Agricultural Sciences
Waterford Union High School

STATISTICS

By C Kohn, L Hitsman, Waterford WI

STATISTICS

- **<u>Statistics</u> is the collection, analysis, and interpretation of the data in order to determine the specific outcome of an experiment or survey.**
  - Scientists use statistics to determine whether or not their experimental results are reliable and repeatable.
  - Scientists also use statistics to determine whether or not they can trust the accuracy of their results.

- **Knowing whether the trends of an experiment are a one-time occurrence or will occur every time is critical in determining the dependability of an experiment.**
  - An experiment with results that vary each time is not dependable.
  - A dependable experiment is one that gets similar results each time it is performed.

RELIABILITY = SIMILARITY

- **Two factors affect whether or not the results of an experiment will be similar each time.**
  - **Amount of Data**: the more data we have, the more likely we will have similar averages from our experiments each time it is performed.
  - **Similarity of Data:** the more alike our data, the more likely the experiment will have similar results each time.

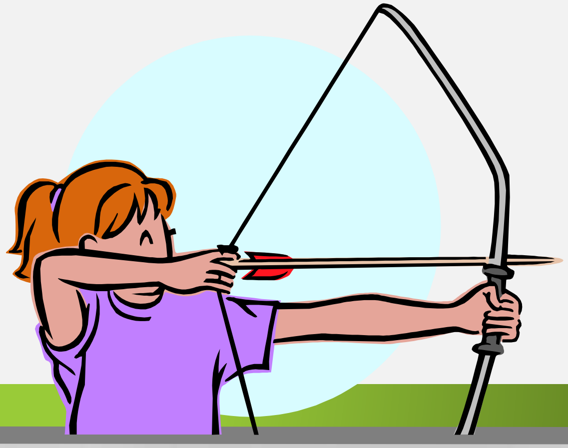
- **Scientists use data to answer to key questions:**
  - 1. How do I know that I am not wrong?
  - 2. How do I know that I will be right *every* time?

opticalres.com

STATISTICS

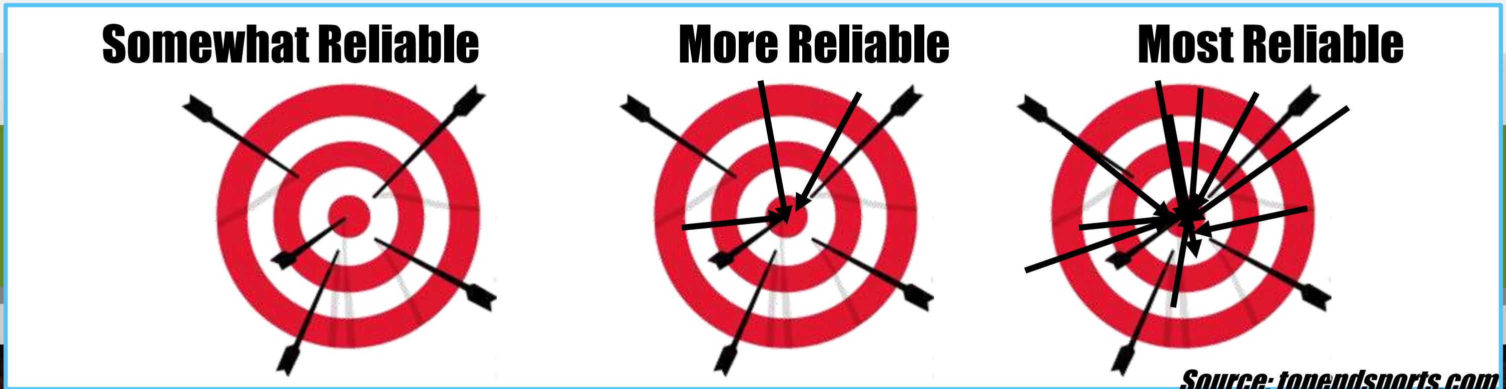
- **Data in research is sort of like target practice.**
  - When you are shooting at a target, you want all of your shots to be close together.
    - *The closer your shots are to each other, the better.*
  - You also need to have lots of shots to ensure that a shooter is accurate and not just "lucky".
    - *If a shooter only takes one shot and they hit a bulls-eye, it might be accuracy or it might be luck.*
    - *If a shooter takes lots of shots and all are on the bulls-eye, we know that the shooter is accurate.*

- **Statistics are similar to target practice: the more numbers we have, and the more similar each number is to each other, the better.**
  - An archer wants lots of arrows all close together on the target.
  - A researcher wants a large amount of measurements, and they want those measurements to be as similar to each other as possible.



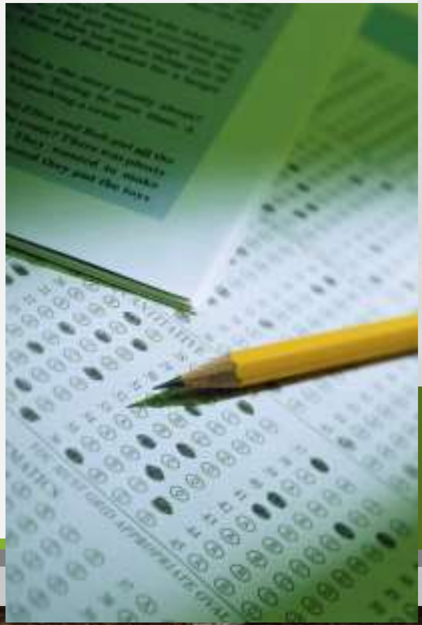
RELIABLE VS. UNRELIABLE

- **An experiment with very little data and/or data that varies a lot will have results that are unreliable.**
  - If we don't have a lot of data, or if that data varies a lot, we will be unable to predict how that experiment would turn out even if we repeated it under the exact same conditions.
- **However, if an experiment has lots of data AND that data is very similar, it is likely that the experiment *will* have reliable results if performed again.**
  - This tells a researcher that their experimental results will be similar and predictable every time their experiment is repeated.
  - These results are then considered reliable.



AVERAGES

- **In an experiment, the most important kind of statistic is the mean (or average) of the data.**
  - Mean is the same as average.
  - It is calculated by dividing the sum of the numbers by the sample size.
    - *Mean = (Sum of Data)/(Sample Size)*
    - *The mean of 1, 2, & 3 would be (1+2+3)/3 = 2. The mean is 2.*



- **The mean is important because provides information about every measurement in a single number.**
  - For example, a teacher would know if their class understood the material in a unit by the average score from the unit quiz.

AVERAGES CONTINUED

- **If the average score of a quiz is high, a teacher does not have to look at each score individually to see if the class did well.**
  - A teacher can look at the average quiz score and assume that the class as a whole understood the material if the average is high.
  - While a couple students might need to relearn the material and retake the quiz, a teacher can assume they can move on to the next unit with a high average.

- **On the other hand, if the average quiz score is low, a teacher knows they should re-teach that unit (even if a couple students scored pretty well).**
  - Because they have the mean quiz score, a teacher can focus primarily on this number (the mean) to determine if their teaching was effective.

AVERAGES CAN CHANGE

- **When we take the average of something, we are using a number that can change as our data changes.**
  - For example, if we wanted to know the mean height of this class. We would…
    1. *Record each person's height.*
    2. *Add all the heights together.*
    3. *Divide by the number of students we have.*

- **However, the mean height of class can change.**
  - If we gained or lost a student, the mean or average height would be a new number.
  - If a student had a growth-spurt, their new height would change the average height if it was calculated a few months later.
  - The "average height" is not one number; it can change each time it is measured!

CLASS AVERAGE

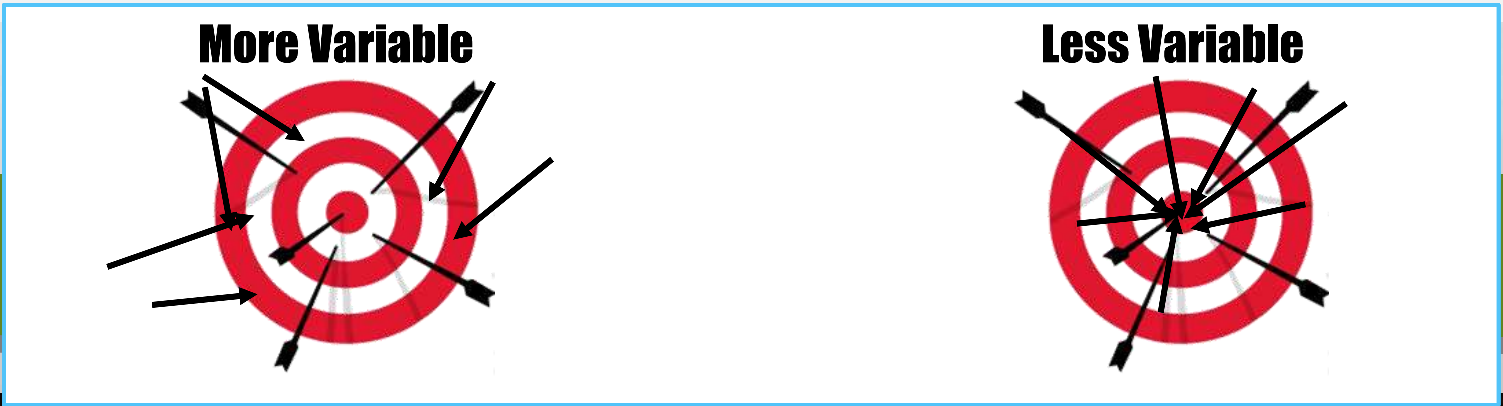
- **If our class did not have very many students, the addition of one more person's height would have a big impact on the calculated average.**
  - On the other hand, if we had 1000 students in our class, the addition of one more person's height would hardly change the calculated average.
- **If a new person's height was very similar to the average, our calculated average would not change much.**
  - On the other hand, if they were 6'7", our calculated average would change a lot more.

www.presentermedia.com

VARIABILITY

- **Whether or not the numbers in data are similar to each other (or vary widely) is known as <u>variability</u>.**
  - Variability is the opposite of similarity in data.

- **The *more* similar our data, the *less* variable it is. The *less* variable the data, the *more* reliable it is.**
  - Just like an archer wants all of their arrows to be close together, we want the numbers in our data to be as similar as possible.
  - In other words, we want our data to have low variability.

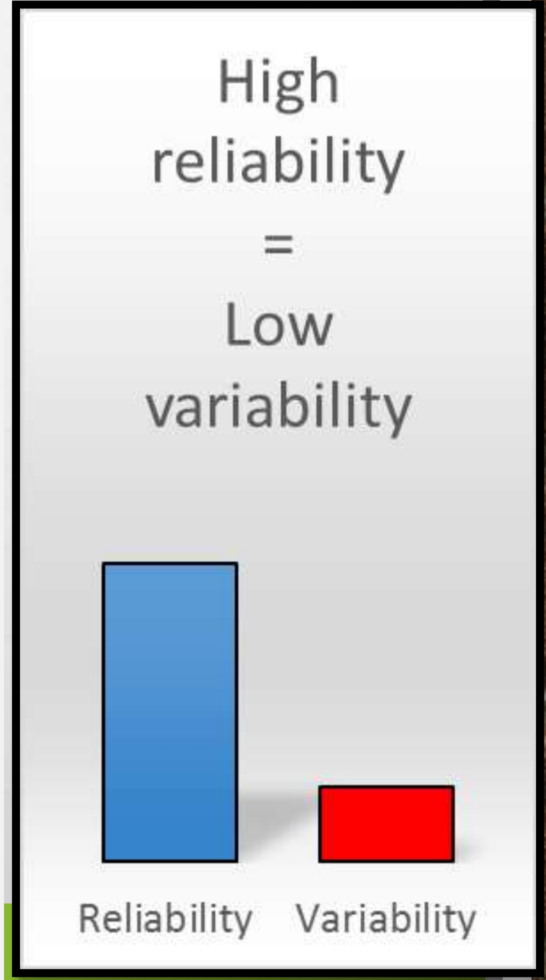


Source: topendsports.com

VARIABILITY

- **The less similar our data, the more variable it is. The more variable the data, the less reliable it is.**
  - We know that if an archer has arrows all over the target, they are not a reliable shooter.
  - Similarly, if the numbers in our data are highly variable and very dissimilar, our experiment does not have reliable results.

- **The lower the variability of our data, and the greater the amount of data that we have, the more reliable and repeatable our data.**
  - An experiment with large amounts of data and with minimal variability is a reliable experiment.
  - An experiment with little data and/or data with lots of variability is not a reliable experiment.



STANDARD DEVIATION

- **Variability can be measured using a statistical method known as Standard Deviation.**
  - Standard deviation is the “average variability”.
  - **Standard deviation** is a measurement of how similar or dissimilar the numbers from an experiment are to each other.

- **A *low SD* score means your data is all very similar.**
  - These corn plants would have low SD

More reliable data

- **A *high SD* score means your data is very dissimilar.**
  - These corn plants would have high SD

Less reliable data

STANDARD DEVIATION

- **Standard deviation is calculated by measuring the difference between the average and each individual measurement and taking the average of these differences.**
  - The lower the standard deviation value, the more reliable the data.
  - The greater the standard deviation value, the less reliable the data.

$$\text{Standard Deviation} = \sqrt{\frac{[(data_a - avg)^2 + (data_b - avg)^2 + \ldots.\ (data_x - avg)^2]}{(n-1)}}$$

STANDARD DEVIATION

- **To calculate standard deviation, we have to determine the average difference between each number and the mean of the data and divide by the amount of data minus 1.**
  - *Don't worry about knowing the specific formula; just focus on knowing what standard deviation values indicate*.

- **For example, if we had radish height measurements of 7, 10, and 13 cm, our average height would be 10 cm.**
  - $SD = \sqrt{\frac{[(7-10)^2 + (10-10)^2 + (13-10)^2]}{(3-1)}} = 3$
  - Our standard deviation in this case is 3 cm.
  - This indicates that on average each measured height varies by 3 cm from the mean.

STANDARD ERROR

- **<u>Standard error</u> is another statistical equation that can be used to calculate the reliability of data from an experiment.**
  - Standard error is very similar to standard deviation.
  - While standard deviation uses the "average variance" to calculate the reliability of data, standard error uses both variability <u>and</u> the size of the sample to calculate reliability.

- **Standard error is calculated by dividing the standard deviation value by the square root of the amount of data that you have.**
  - For example, if your mean was 10, if your standard deviation was 3, and your sample size was 9, your standard error would be $3/\sqrt{9} = 3/3 = 1$.
  - In this case, our standard error value is 1.

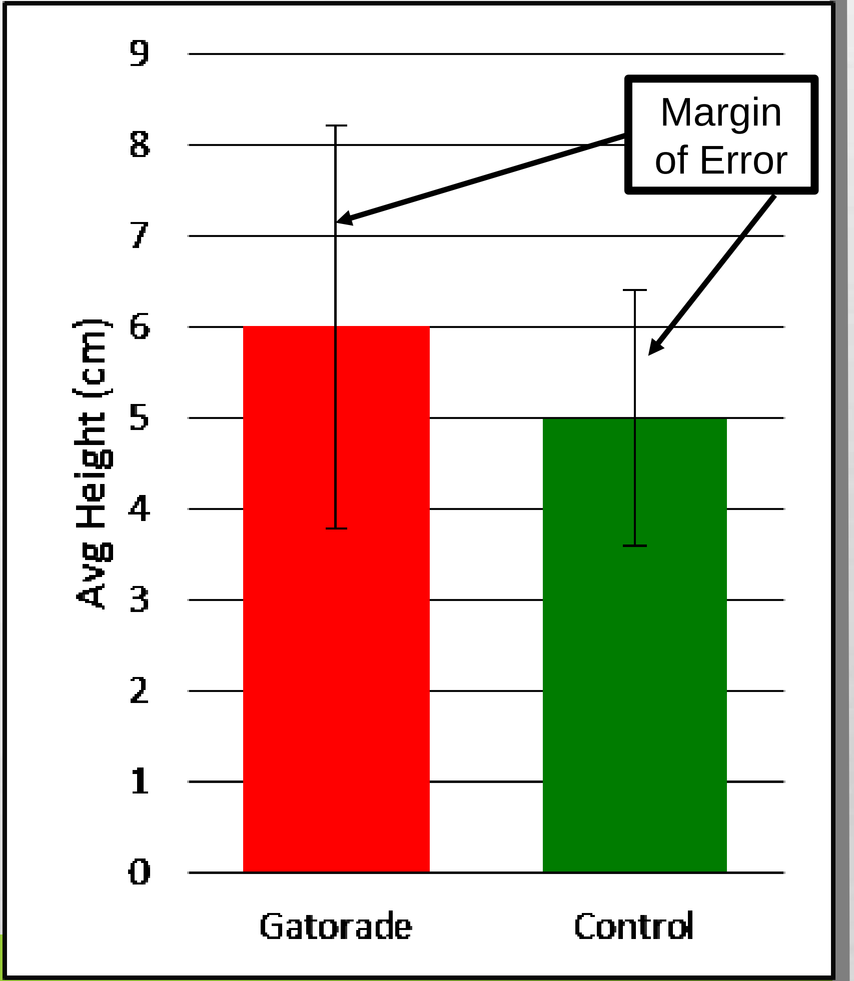
STANDARD ERROR

- **Because standard error incorporates both the sample size and the variability of the data, it is more valuable than standard deviation for calculating reliability.**
  - The lower the standard error value, the more reliable the data.
  - The greater the standard error value, the less reliable the data.

- **Standard error can be used to calculate the <u>margin of error</u>.**
  - <u>Margin of error</u> shows the range in which our mean (average) would be found any time we repeated an experiment under the same conditions.
  - The smaller the margin of error, the more reliable the results.

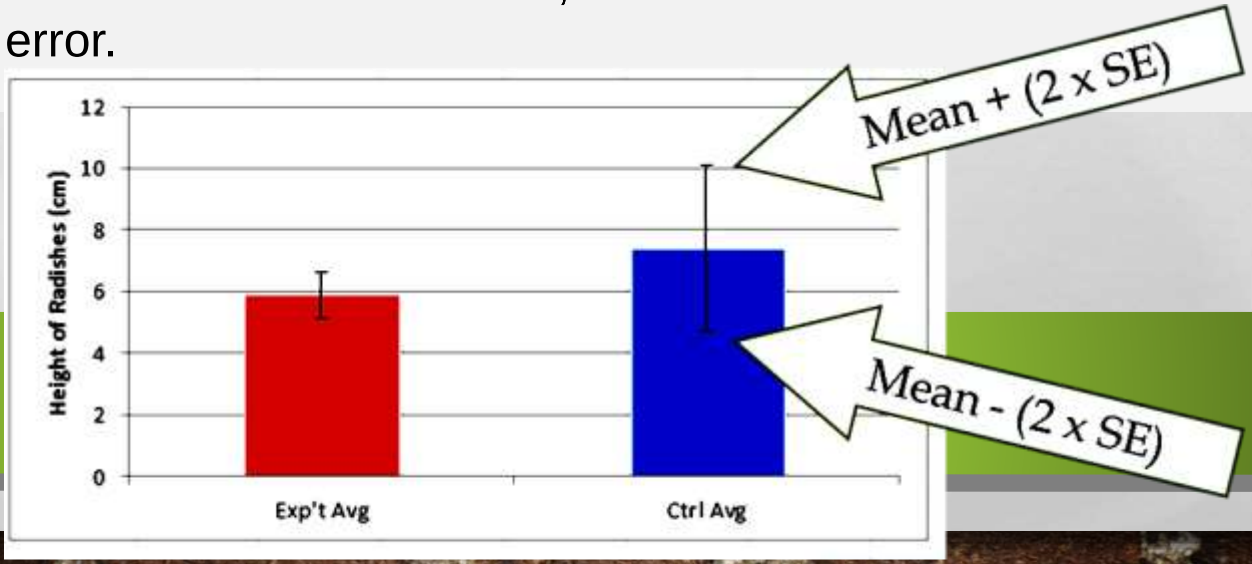
MARGIN OF ERROR

- **The margin of error is the range in which we would find the new results if the experiment was repeated under the same conditions.**
  - For example, if we repeated the Gatorade-radish experiment under the same conditions, the margin of error tells us the range in which we would find the average height for the second experiment.
  - This helps a scientist to determine how similar and predictable their results will be each time an experiment is performed.



MARGIN OF ERROR CONTINUED…

- **Typically the margin of error is equal to: Mean +/- (2 x Standard Error).**
  - For example, in our previous example, the mean was 10 cm and the standard error value is 1 cm.
  - This would mean that our margin of error would be 12 to 8 cm.
    - *[10 +/- (2 x 1)] = 12 or 8*
- **If margin of error is calculated by doubling the standard error value, it provides over a 95% confidence interval.**
  - This means that there is over a 95% chance if an experiment were repeated under the same conditions, the new results would be inside the margin of error.

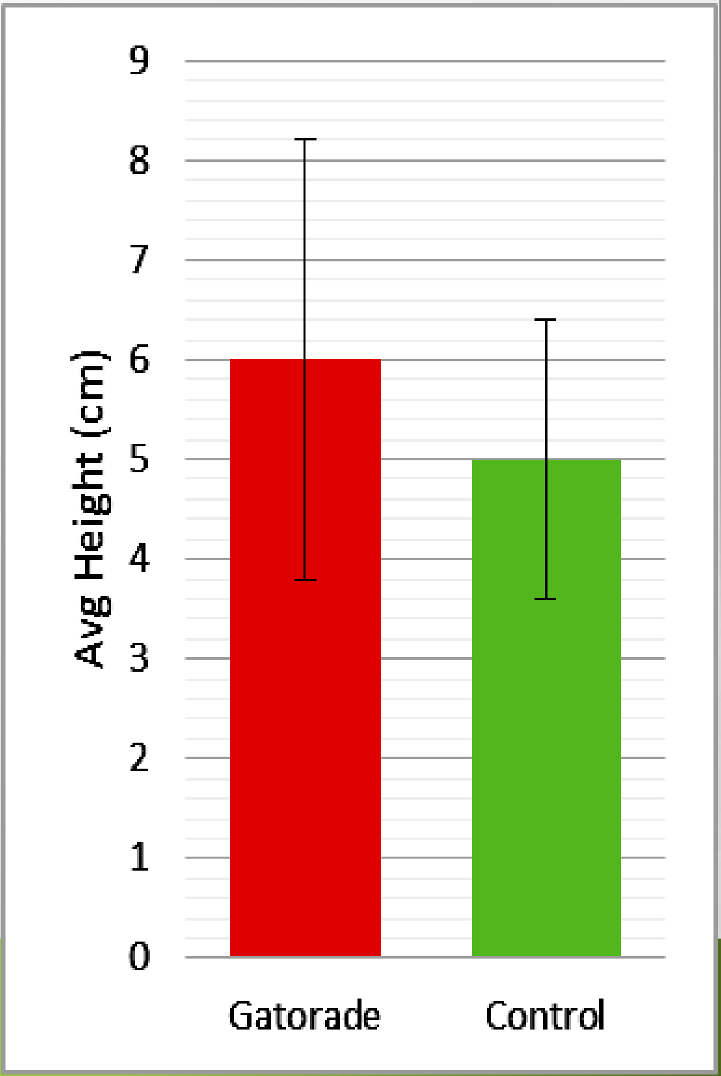


ERROR BARS

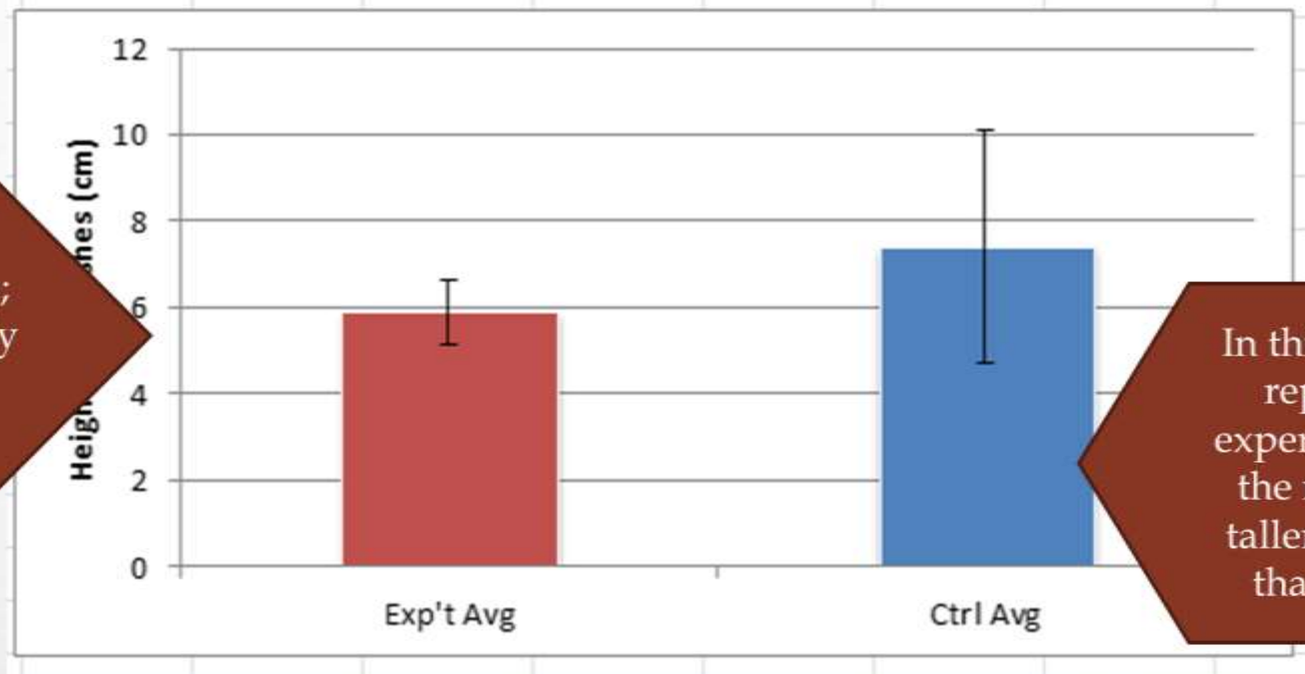
- **Standard Error and the Margin of Error are used to create Error Bars.**
  - **Error bars** are a visual depiction of your Margin of Error.
- **Both margin of error and error bars enable us to visually determine if two averages have a statistically significant difference.**
  - In other words, the error bars show whether or not two groups are different enough to always be different (if we repeated the same experiment under the same conditions).
- **If the error bars and margin of error do not overlap, then the average of one group will always be greater than the average of another group.**
  - Error bars do not overlap if the greatest error bar value of one group is less than the lowest error bar value of another group.
  - Visually, this would mean that the top of one error bar is beneath the bottom of another error bar.

ERROR BARS & MARGIN OF ERROR

- **If the error bars and margin of error DO overlap, this means that the two groups are too similar to say that there is any real difference.**
  - This would occur when the top of one error bar is above the bottom of another error bar.
  - This means that if an experiment were repeated, it would be impossible to predict which group would have the greatest average value in the second experimental trial.
- **For example, if radishes treated with Gatorade had an average height whose margin of error/error bars overlapped with that of the control, this would mean that…**
  - There is no significant difference between average height of the Gatorade-treated radishes and the average height of the control radishes.
  - If we were to repeat the experiment under the same conditions, we would be unable to predict whether the Gatorade-radishes or the control radishes would be taller on average.

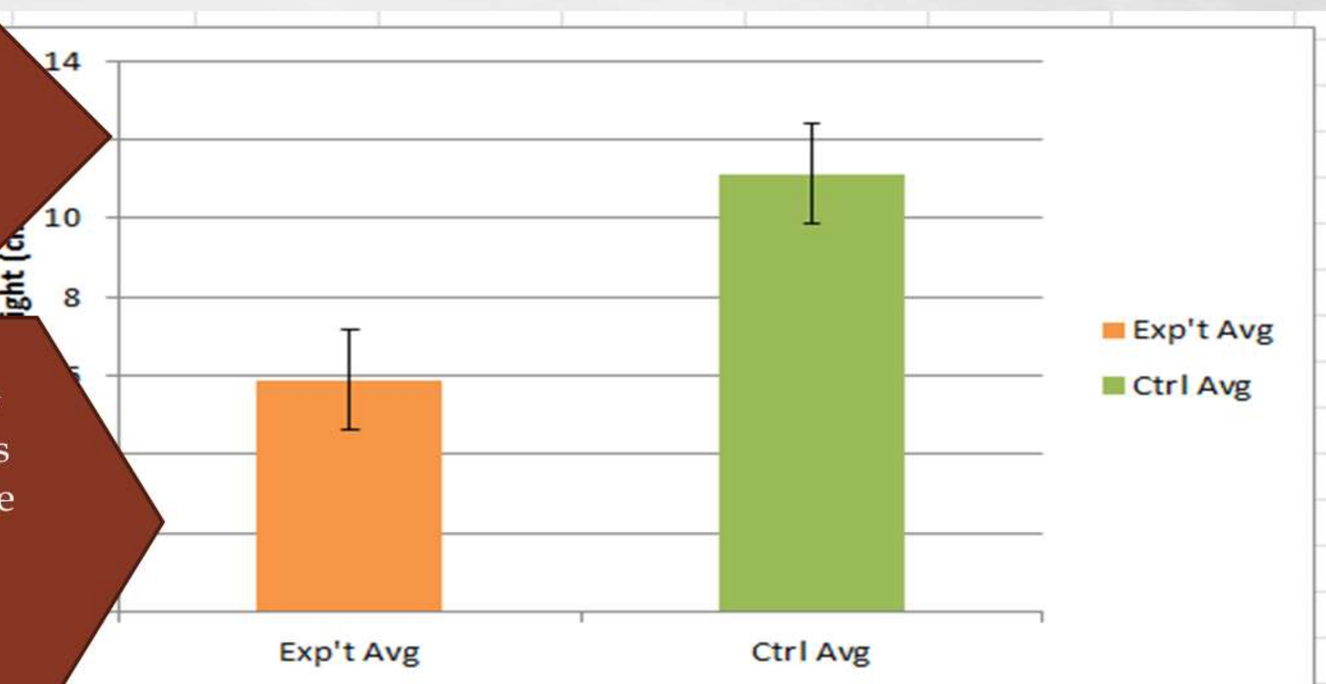


Error Bars overlap; they are statistically the same.



In this case, if we repeated the experiment again, the red *could* be taller on average than the blue.

Error Bars do not overlap; they are statistically different.



No matter how many times we repeated this experiment, the orange would always be shorter on average than the green.

WHAT TO TEST?

- **<u>Null Hypothesis</u> is the statement that you want to test.**
  - It is the hypothesis that the researcher tries to disprove, reject, or nullify.
  - The null hypothesis may represent an existing hypothesis that has not been proven correct or incorrect.
- **For example, if we were testing whether or not radishes that were treated with Gatorade would grow taller, our null hypothesis would be:**
  - "There is no difference between the height of radishes given Gatorade and the height of the control radishes."
- **The <u>alternative hypothesis</u> is what we are attempting to prove or demonstrate.**
  - Alternative hypothesis is what you accept if the null hypothesis is rejected.
- **For example, our alternative hypothesis in the earlier situation would be:**
  - "Radishes treated with Gatorade will grow taller than the control radishes."

SIGNIFICANCE OF YOUR RESULTS

- **<u>P-value:</u> this measurement enables you to determine whether or not you can accept or reject the null hypothesis.**
  - P-value is a number between 0 and 1 and is interpreted in the following way:
- **Small p-value (lower than 0.05) = STRONG evidence against the null hypothesis.**
  - A p-value under 0.05 means you would reject the null hypothesis and accept the alternative hypothesis.
- **Large p-value (over 0.05) indicates that you have WEAK evidence against the null hypothesis.**
  - If p-value greater than 0.05, you cannot reject the null hypothesis.
- **A p-value above 0.05 does not indicate anything about your alternative hypothesis.**
  - It only means that we cannot prove that the null hypothesis is incorrect.

SUMMARY OF CONCEPTS TERMS:

- **Most Important – Reliability of data depends on two factors:**
  - 1) **How much data we have**. More data is more reliable.
  - 2) **How similar the measurements are to each other**. More similar data is more reliable.
- **Mean: the average of the data.**
  - The mean is found by adding up all the measurements and dividing by the number of measurements that you have.
  - Whether or not a mean (average) is reliable depends on how much data was used to create the mean and how similar the measurements are to each other.
- **Variability: this is the measure of how much your measurements are different from each other.**
  - The more similar your measurements, the lower your variability.
  - The less similar your measurements, the greater your variability.

SUMMARY OF CONCEPTS TERMS:

- **<u>Standard Deviation</u>: this is the measurement of how much variability you have.**
  - Standard Deviation tells us how much each measurement differs from the mean on average. The <u>lower</u> the standard deviation, the <u>lower</u> the variance, the <u>more reliable</u> the results.
- **<u>Standard Error</u>: this is a measurement of the reliability of data that includes both variability and the amount of data in an experiment.**
  - Standard Error is a more valuable measurement of the reliability of data because it includes both size and variance. The <u>lower</u> the SE, the more reliable the results.
- **<u>Margin of Error</u>: the range in which we would almost always find the mean if we repeated the experiment under the same conditions.**
  - Margin of Error is equal to: Mean +/- (2 x Standard Error).
  - This gives us over a 95% likelihood of finding an accurate margin of error.
- **<u>Error Bars</u>: a visual depiction of margin of error on a graph.**
  - Error bars are drawn on a graph based on the margin of error.

SUMMARY OF CONCEPTS TERMS:

- **<u>Statistically significant difference</u>: this means that two means (averages) are different enough that they will always be different.**
  - Two groups are significantly different if their error bars do not overlap.
  - This means that the highest value in the margin of error of one group's average does not overlap with the lowest value of the other group average's margin of error.

- **<u>Null Hypothesis</u>: this is the hypothesis you are trying to disprove.**
- **<u>Alternative Hypothesis</u>: this is the hypothesis that you are trying to prove is correct by disproving the null hypothesis.**

- **<u>P-value</u>: this tells us whether or not we can reject the null hypothesis and accept the alternative hypothesis.**
  - If the p-value is below 0.05, we can reject the null hypothesis and accept the alternative hypothesis.
  - If the p-value is above 0.05, we cannot reject the null hypothesis.