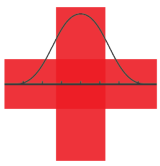
AP Statistics

Summer 2018

Lindsey Gallas
East Kentwood High School
Kentwood, Michigan
lindsey@statsmedic.com



STATS MEDIC

Agenda:

1. Introductions

2. General Course Information and Resources

3. Content

    - Unit 1: Exploring Data
    - Unit 2: Modeling Distributions of Data
    - Unit 3: Describing Relationships
    - Unit 4: Designing Studies and Experiments
    - Unit 5: Probability
    - Unit 6: Random Variables
    - Unit 7: Sampling Distributions
    - Unit 8: One Sample Confidence Intervals
    - Unit 9: One Sample Significance Tests
    - Unit 10: Two Sample Inference
    - Unit 11: Chi-square Tests
    - Unit 12: More About Regression

4. Questions/Work Time

<u>Goals for the week:</u>

- Participants will be exposed to relevant resources and instructional strategies that can enhance the quality of their AP Statistics course.
- Participants will actively participate in activities that develop deeper understanding of statistical concepts.
- Participants will learn about the format, content, rubric, and grading of the AP Statistics Exam.
- Participants will have a better understanding of statistical inference.

Day	Content	Classroom Activities
1	General Course Information and Resources	
	Ch. 1: Exploring Data	Smelling Parkinson's Disease
	Ch. 3: Describing Relationships	How Many Rubber Bands Does Barbie Need?
		How Good are the Predictions for Barbie
		Barbie Bungee Finale: Drop only
	Ch. 4: Studies and Experiments	What is the Average Word Length of a Beyonce Song?
	Ch. 6: Random Variables	Is it Smart to Foul at the End of the Game?
	Ch. 7: Sampling Distributions	What's the Proportion of Orange Reese's Pieces?
	Ch. 8: One Sample Confidence Intervals	How Many States Can you Name?
	Ch. 9: One Sample Significance Tests	Is Mrs. Gallas a Good Free Throw Shooter?
	Ch. 10: Two Sample Inference	Is Yawning Contagious?

SMELLING PARKINSON'S DISEASE

INTRODUCTION

As reported by the Washington Post, Joy Milne of Perth, UK, smelled a "subtle musky odor" on her husband Les that she had never smelled before. At first, Joy thought maybe it was just from the sweat after long hours of work. But when Les was diagnosed with Parkinson's 6 years later, Joy suspected the odor might be a result of the disease.

Scientists were intrigued by Joy's claim and designed an experiment to test her ability to "smell Parkinson's." Joy was presented with 12 different shirts, each worn by a different person, some of whom had Parkinson's and some of whom did not. The shirts were given to Joy in a random order and she had to decide whether each shirt was worn by a Parkinson's patient or not.

1. Why would it be important to know that someone can smell Parkinson's disease?

2. How many correct decisions (out of 12) would you expect Joy make if she couldn't really smell Parkinson's and was just guessing?

3. How many correct decisions (out of 12) would it take to *convince* you that Joy really could smell Parkinson's?

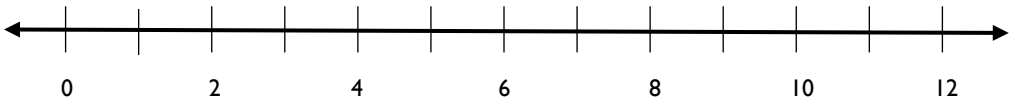
Simulating the Experiment

Although the researchers wanted to believe Joy, there was a chance that she may not really be able to tell Parkinson's by smell. It's logical to be skeptical of claims that are very different than our experiences. If Joy couldn't really distinguish Parkinson's by smell, then she would just have been guessing which shirt was which. The researchers were not willing to commit time and resources to a larger investigation unless they could be convinced to that Joy's wasn't just guessing. To investigate the idea that Joy was just guessing which shirt was worn by which type of person, we will begin by assuming that Joy was just guessing.

4. Mrs. Gallas will hand you 12 cards (shirts) that have been shuffled into a random order. Don't turn them over yet! On the back of some of them is "Parkinson's" and on the back of others is "No Parkinson's." For each card, guess Parkinson's or No Parkinson's. Once you have made your guess, turn the card over and see if you were correct. Repeat this for each card and record the number of correct identifications (out of 12) below.

Tally of correct identifications	Number of correct identifications	Proportion of correct identifications

5. Create a dotplot of the number of correct identifications with the rest of the class. Record the results below.



6. In the actual experiment, Joy identified 11 of the 12 shirts correctly. Based on the very small-scale simulation by you and your classmates, what proportion of the simulations resulted in 11 or more shirts correctly identified, assuming that the person was guessing?

Name: _____ Hour: _____

Lesson 3.1: Day 1: How many rubber bands does Barbie need?

How many rubber bands should we attach to Barbie so that she has the absolute most fun without smashing her head if she were to jump from the balcony in the front foyer of the school (5.3 meters above the ground)? Here's the catch: You many only use 7 rubberbands to figure this out.

Complete the table:

# Rubber bands	0	1	2	3	4	5	6	7
Lowest point of fall								

Use your group's data to complete the following:

1. Identify the explanatory and response variables.

2. How many variables do we have? Are they categorical or quantitative?

3. Use the applet at www.stapplet.com to make a scatterplot. Draw below.

4. Describe the relationship displayed in the scatterplot.

Name: _____ Hour: _____

Lesson 3.1 – Displaying Relationships: Scatterplots

Big Ideas:

Check Your Understanding:

1. Is there a relationship between the amount of sugar (in grams) and the number of calories in movie-theater candy? Here are the data from a sample of 12 types of candy.

Name	Sugar (g)	Calories	Name	Sugar (g)	Calories
Butterfinger Minis	45	450	Reese's Pieces	61	580
Junior Mints	107	570	Skittles	87	450
M&M'S®	62	480	Sour Patch Kids	92	490
Milk Duds	44	370	SweeTarts	136	680
Peanut M&M'S®	79	790	Twizzlers	59	460
Raisinets	60	420	Whoppers	48	350

a. Identify the explanatory and response variables. Explain your reasoning.

b. Make a scatterplot to display the relationship between amount of sugar and the number of calories in movie-theater candy.

c. Describe the relationship shown in the scatterplot.

Name: _____ Hour: _____ Date: _____

Lesson 3.2: Day 1: How good are the predictions for Barbie?

Here is the data from one of the groups. The group forgot to record their measurement for 5 rubber bands.

Number of rubber bands	0	1	2	3	4	5	6	7
Distance traveled (cm)	25	32	41	49	55	?	69	78

1. Go to stapplet.com to make a scatterplot. Then click "Calculate least-squares regression line". This is the line that best models the data. Write the equation below.

2. Use the regression line to predict the distance Barbie travels for 5 rubber bands. Show work.

3. One of the group members later found the measurement for 5 rubber bands was 64 cm. Was the prediction from #2 too high or too low? How far off?

4. Predict the distance that Barbie would travel if the group used 20 rubber bands. Would you trust this prediction more or less than the prediction you made in #2?

5. What is the y-intercept of the equation of the regression line? What does it mean?

6. What is the slope of the equation of the regression line? What does it mean?

Name: _____ Hour: _____ Date: _____

Lesson 3.2 – Prediction, Residuals, Interpreting a Regression Line

Big Ideas:

Check Your Understanding:

1. Some data were collected on the weight of a male white laboratory rat for the first 25 weeks after its birth. A scatterplot of *y* = weight (in grams) and *x* = time since birth (in weeks) shows a fairly strong, positive linear relationship. The regression equation $\hat{y} = 100 + 40x$ models the data fairly well.

a. Interpret the slope of the regression line.

b. Does the value of the y intercept have meaning in this context? If so, interpret the y intercept. If not, explain why.

c. Predict the rat's weight at 16 weeks old.

d. Calculate and interpret the residual if the rat weighed 700 grams at 16 weeks old

e. Should you use this line to predict the rat's weight at 2 years old? Use the equation to make the prediction and discuss your confidence in the result. (There are 454 grams in a pound.)

Name: _____ Hour: _____ Date: _____

Barbie Bungee – The Finale

It's finally time to jump Barbie! At the end of the hour we will be dropping Barbie from the staircase in the foyer which is 17 ft. (5.2 m). Before we drop her, we will use everything we've learned this chapter to calculate the best possible length of bungee cord.

Write in your group's data in the table below.

Number of rubber bands	0	1	2	3	4	5	6	7
Lowest point head reaches (cm)								

1. Identify which variable is the explanatory variable and which is the response variable?

2. Use the Applet to create a scatterplot.

3. Describe your distribution (DUFS).

4. Estimate the *r* value of your distribution.

5. What would happen to the correlation (r) if you graphed the scatterplot with the lowest point on the horizontal axis and # rubber bands on the vertical axis?

Name: _____ Hour: _____ Date: _____

6. Calculate the correlation using SPA applets. Write it below. What is the unit of the correlation?

7. Use the Applet to find the least squares regression line for your data. Write the equation below.

8. What is the slope of your LSRL? Interpret the slope.

9. What is the *y*-intercept of your line? Interpret.

10. Use the LSRL to calculate and interpret the residual for 4 rubber bands.

11. Sketch the residual plot for your LSRL.

12. Find the r^2 value and interpret it.

13. Find the standard deviation of the residuals and interpret it.

14. Is the linear regression an appropriate model? Explain.

15. Use your model to predict the number of rubber bands Barbie will need in order to have the most exciting yet safe bungee jump from 17 ft. (518 cm)

Name: _____ Hour: _____ Date: _____

Lesson 4.1: What's the average word length of a Beyoncé song?

1. Quickly circle a random sample of 5 words. Write them below. How many letters in each word?

2. What is the average word length of your sample? _____.

3. Put your average on the dotplot on the white board at the front of the room. Copy the class dotplot below.

4. Find a new sample of 5 words using a random number generator. Put your average on the dotplot on the white board at the front of the room. Copy the class dotplot below.

5. How is the dotplot from #4 different than the dotplot for #3? Which do you think is a better estimator of the true mean word length?

6. What do you think the true mean word length is for "Crazy in Love"?

7. It is known that Beyonce wrote the lyrics for all of the Destiny's child songs. The average word length for these songs is 3.64 letters. Based on your samples, do you have good evidence that Beyonce did not write the lyrics for "Crazy in Love". Explain.

Name: _____ Hour: _____ Date: _____

Lesson 4.1 – Sampling Methods

Important ideas:

Check Your Understanding

1. In June 2008 *Parade* magazine posed the following question: "Should drivers be banned from using all cell phones?" Readers were encouraged to vote online at www.parade.com. The July 13, 2008, issue of *Parade* reported the results: 2407 (85%) said "Yes" and 410 (15%) said "No."

   a. What type of sample did the *Parade* survey obtain?

   b. Explain why this sampling method is biased.

   c. Is 85% likely to be greater than or less than the percentage of all adults who believe that cell-phone use while driving should be banned? Why?

2. To help eliminate bias, a reporter from *Parade* decides she will go out and ask people in person if they think drivers should be banned from using cell phones. She lives close to the local high school so she goes to the parking lot at 3:00 pm and asks the first 100 people she sees.

   a. What type of sample did the reporter obtain?

   b. Explain why this sampling method is biased.

3. How could *Parade* magazine avoid the bias described above?

Name: _____ Hour: _____ Date: _____

Lesson 6.3: Day 1: Is it smart to foul at the end of the game?

In the 2005 Conference USA basketball tournament, Memphis trailed Louisville by two points. At the buzzer, Memphis's Darius Washington attempted a 3-pointer; he missed but was fouled, and went to the line for three free throws. Each made free throw is worth 1 point. Was it smart to foul?

1. What are all the possible ways the shots could fall (e.g. make-miss-miss, etc.)?

2. Darius Washington was a 72% free-throw shooter. Find the probability that Memphis will win, lose or go to overtime. When you have found the probabilities put them in the table in #3.

Win	Lose	Overtime

3. Prior to watching each shot, calculate the probability that Memphis wins the game in regulation, loses the game in regulation, or sends the game into overtime.

		Shots Remain.	Probability Memphis Win	Probability Memphis Lose	Probability Overtime
75	73				
75					
75					

4. Washington is a 40% 3-point shooter. Do you think Louisville was smart to foul? Why or why not?

Name: _____ Hour: _____ Date: _____

Lesson 6.3 Day 1– Binomial Random Variables

Important ideas:

Check Your Understanding

1. For each of the following situations, determine whether or not the given random variable has a binomial distribution. Justify your answer.

a. Shuffle a deck of cards. Turn over the top card. Put the card back in the deck, and shuffle again. Repeat this process 10 times. Let X = the number of aces you observe.

b. Choose 5 students at random from your class. Let Y = the number who are over 6 feet tall.

2. Pedro drives the same route to work on Monday through Friday. His route includes one traffic light. According to the local traffic department, there is a 55% chance that the light will be red on a randomly selected work day. Suppose we choose 10 of Pedro's work days at random and let Y = the number of times that the light is red.

a. Explain why Y is a binomial random variable.

b. Find the probability that the light is red on exactly 7 days.

Name: _____ Hour: _____ Date: _____

Lesson 7.2: What's the proportion of orange Reese's Pieces?

If we take a sample of Reese's Pieces, what proportion of the candies will be orange?

Suppose a large bag of Reese's Pieces has 1000 pieces. The manufacturer says that exactly 40% of the candies are orange. If we select a sample of 50 pieces, how many will be orange? Let X = the number of orange candies in the sample.

1. What type of probability distribution does X have? Justify.

2. Draw a sample of 50 Reese's Pieces using the applet. How many pieces were orange? Repeat this 5 times. Write the values below.

3. Write the values on sticker dots and add it to the dotplot on the board. Sketch the dotplot below.

4. What does each dot represent?

5. What is the mean and the standard deviation for the distribution of X? Show work.

6. What is the approximate shape of the sampling distribution for X? Explain and sketch it below.

Name: _____ Hour: _____ Date: _____

Instead of finding the number of candies that are orange, we will now find the **proportion** of candies that are orange.

7. Use your samples from #2 and turn each number of orange candies into the **proportion of orange candies** in the sample ($\hat{p}$). Write the proportions below and add them to the second dotplot on the board.

8. Sketch the dotplot below.

9. What does each dot represent?

10. Find the new mean and standard deviation. Show work.

11. What is the approximate shape of the sampling distribution for $\hat{p}$? Explain and sketch it below.

12. We know that bags of Reese's Pieces contain exactly 40% that are orange. If we select a random sample of 50 candies, what is the probability that the sample proportion will be 50% or greater?

Name: _____ Hour: _____ Date: _____

Lesson 7.2 – The Sampling Distribution of $\hat{p}$

Important ideas:

Check Your Understanding

Suppose that 75% of young adult Internet users (ages 18 to 29) watch online videos. A polling organization contacts an SRS of 1000 young adult Internet users and calculates the proportion $\hat{p}$ in this sample who watch online videos.

1. Identify the mean of the sampling distribution of $\hat{p}$.

2. Calculate and interpret the standard deviation of the sampling distribution of $\hat{p}$. Check that the 10% condition is met.

3. Is the sampling distribution of $\hat{p}$ approximately Normal? Check that the Large Counts condition is met.

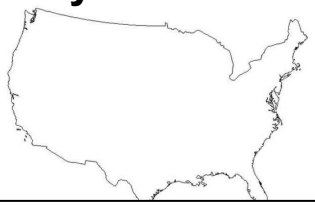
4. Find the probability that the random sample of 1000 young adults will give a result within 2 percentage points of the true value.

5. If the sample size were 9000 rather than 1000, how would this change the sampling distribution of $\hat{p}$?

Name: _____ Hour: _____ Date: _____

Lesson 8.3: Day 2: How many states can you name?





How many states can you name in one minute? We will use this class as a random sample of all AP Stats students to estimate a 95% confidence interval for the mean number of states an AP Stats student can name in one minute.

1. When the timer starts, list as many states as you can on a piece of paper. Write the number of states you listed on the board.

2. What type of data is this? Categorical or quantitative?

2. Enter the class data at stapplet.com. Find the sample mean and standard deviation. Sketch the dotplot of the sample data.

$n =$ $\bar{x} =$ $s_x =$

3. Construct a 95% confidence interval to estimate the mean # of states a senior can name.

STATE: State the parameter you want to estimate and the confidence level.

Parameter: _____ Confidence level: _____

PLAN: Identify the appropriate inference method and check conditions.

Name of procedure:

Check conditions:

DO: If the conditions are met, perform the calculations.

General Formula for any confidence interval:

Specific Formula for this confidence interval:

Plug numbers into the formula:

Answer:

CONCLUDE: Interpret your interval in the context of the problem.

Interpret:

Name: _____ Hour: _____ Date: _____

Lesson 8.3 Day 2 – The Four Step Process

Important ideas:

Check Your Understanding

1. Administrators at your school want to estimate how much time students spend on homework, on average, during a typical week. They want to estimate μ at the 90% confidence level with a margin of error of at most 30 minutes. A pilot study indicated that the standard deviation of time spent on homework per week is about 154 minutes. How many students need to be surveyed to meet the administrators' goal?

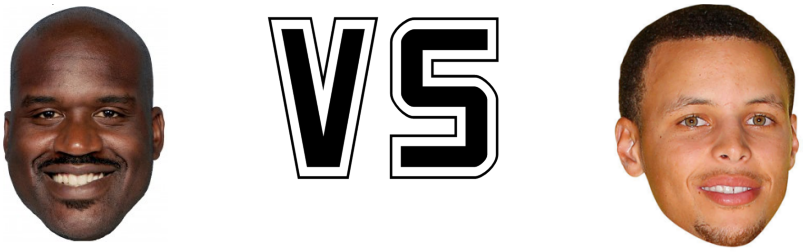
2. Biologists studying the healing of skin wounds measured the rate at which new cells closed a cut made in the skin of an anesthetized newt. Here are data from a random sample of 18 newts, measured in micrometers (millionths of a meter) per hour:

   29 27 34 40 22 28 14 35 26 35 12 30 23 18 11 22 23 33

   Calculate and interpret a 95% confidence interval for the mean healing rate μ.

Name: _____ Hour: _____ Date: _____

Chapter 9 Intro: Is Mrs. Gallas a good free throw shooter?



Mrs. Gallas claims she is an 80% free throw shooter. To prove her skills she shoots 50 free throws and makes 32 shots. Is Mrs. Gallas exaggerating about her free throw skills?

1. Identify the population, parameter, sample and statistic.

Population: _____ Parameter: _____

Sample: _____ Statistic: _____

2. There are two possible explanations for why Mrs. Gallas only made 32/50 shots.

1.)

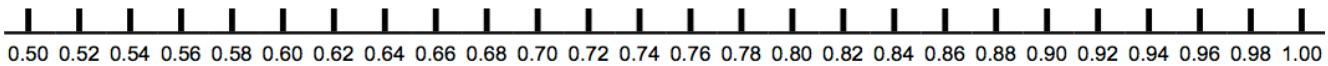
2.)

To test Mrs. Gallas' claim, we will **assume #1, she is an 80% free throw shooter**, and examine the likelihood that she makes 32/50 shots through simulation.

3. Use the spinner provided to simulate 50 free throws **shot by an 80% free throw shooter** by spinning 50 times. What is your sample proportion of shots made?

4. Repeat for another sample of 50 spins. Calculate the sample proportion.

5. Add your sample proportions to the dotplot on the board. Each person in your group should add two dots to the board. Sketch the dotplot below.



TheStatsMedic

Name: _____ Hour: _____ Date: _____

6. What does each dot represent?

7. One student says, “Each dot represents the proportion of free throws made out of 50 free throws shot by Mrs. Gallas.” Is this correct? Explain.

8. What percentage of the dots represent a percentage of 64% or less?

Interpret this percentage in context.

9. Based on your answer to Question 8, does the observed $\hat{p}$ = 0.64 result give convincing evidence that Mrs. Gallas is exaggerating? Or is it plausible that an 80% shooter can have a performance this poor by chance alone?

Name: _____ Hour: _____ Date: _____

Lesson 10.1: Day 1: Is Yawning Contagious?

Mythbusters investigated this question. Here's a brief recap. Each subject was placed in a booth for an extended period of time and monitored by hidden camera. 34 subjects were given a "yawn seed" by one of the experimenters: that is, the experimenter yawned in the subject's presence before leaving the room. The remaining 16 subjects were given no yawn seed.

1. Draw an outline of *Mythbuster's* experiment.

50 subjects

2. Here are the Mythbusters results.

	Subject Yawned?		
Yawn seed?	Yes	No	Total
Yes	10	24	34
No	4	12	16
Total	14	36	50

Call p_1 the true proportion of people who given the yawn seed will yawn. $\hat{p}_1$ = _____

Call p_2 the true proportion of people who given no yawn seed will yawn. $\hat{p}_2$ = _____

What is the difference in proportions $\hat{p}_1 - \hat{p}_2$? _____

3. Do the data provide *some* evidence that yawning is contagious? Why?

4. Adam Savage and Jamie Hyneman, the cohosts of *Mythbusters* used these data to conclude that yawning is contagious. Do you agree?

Name: _____ Hour: _____ Date: _____

In this Activity, your class will investigate whether the results of the experiment are statistically significant OR if they could have occurred purely by chance due to random assignment.

4. What is the null hypothesis?

The 50 people in the experiment are represented by the cards. A person is either a yawner or a non-yawner, no matter which treatment they are randomly assigned.

5. Shuffle the 50 cards and put them into two piles, one group of 34 that gets the yawn seed and one group of 16 that does not get the yawn seed. Record the proportion of people who yawned in each group. You will do this three times.

Trial	Proportion who yawned in yawn seed group, $\hat{p}_1$	Proportion who yawned no yawn seed group, $\hat{p}_2$	Difference in proportions, $\hat{p}_1 - \hat{p}_2$
1			
2			
3			

6. Make a class dotplot of the difference in proportions. Sketch below:

7. In what percent of the class's trials did the difference in proportions equal or exceed 29% - 25% = 4% (what *Mythbusters* got in their experiment)?

8. What conclusion can you draw about whether yawning is contagious?

Name: _____ Hour: _____ Date: _____

Lesson 10.1 Day 1: Sampling Distribution for a Difference in Proportions

Important ideas:

Check Your Understanding

Your teacher brings two bags of colored goldfish crackers to class. Bag 1 has 25% red crackers and Bag 2 has 35% red crackers. Each bag contains more than 1000 crackers. Using a paper cup, your teacher takes an SRS of 50 crackers from Bag 1 and a separate SRS of 40 crackers from Bag 2. Let $\widehat{p_1} - \widehat{p_2}$ be the difference in the sample proportions of red crackers.

(a) What is the shape of the sampling distribution of $\widehat{p_1} - \widehat{p_2}$? Why?

(b) Find the mean of the sampling distribution.

(c) Calculate and interpret the standard deviation of the sampling distribution.