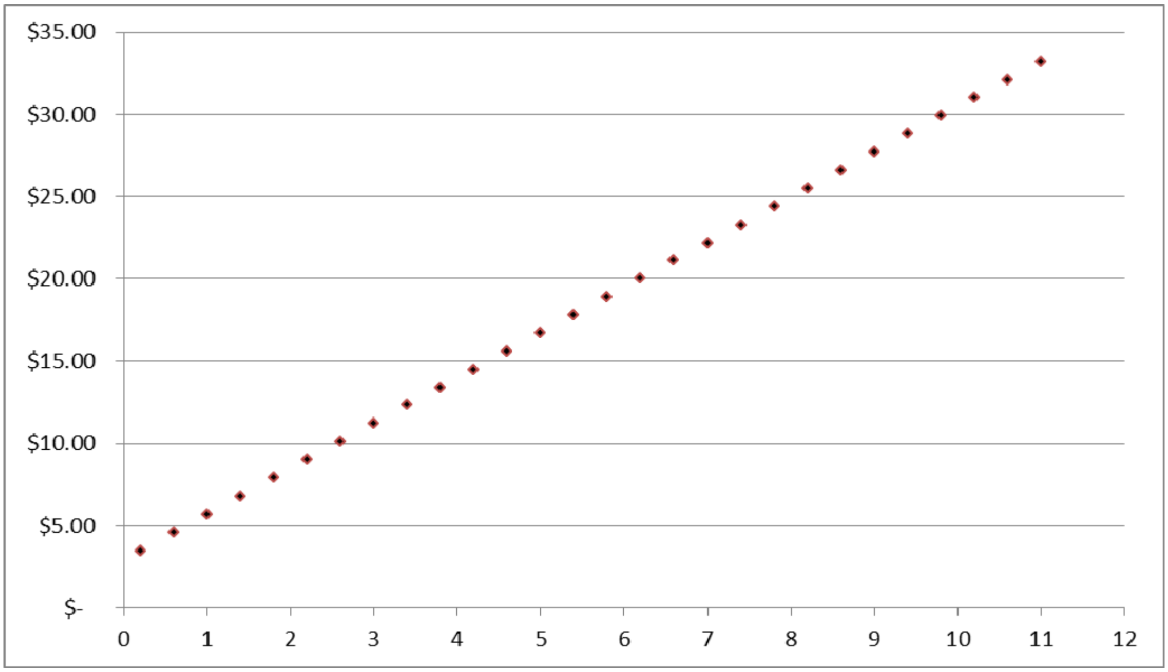
B) the level of confidence is 95%, $\sigma = 19.68$, and $E = 0.775$

C) the level of confidence is 90%, $\sigma = 5.44$, and $E = 0.325$

D) the level of confidence is 90%, $n = 544$, and $E = 0.325$

4.2: Lines and Applications

The following graph shows the relationship between a trip in miles (the x-axis) and the cost in dollars (the y-axis) for a taxi company. Draw a line that fits the data well on the graph.



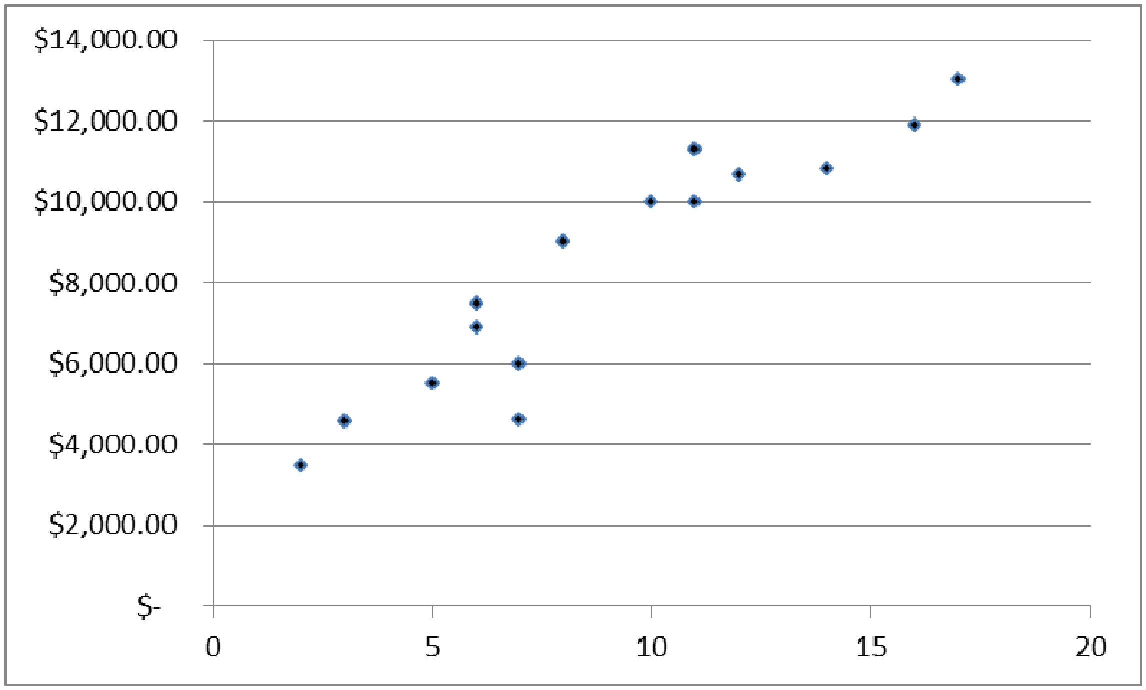
Interactive Example (1): Find the y-intercept and the slope. Use these to create the equation of the cost. Then use the equation to find the cost of a 25 mile cab ride.

EXPLORE (1)! Refer to the previous graph:

A) In the previous problem, what does the slope of the line represent?

B) What does the y-intercept of the line represent?

C) Use a sentence to explain what the point (14, 41.45) represents related to this problem.



<u>EXPLORE (2)!</u> Above are prices of a Ford Mustang found on TrueCar. The y-axis has the price in dollars and the x-axis has the model years after 1997.

A) Draw the line above that best represents the data.

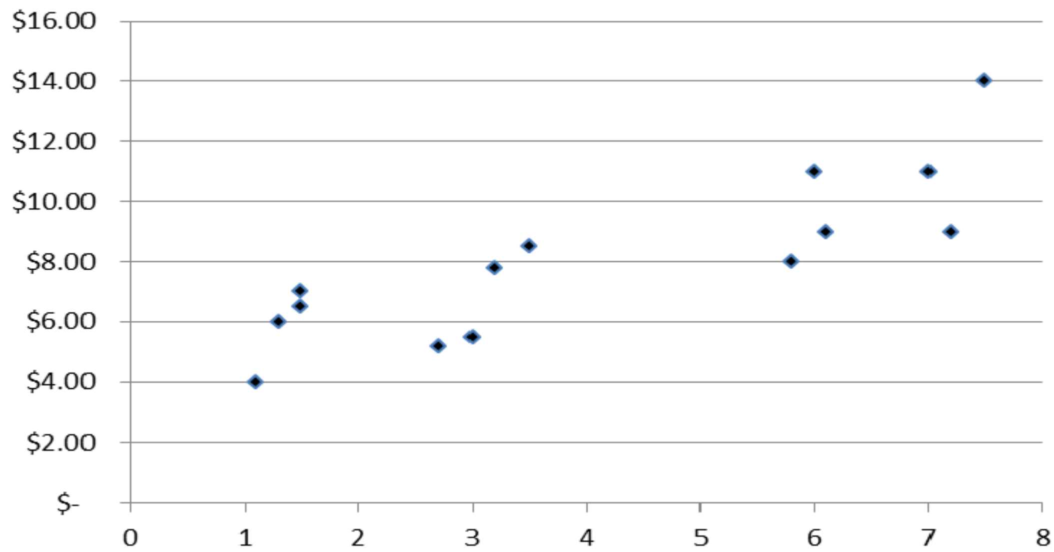
B) Use the line to find the slope and y-intercept. Interpret each of them in the context of the graph.

C) Create a linear equation using the information you found in the previous part.

D) Use your linear equation to approximate the cost of a 2006 Ford Mustang.

E) Use your linear equation to approximate the cost of a 2020 Ford Mustang.

When you have graphed the line, you may see a positive slope or a negative slope. This slope indicates that there is a **positive relationship** between the variables (as one increases, so does the other). If there is a negative slope, then there is a **negative relationship** between the variables (as one increases, the other decreases).



EXPLORE (3)! Uber prices for trips based on mileage are given above. Create a line that is the best fit for these even though it won't be a perfect fit.

A) Draw the line above that best represents the data. As the miles go up, the cost ___________?

B) Use the line to find the slope and y-intercept. Interpret each in the context of the graph.

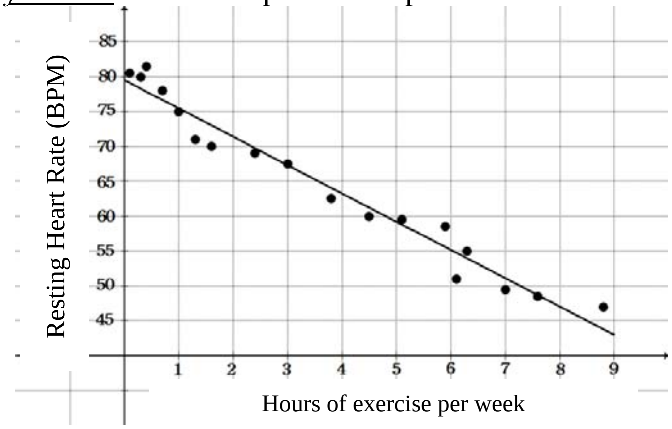
C) Create a linear equation using the information you found in the previous part.

D) Use your linear equation to approximate the cost of a 4 mile trip with Uber.

E) Use your linear equation to approximate the cost of a 20 mile trip with Uber.

F) Are you more confident with your estimate for the 4 mile or 20 mile trip? Explain.

Example (2): For the following scatterplot, draw a line and create the equation of the line *as a function*. Then interpret the slope of the line with a sentence.



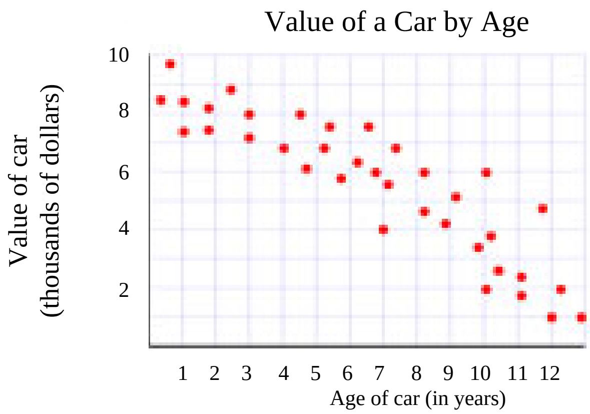
The y-intercept is close to 80, so let's approximate as 79 for the point (0,79). Another point is (6,55).

So we'll have to find the slope: y-value decreases by 24 and x-value increases by 6. $m = \frac{-24}{6} = -4$.

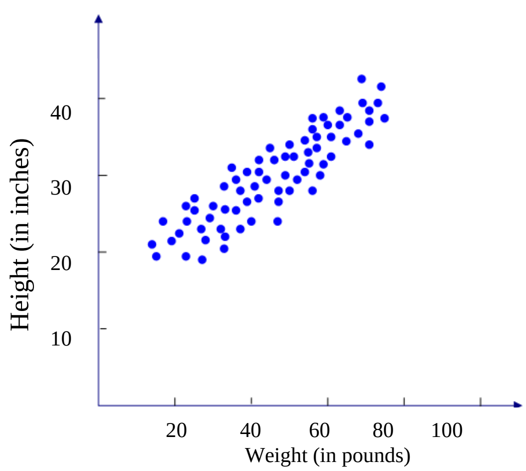
Our function would be: $HR(x) = -4x + 79$ where HR is beats per minute and x is hours of exercise per week. The slope means that for each additional hour of exercise that you do per week, your resting heart rate will go down by 4 beats per minute.

EXPLORE (4)! Try some of your own. Draw a line – create a function – interpret the slope and y-intercept for both.

A) **



B)



Follow up: Based on your model in part (B), predict the height of someone weighing 125 lbs?

<u>4.3 Basics of Statistics – Weighted Averages</u>

As we are preparing you for statistics, there are certain statistical concepts that are fairly basic and some that you've possibly already seen. For us, we'll be thinking of **<u>descriptive statistics</u>** – meaning that we will get some data and just describe what's going on with it. Later in your coursework, you may be asked to deal with **<u>inferential statistics</u>** – which is where we look at the data and try to figure out some additional information, or possibly predict what would happen in other situations.

In order to describe information, or data, we may want to see what is happening on average. However, the word average is problematic because there are different interpretations of average. The 3 most popular meanings are:

- The middle value. Mathematicians call this the **<u>median</u>**, and there can be only one. In order to find the median, we need to put the data in order (big to small, or small to big).
- The value occurring the most. Mathematicians call this the **<u>mode</u>**, and there can be more than one.
- The sum of all the numbers (data) divided by the number of data values. Mathematicians call this the **<u>mean</u>**, and there can be only one.

<u>Example (1):</u> Find the mean, median, and mode of: 5, 8, 7, 11, 11, 9, 8, and 11.

<u>Solution:</u>

- <u>Mode:</u> The mode here is 11 because it occurs more than any other data value.
- <u>Median:</u> First, put the numbers in order: 5, 7, 8, 8, 9, 11, 11, 11. The middle number would be between the 8 and the 9, so we just find halfway between these. $\frac{8+9}{2}=\frac{17}{2}=8.5$. For our data set, half the numbers will be above 8.5 and half will be below 8.5.
- <u>Mean:</u> To find the mean, add up all the data values. Then divide by 8 because there are 8 numbers. It looks a little strange, but the formula we use is $\bar{x}=\frac{\Sigma x}{n}$. Σx is a shorthand way of saying "add up all the numbers." Here, $\bar{x}=\frac{5+8+7+11+11+9+8+11}{8}=\frac{70}{8}=8.75$.

<u>EXPLORE (1)!</u> Try some on your own – find the mean, median, and mode.

A) 18, 10, 6, 18

B) 18, 10, 6, 18, 50

C) Pop up your computers and find the mean, median, and mode for: 12, 12, 14, 15, 18, 18, 19, 20, 21, 23, 11, 16, 18, 23, 31, 14, 20, 18, 12.

There are some interesting applications of these principles. See if you can interpret these ideas.

<u>Interactive Examples (2):</u>

A) The median price of a home in Carlsbad is $550,000. Interpret this with a sentence.

B) The mode of the home prices in Carlsbad is $389,000. Interpret this with a sentence.

C) The mean price of a home in Carlsbad is $780,000. Interpret this with a sentence.

D) How could the mean be higher than the median? Explain.

E) In a recent classified ad, jobs at a fast food establishment were posted. The average hourly rate for the store was $26 per hour. Does this seem like a good place to work (based on the wage)? Explain – and include what "average" you think was used in the ad.

F) If a store had 35 employees making $9 per hour, 3 supervisors make $50 per hour, a general manager making $200 per hour, and an owner making $375 per hour. What is the mean rate of pay? Does this example change your response to the previous part?

The last example on the previous page was an example of a **weighted average**. This is where there are many data values at one number and adding those up is easier to do with multiplication.

Score on Quiz	6	7	8	9	10
Number of Students	2	5	7	4	2

To find the mean score on the quiz with the information above, we need to find the sum of all the scores. Instead of adding 6 + 6 + 7 + 7 + 7 + 7 + 7 + 8 + 8 + …. (which is really long), let's do some multiplication!

$\bar{x} = \frac{6(2)+7(5)+8(7)+9(4)+10(2)}{20} = \frac{159}{20} = 7.95$. This is much faster than doing all that addition!

Interactive Example (3): Find the mean score on the quiz.

Score on Quiz	6	7	8	9	10
Number of Students	2	8	17	5	3

We also saw weighted averages when computing expected value in a previous unit.

For roulette, we saw the expected value of a 2 number bet on a standard 00 wheel – we could create a table of outcomes. For a $1 bet, this pays out at 17:1 so the table looks like:

Option	Number of ways	Value
Win	2	$17
Lose	36	$ – 1

Example (4): Find the weighted average of this bet.

Solution: $\bar{x} = \frac{2(\$17)+36(\$-1)}{38} = \frac{-2}{38} \approx \$-0.052632$

So considering all options, the weighted average becomes exactly the same idea as Expected Value from the probability unit.

EXPLORE (2)! Find the weighted average (expected value) of a standard (00-wheel) roulette bet of:

A) 3 numbers (pays 11:1)

B) 12 numbers (pays 2:1)

C) Red – covers 18 numbers (pays 1:1)

Weighted averages also come up when calculating grades. GPA is a weighted grade point average. Each credit earns a grade, and those are averaged with the mean.

<u>Interactive Example (5):</u> Calculate the GPA for the student below.

Class	Credits	Grade (letter)	Grade (points)	Quality Points
Math	4	B	3.0	
English	3	C	2.0	
Chemistry	4	A	4.0	
Yoga	1	D	1.0	
Totals				

A) Calculate the GPA: $\text{GPA} = \dfrac{\text{Sum of Quality Points}}{\text{Credits}} =$

Once this is done for a term, we could keep track of Quality Points and Credits, then calculate an overall GPA. Each term, you will have both: term GPA and overall GPA.

B) What happens to a student who has the same letter grades, but in different classes? Will their GPA be the same?

<u>EXPLORE (3)!</u> Calculate the GPA for the student below.

Class	Credits	Grade (letter)	Grade (points)	Quality Points
Math	4	B	3.0	
English	3	C	2.0	
Chemistry	4	D	1.0	
Yoga	1	A	4.0	
Totals				

A) Calculate the GPA: $\text{GPA} = \dfrac{\text{Sum of Quality Points}}{\text{Credits}} =$

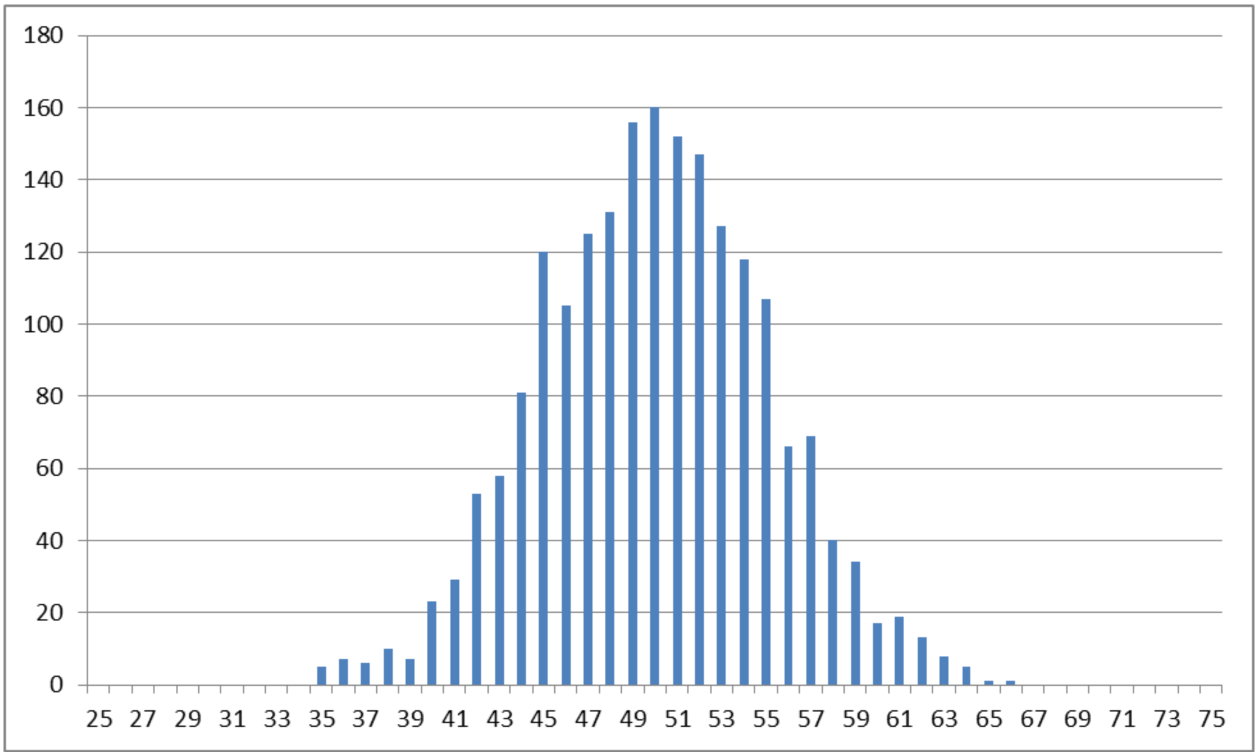
B) If you were going to get a low grade, explain what class you want to get the low grade in?

4.4: Likelihood Analysis

Previously in the probability section, there was a problem related to flipping a coin 100 times and counting heads and tails. If we repeat the experiment over and over again (each experiment has coin flipped 100 times), then the experimental probability keeps going.

The excel spreadsheet on sent out by your instructor was made to mimic this and display the results. Open that spreadsheet and discuss the results. The graphs are listed below.

One coin being flipped 100 times results in one trial where the number of heads was counted. Then Scott used the computer to perform 2,000 trials, and each trial records the number of heads that showed up in the 100 coin flips. In total, this would be 200,000 coin flips which is really challenging to perform by hand!



EXPLORE (1)!

A) Based on the picture, how likely is it that you end up with 50 heads exactly?

B) Would you consider this likely or not?

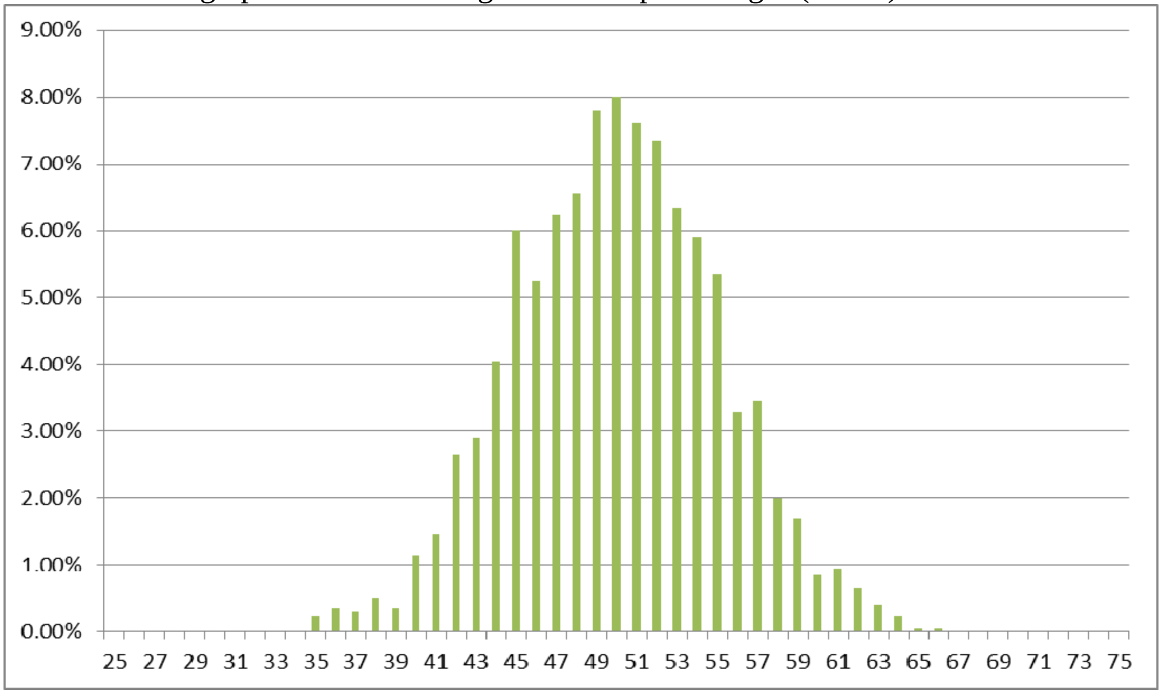
C) Is it the most likely event in the experiment?

EXPLORE (2)! When we have a set of trials and the data to this point, it is sometimes helpful to find the cumulative percentages – which means we add up the totals for certain points.

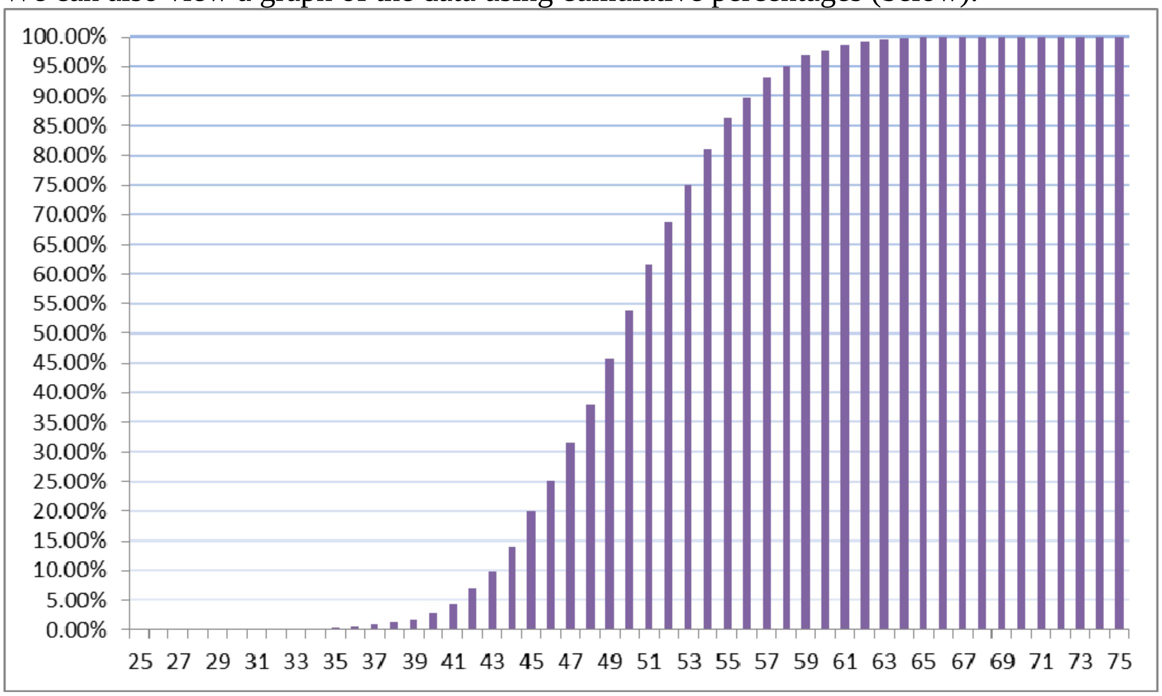
For each of the dark outlined boxes, find the sum of all the percentages up to that point. The first is done for you.

Number of Heads in a Trial of 100 coin flips	Percentage of All Trials	Cumulative percentages
34	0.00%	
35	0.25%	
36	0.35%	
37	0.30%	
38	0.50%	1.4%
39	0.35%	
40	1.15%	
41	1.45%	
42	2.65%	
43	2.90%	
44	4.05%	
45	6.00%	
46	5.25%	
47	6.25%	
48	6.55%	
49	7.80%	
50	8.00%	
51	7.60%	
52	7.35%	
53	6.35%	
54	5.90%	
55	5.35%	
56	3.30%	
57	3.45%	
58	2.00%	
59	1.70%	
60	0.85%	
61	0.95%	
62	0.65%	
63	0.40%	
64	0.25%	
65	0.05%	
66	0.05%	
67	0.00%	

We can view a graph of the data using individual percentages (below).



We can also view a graph of the data using cumulative percentages (below).



EXPLORE (3)! What is the probability of flipping 100 coins and obtaining…

A) 47 or fewer heads

C) 62 heads or more

B) Exactly 47 heads

D) Exactly 62 heads.

EXPLORE (4)! Remember that about 5% was where we thought an event was "strange" in the probability unit.

A) If we used the full 5% on the bottom end, where would you say it is "rare" to have the outcome?

B) If we used the full 5% on the top end, where would you say it is "rare" to have the outcome?

C) If we think of it this way, and cut it in half (some above some below), we'd have 2.5% on the top end and 2.5% on the bottom end. Using the previous charts, where would you say it is "rare" to have the outcome?

EXPLORE (5)! Relate these back to the coin flipping from the previous few pages.

A) If someone says that they flipped a coin 100 times and ended with 46 heads, does this seem reasonable? Explain why or why not and link this to the likelihood it will happen.

B) If someone says that they flipped a coin 100 times and ended with 38 heads, does this seem reasonable?Explain why or why not and link this to the likelihood it will happen.

C) If someone says that they flipped a coin 100 times and ended with more than 55 heads, does this seem reasonable? Explain why or why not and link this to the likelihood it will happen.

In a statistics class, the notation is often used as $P(X)$ which represents the **probability of a random variable** X. If we let X represent the number of heads that are flipped out of 100 coin flips, then $P(X=50)$ would represent the probability that we will flip 50 heads out of 100 coins. Using the tables on the previous charts, $P(X=50)=\frac{160}{2{,}000}=0.08$. In a sentence, we would say that the probability of flipping 100 coins and getting 50 heads is about 8%.

EXPLORE (6)! Determine the values of the following based on the coin flipping examples and interpret the result with a sentence. Use the appropriate table from this section.

A) ** $P(X=52)$

D) ** $P(X \geq 39)$

B) $P(X=40)$

E) $P(X \geq 60)$

C) $P(X<40)$

F) $P(X<55)$

4.5: Critical Thinking with Numbers – Simpson's Paradox

As we get closer to statistics, the ability to work with numbers remains important, but what will become even more important is the ability to interpret the results. We conclude with an example showing the importance of interpretation.

In 1973, the University of California, Berkeley was sued for gender bias. Here is the main reason for the suit: there were more men applying to graduate school, but they were being accepted at a higher rate. The difference was so high (~10%) that it was claimed this was unlikely to be due to chance.

	Applicants	**Admitted**
Men	8442	44%
Women	4321	35%

When you look at the big picture like this, one conclusion can be made – those admitting students were biased against women. However, it was important to notice that this is the big picture and rates should be considered for individual departments. The individual department data shows a much different picture.

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

The interesting thing here is that the specific data actually shows many departments had a bias in favor of women. This type of phenomenon is known as **Simpson's Paradox**, and occurs when one group appears better in multiple smaller cases, yet when combined, the other group appears better.

This exact idea has been found in sports as well – baseball is a great example! In baseball, a player obtains a batting average determined by their hits divided by at-bats, then rounded to 3 decimal places. For example, if you had 27 hits in 79 at bats, your batting average would be

$$\text{BA} = \frac{21}{79} = 0.2658227848... \approx 0.266 .$$

Ken Ross (University of Oregon) wrote a book about Math and Baseball showing another example. In 1995 and 1996, Derek Jeter and David Justice both played Major League Baseball (MLB).
Here are their individual batting averages:

	1995		**1996**	
Derek Jeter	12 out of 48	0.250	183 out of 582	0.314
David Justice	104 out of 411	**0.253**	45 out of 140	**0.321**

If we thought about combing those together, who do you feel *should* have the higher batting average?

Well, it sure seems that David Justice should have the higher batting average since he had the higher batting average in both 1995 and 1996. Let's just check and see:

	1995		**1996**		**Combined**	
Derek Jeter	$\frac{12}{48}$	0.250	$\frac{183}{582}$	0.314	$\frac{195}{630}$	**0.310**
David Justice	$\frac{104}{411}$	**0.253**	$\frac{45}{140}$	**0.321**	$\frac{149}{551}$	0.270

Was this what you expected – or the opposite? To think that someone wins a head-to-head battle in each year but loses overall, by quite a bit actually, is counter-intuitive. This is why we need the ability to look deeper and not take things at face value.

EXPLORE (1)! Determine the correct answer to the following question. In a school district, there are two high schools. In the *first* high school, the graduation rate of girls is higher than the graduation rate of boys. In the *second* high school, the graduation rate of girls is also higher than the graduation rate of boys. Does it follow that the graduation rate of girls for the *district* is higher than boys? (circle your choice)

A) For the whole district, the graduation rate of girls is *higher* than the graduation rate of boys.
B) For the whole district, the graduation rate of girls is *lower* than the graduation rate of boys.
C) For the whole district, the graduation rate of girls is *the same as* the graduation rate of boys.
D) There is not enough information to determine which rate is higher.

For more information about Simpson's Paradox and the mathematics behind it, see the article below:
http://www.ima.umn.edu/~jberwald/research/Syn_bandyopadhyay.pdf
Page 18 of this article shows the example above with the graduation rates.

EXPLORE (2)! In the article above, the authors indicated that (A) was the most likely choice. Look up the answers to these questions in the pdf link above.

- What percent answered (A) in their survey of philosophy students?
    - Between 60% and 70%
    - Between 70% and 80%
    - Between 80% and 90%
    - Between 90% and 100%

- What percent answered (D)?
    - Between 0% and 5%
    - Between 5% and 10%
    - Between 10% and 15%

<u>4.6: Wrap-up and Review</u>

Key concepts in the courses:

- Gaining an appreciation for number sense
- Solving linear equations
- Graphing lines and using slope and intercept to interpret the relationships between input and output
- Likelihood and probability
- Expected values and weighted averages
- Relationships using Venn diagrams
- Understanding conditional statements and using correct reasoning for arguments
- Financial awareness and the basics of investing, borrowing, and saving for retirement
- A clear understanding of the time-value of money
- Using tables, charts, graphs, and formulas to solve problems
- Being comfortable with a computer for data entry (and financial predictions)
- Through repeated problem types, seeing that there is often more than one way to solve a problem
- Gaining more confidence in computation, interpretation, and understanding the concepts and usefulness of mathematics

You made it! Thank you for taking this mathematical journey with us. Remember that you are in charge of how you approach a problem – there often is more than one way to solve it! Always ask why when you don’t understand; there is a reason why we do things in math and statistics.

Math is not your enemy and maybe even some times it is your friend. Math can help show you how to save money for your future, think critically and logically, and predict risk with probability. We hope that you found the information useful for future classes as well as in everyday life. And maybe, just maybe, you had a little fun in the process.

While it may be a goal for you to enjoy math more, we will be very happy if you hate math less. Now, go out and rock that statistics class. YOU CAN DO IT! ☺

INDEX (in alphabetical order):