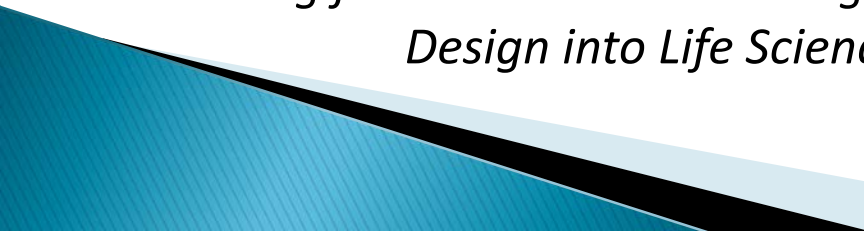# Is it significant?

Making Experimental Data Meaningful

In this talk we will be collecting data, analyzing data using statistics, and thinking about how to represent data in meaningful ways.

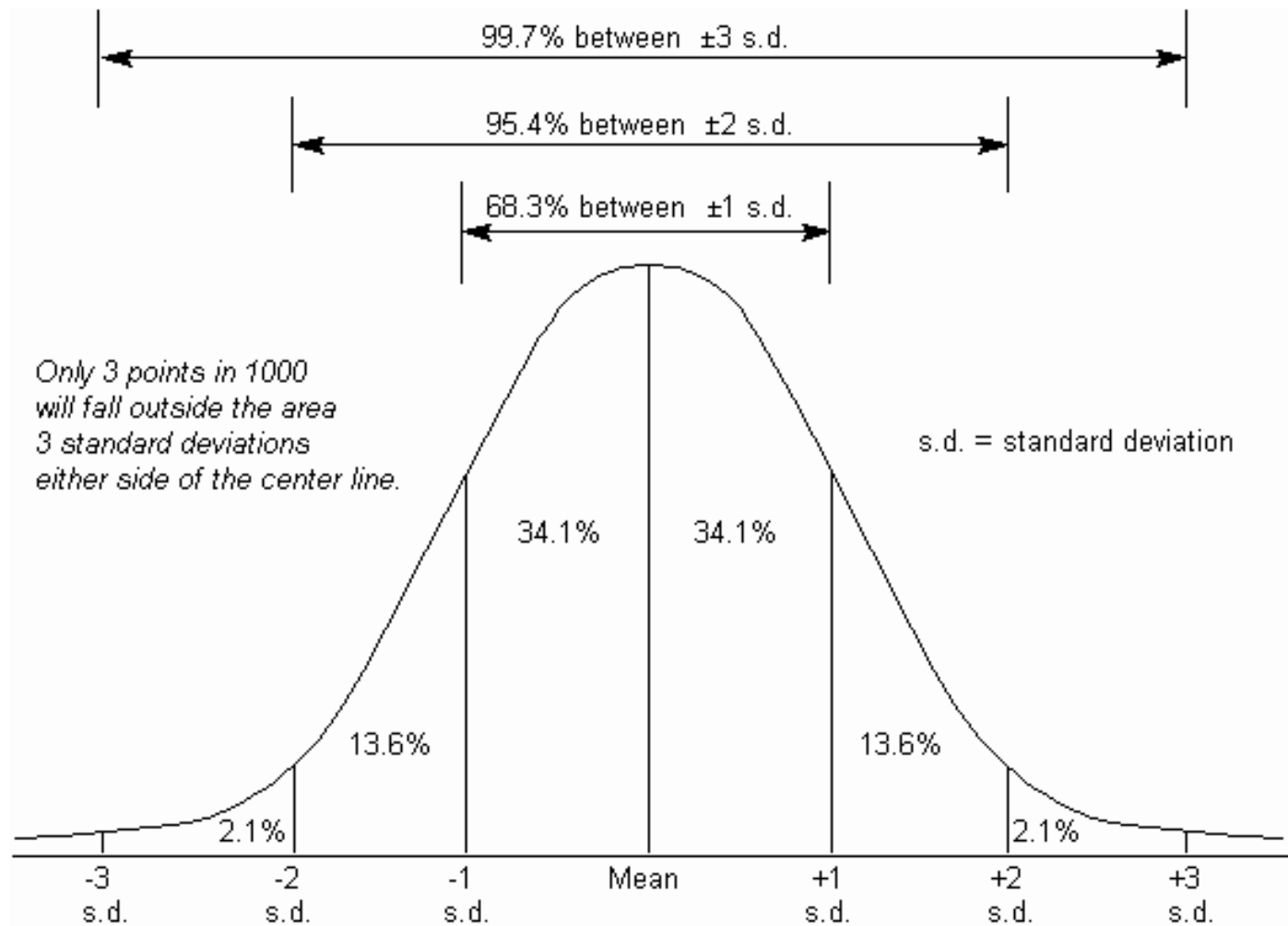**Presenter: Kathy Daniels (Mississinewa HS, Gas City, IN)**

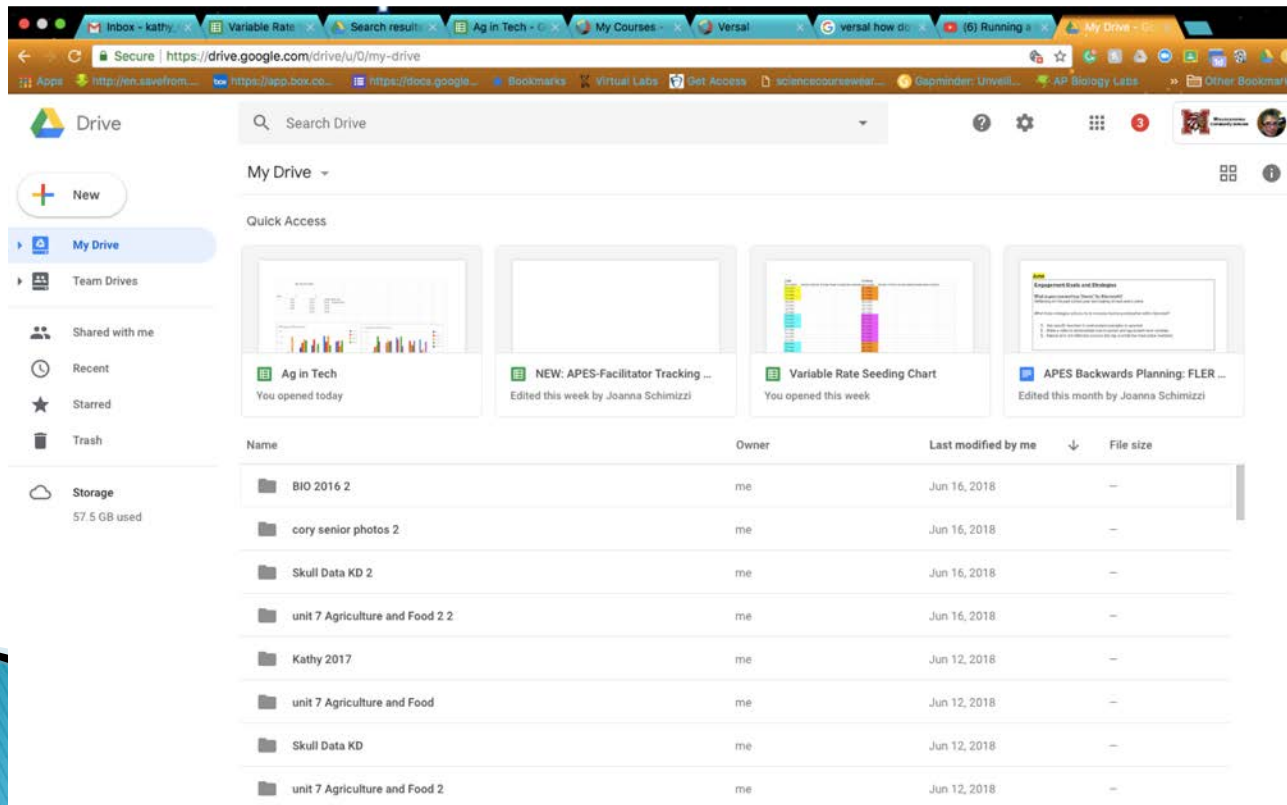# How do students (& teachers) think about probability?

- A coin...

  ◦ What do we know?

  ◦ If I flip the coin...what can I expect?

  ◦ When do we start to feel suspicion if we don't get what we expect?
  ◦ ½ = 0.50 = 50%
  ◦ ½ x ½ x ½ = 0.125 = 12.5%
  ◦ ½ x ½ x ½ x ½ = 0.0625 = 6%
  ◦ ½ X ½ X ½ X ½ X ½ = 0.03125 = 3%

99.7% between ±3 s.d.

95.4% between ±2 s.d.

68.3% between ±1 s.d.

Only 3 points in 1000 will fall outside the area 3 standard deviations either side of the center line.

s.d. = standard deviation

34.1%   34.1%

13.6%   13.6%

2.1%   2.1%

| -3 s.d. | -2 s.d. | -1 s.d. | Mean | +1 s.d. | +2 s.d. | +3 s.d. |

syque.com

Google Sheets – A quick how-to

IF you have a Google account

▸ Open your browser (Chrome is preferred)
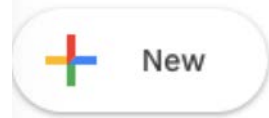
▸ Go to the address bar and type Google drive

To make a new spreadsheet

- Click on

- Then select Google Sheets
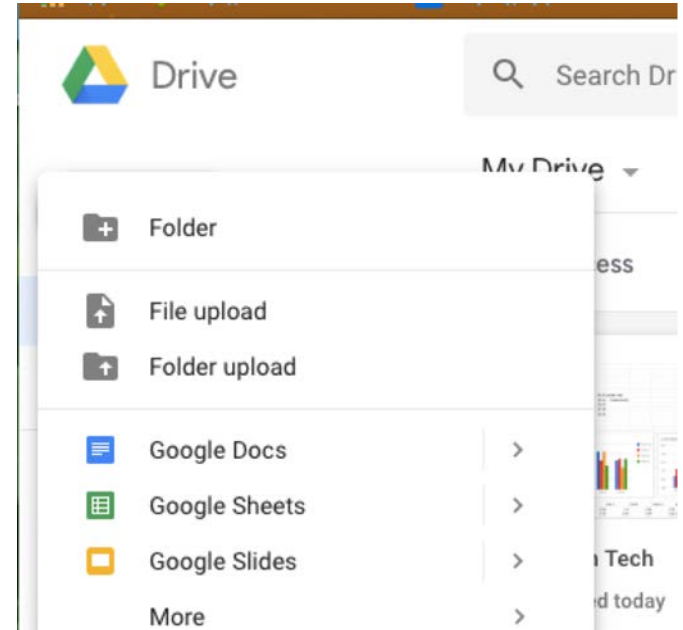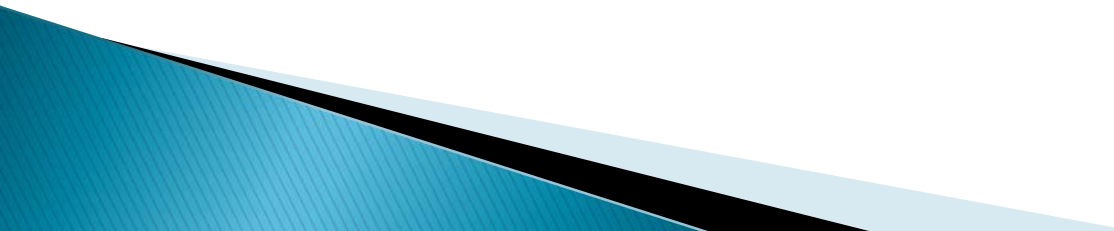
Before we add data…

Give your new spreadsheet a name and share it with me at kathy_daniels@olemiss.k12.in.us
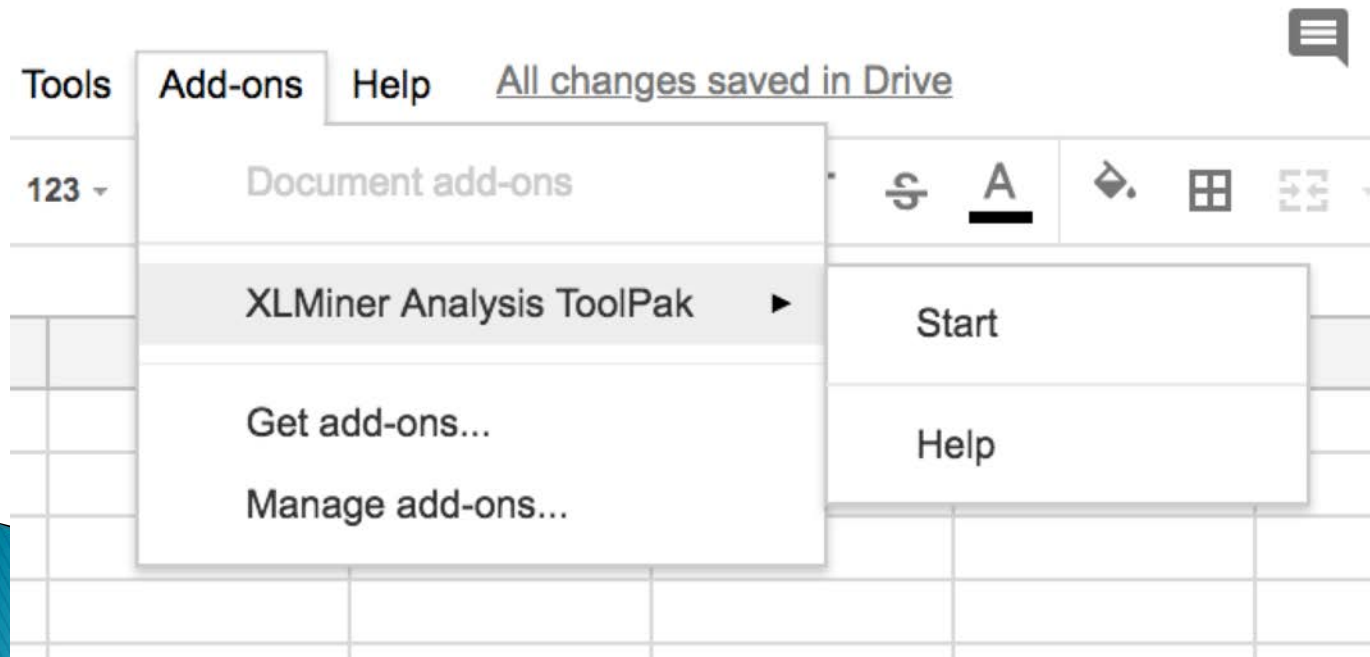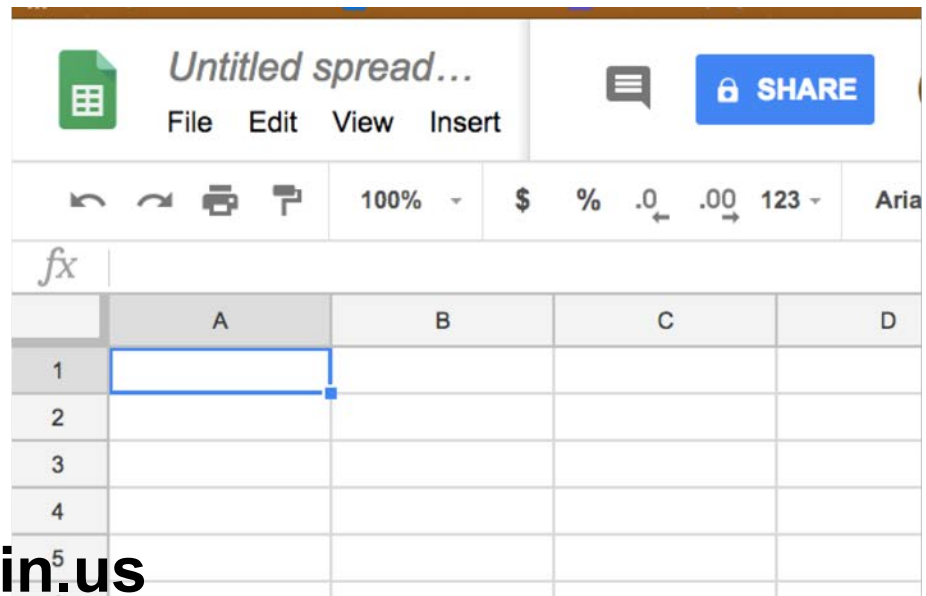
The dog kennel

- Imagine you have agreed to watch your friend's dog kennel business for the weekend while they take a short vacation.
- Your kennel has 9 dogs this weekend.
- Your job is to feed, exercise, and clean up after the dogs for three days – and to keep them healthy for their owners.
- You will keep a spreadsheet of their weight to show your friend how responsible you are!

Make a spreadsheet

	A	Weight on Arrival	Weight on Departure	Weight Change	% change
1		Weight on Arrival	Weight on Departure	Weight Change	% change
2	Coco	5	2	-3	-60.00
3	Tootsie	12	10	-2	-16.67
4	Buttons	16	8	-8	-50.00
5	Lacy	10	12	2	20.00
6	Sammy	3	4	1	33.33
7	Solo	7	10	3	42.86
8	Waylow	9	8	-1	-11.11
9	Red	25	22	-3	-12.00

What kinds of graphs might be helpful in determining the health of the dogs at the end of the weekend?

Types of Descriptive Stats students can use

- Graphs
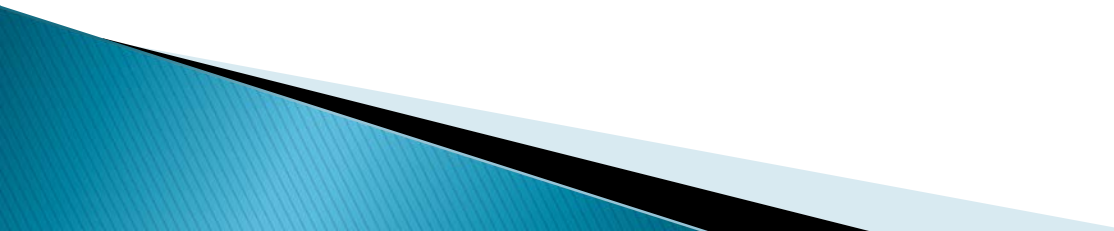 - Line, bar, pie, histograms
  - Linear Regression
- T Tests
  - Paired, Unpaired
- ANOVA
- Chi Square

How do I pick what analysis tools to use?
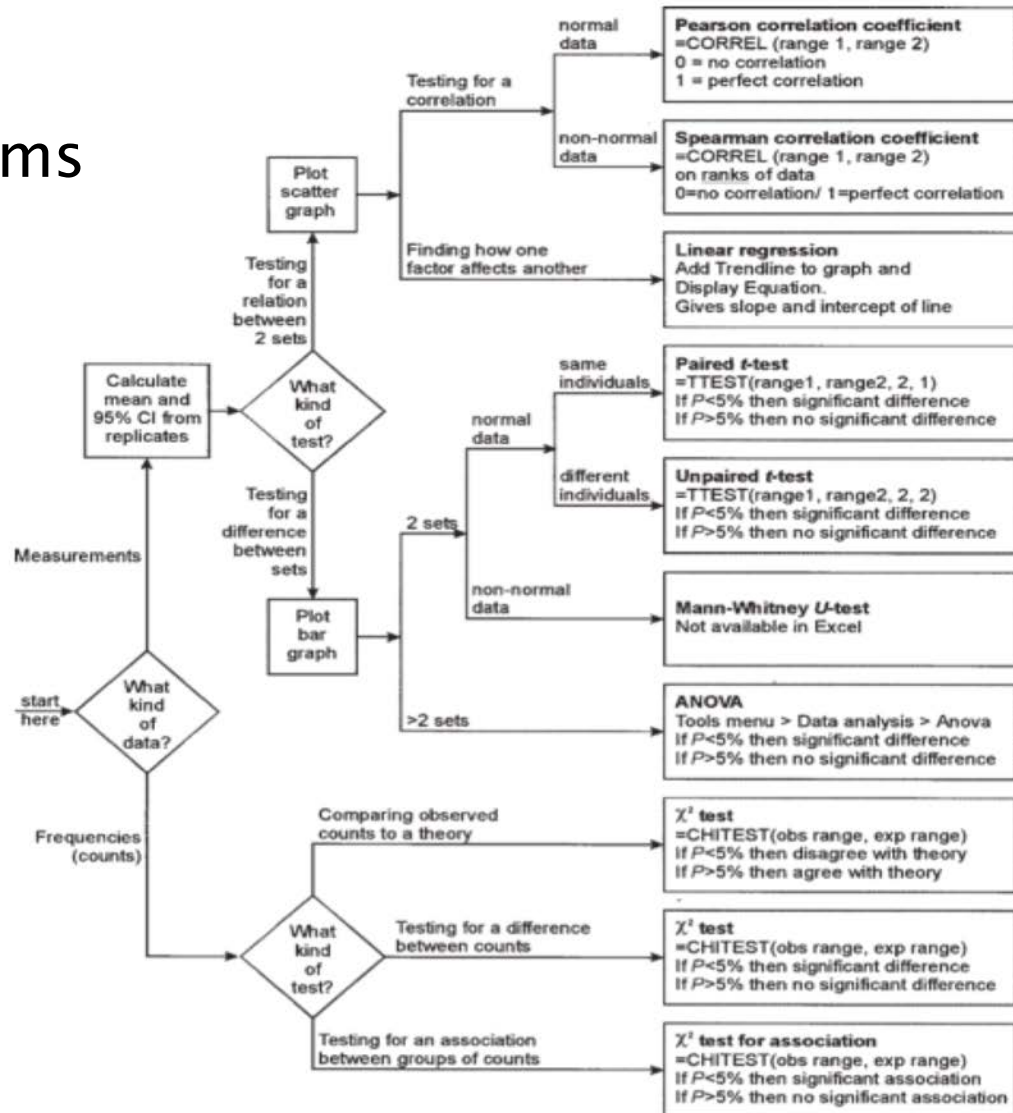
What can farmers do to increase crop yields?
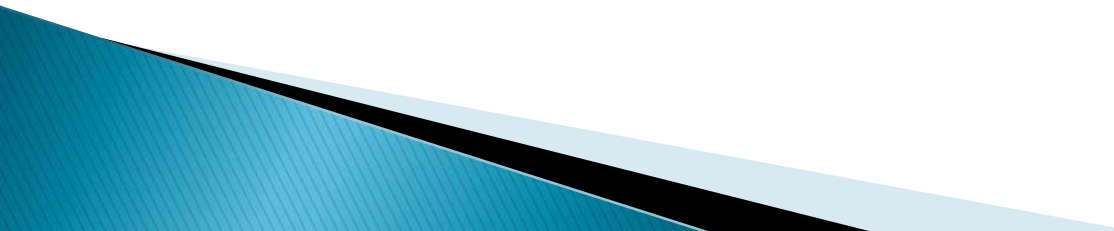
Think about costs, benefits, and possible issues

- Fertilizer

- Herbicides/Pesticides

- Seed Choice

- Field Conditions (Soil Testing / Drainage)

How do farmers make their choices?

What defines "good enough"?

Who makes the decisions for the farm?

How can we help our students think about these factors objectively?

What is a T-test?

- A test designed to determine the significance of the data collected during an experiment.
  - Can compare the mean of one population to a fixed constant (One-Sample t-test)
  - Can compare the mean of 2 independent populations to each other (Two-Sample t-test)
  - Can compare 2 linked measurements, such as before and after of the same individuals (Paired t-test)

Two-Tailed Versus One-Tailed Hyphothesis Tests
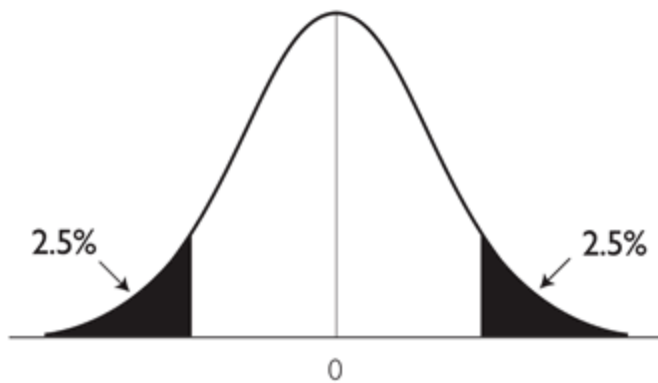
Figure A:
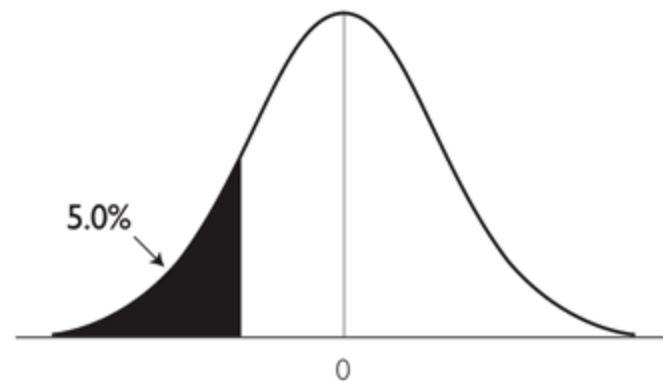Two-Tailed Test

Figure B:
One-Tailed Test
(Left-Tailed Test)

2.5%   2.5%

0

5.0%

0

Source: The Heritage Foundation.

To DO a t-test on Google Sheets

- Open the XLMiner ToolPak
- Choose t-test Paired Two Sample for Means

Calculate Average

Fill in data ranges and select output location
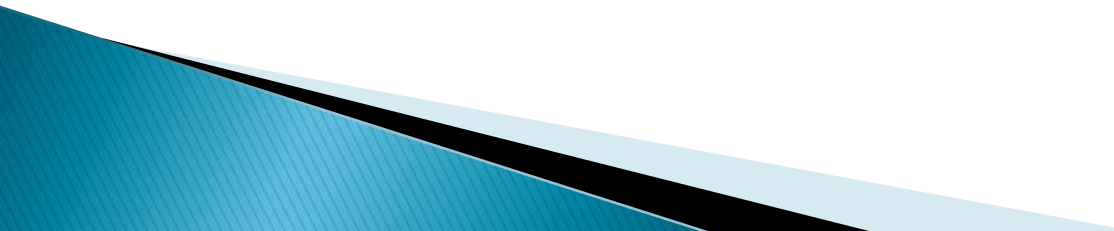
Block	Variety A	Variety B
1	47	45
2	46	50
3	51	60
4	48	57

t-Test: Paired Two Sample for Means

	Variety A	Variety B
Mean	48	53
Variance	4.666666667	46
Observations	4	4
Pearson Correlation	0.7962775716	
Hypothesized Mean Differe	0	
df	3	
t Stat	-1.912730139	
P(T<=t) one-tail	0.0758589777	
t Critical one-tail	2.35336342	
P(T<=t) two-tail	0.1517179554	
t Critical two-tail	3.182446305	

What does the p value tell us?

- If the "p is low, reject the H_0"
- If the p value is generally less than 0.05 then we can be relatively certain the difference seen is significant, our alternate hypothesis, H_a, is supported
  - 5% chance the difference is due to random chance
  - 95% chance the difference is due to treatment or variable being tested
- Generally H_0 is that there is *no effect* and H_a is that there is some effect

ANOVA

- Analysis of Variance and allows you to see if two variables in an experiment influenced one another in the experiment.
  - One-way ANOVA compares data from single category but with different levels of the variable
  - Two-way ANOVA compares data from 2 different categorical levels

To do an ANOVA

▸ Follow same steps as in the t-test except select Anova: Two-Factor without replication

▸ Then for Input Range select entire table

▸ Check Labels in First Row box

▸ Select location for output analysis

What do these p values mean?

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
1	3	155	51.66666667	58.33333333
2	3	164	54.66666667	74.33333333
3	3	160	53.33333333	26.33333333
4	3	159	53	21
Treatment A	4	192	48	8.666666667
Treatment B	4	206	51.5	5.666666667
Treatment C	4	240	60	8.666666667

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	13.66666667	3	4.555555556	0.4939759036	0.6995242793	4.757062664
Columns	304.6666667	2	152.3333333	16.51807229	0.003631223645	5.14325285
Error	55.33333333	6	9.222222222			
Total	373.6666667	11				

LSD

▶ **Least Significant Difference**
  ◦ How is it calculated?

  ◦ What does it mean?

So tonight & tomorrow you need to think about Organizing, Graphing, Testing & Interpreting Data

1. Think and talk about how to organize the data _before_ making your graphs or performing your statistical analysis

2. Compare the different visual representations (plots) of the data – what graph really gives you a helpful and accurate "picture" of the data?

3. Select one plot that your group feels is the best representation of the data and be prepared to justify your choice

1. Select one statistical test (t-test or ANOVA) that you could do on your data set

Versal Self-guided course in Statistics

https://versal.com/c/ybyfr3/summary

To contact me:

Kathy_daniels@olemiss.k12.in.us