The Practice of Statistics, 6e

Bedford, Freeman and Worth

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

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

4. Questions/Work Time

<u>Goals for the training:</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?

- Early research.
- Could lead to figuring out what causes it.
- Early detection.

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?

6 – she has a 50/50 chance of guessing correctly. The person does or does not have Parkinson's disease.

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

A lot, 10, 11, or 12 possibly.

Answers may vary.

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.

Why is this important?

If she's guessing what's the probability she guesses correctly?

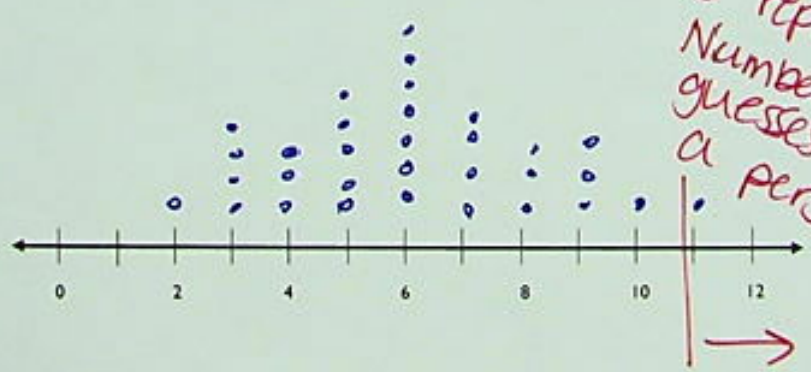
4. Mr. Wilcox will hand you 12 cards (shirts) that have been shuffled into a random order. Don't turn them over yet! On the back of some of them is "Parkinson's" and on the back of others is "No Parkinson's." For each card, guess Parkinson's or No Parkinson's. Once you have made your guess, turn the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

Each person in the group should complete a set of 12.

Tally of correct identifications	Number of correct identifications	Proportion of correct identifications
卌 |	6	6/12 = .5

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

What do the dots represent? Number of correct guesses made by a person guessing.



How many dots are at 11 or 12? Make %.

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?

1/32

Are you convinced? What percent would convince you?

Name: ______________________________________ Hour: _____

Barbie

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
Distance traveled								

Use your group's data to complete the following:

Used to predict (input)

Outcomes of study (output)

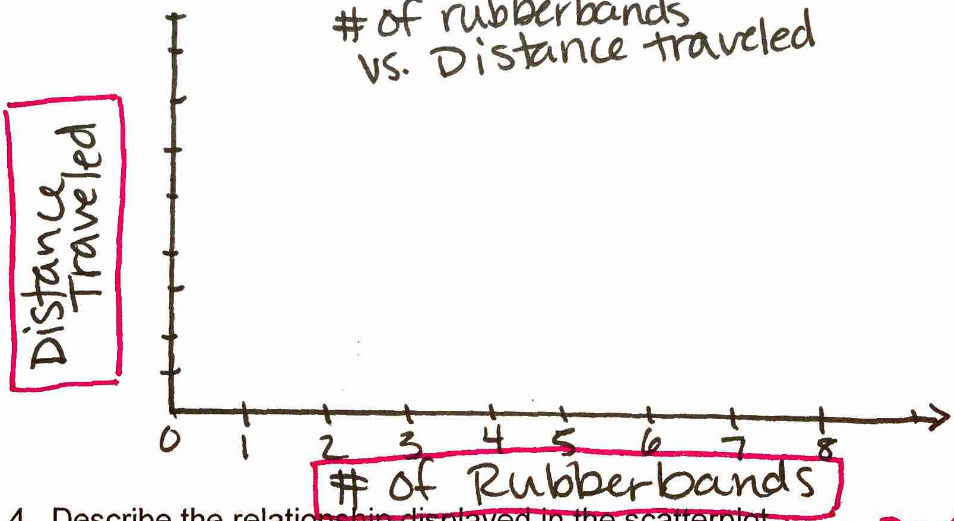
1. Identify the explanatory and response variables.

   # rubberbands – explanatory
   Distance traveled - Response

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

   2 quantitative

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



Answers vary

4. Describe the relationship displayed in the scatterplot.

Direction
Unusual features
Form
Strength (strong/weak)

Positive Direction

The dots are increasing. They form a tight linear pattern.
strength form
There are no unusual features or outliers.

TheStatsMedic

Name: ______________________________ Hour: _____

Lesson 3.1 – Displaying Relationships: Scatterplots

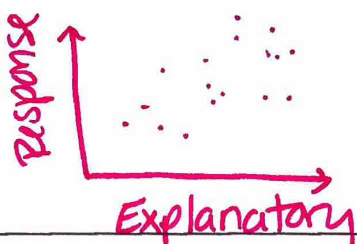
Big Ideas:

LT#1

Explanatory → used to predict

Response → outcome, responds to explanatory.

LT # 2 & 3



Describing:

Direction (+/-/None)

Unushal Features

Form (Linear/Nonlinear)

Strength

Check Your Understanding:

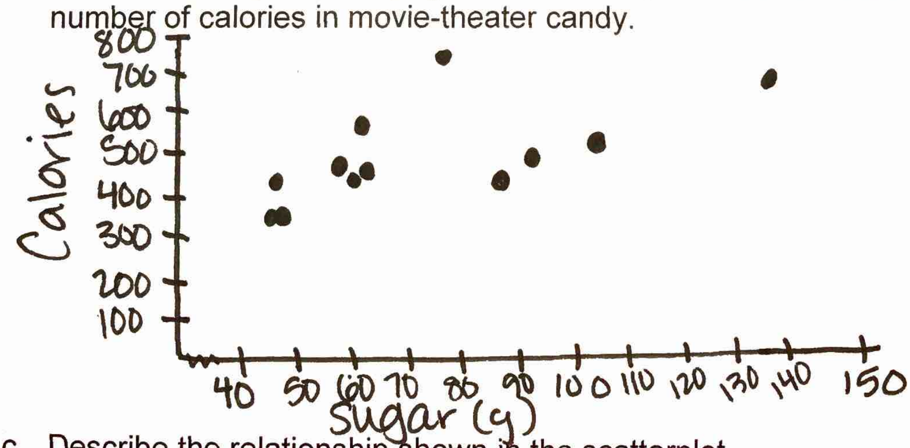
1. Is there a relationship between the amount of sugar (in grams) and the number of calories in movie-theater candy? Here are the data from a sample of 12 types of candy.

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.

Sugar is explanatory and calories are the response. If you add sugar the calories will go up.

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

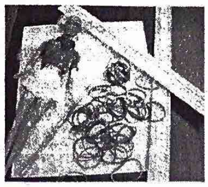


c. Describe the relationship shown in the scatterplot.

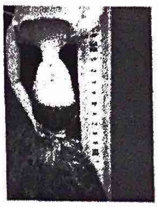
The scatterplot shows a positive, fairly strong linear pattern. There is a possible outlier with Peanut M&M's at (79, 790).

Name: ______________________________________ Hour: _____ Date: ______

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



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

"y-hat" means predicted y

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.

$\hat{y} = 25.333 + 7.464x$

$\widehat{\text{Distance}} = 25.333 + 7.464(\text{Rubberbands})$

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

$25.333 + 7.464(5)$

$\widehat{\text{Distance}} = 62.653$ cm

Residual = Actual − Predicted

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?

$64 - 62.653 = 1.347$ cm

It was 1.347 grams too low.

The predicted distance

Extrapolation

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?

$25.333 + 7.464(20) = 174.613$ cm

Less because it is far away on the x-axis.

(0, y-int)

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

25.333, when we use 0 rubberbands, ($x=0$) the predicted distance traveled is 25.333 cm. (y-int.)

slope = change in y / change in x

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

7.464, When we add 1 rubberband, (when x increases by 1) the predicted distance traveled increases by 7.464 cm. (slope)

TheStatsMedic

Name: ______________________________ Hour: _____ Date: ______

Lesson 3.2 – Prediction, Residuals, Interpreting a Regression Line

Big Ideas: LT#1 Predictions	LT #2 Residuals	LT #3 y-int & slope
$\hat{y} = a + bx$ ($\hat{y}$: Predicted, a: y-int, b: slope) - Be careful of extrapolation!	Resid = Actual − Predicted R = A − P The actual y-context was resid. higher/lower than predicted for x = #.	y-int: When x = 0 context the predicted y-context is y-int. Slope: With each additional x-context the predicted y-context increases/decreases by slope.

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.

With each additional week, the predicted weight increases by 40 grams.

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

Yes, when a rat is 0 weeks old, the predicted weight is 100 grams.

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

$\widehat{\text{Weight}} = 100 + 40(16)$
$= 740$ grams

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

Residual = 700 − 740 = −40 grams
The actual weight is 40 grams lower than predicted when x = 16 weeks.

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.)

No, that would use x = 104 weeks and our data is from the first 25 weeks. This is extrapolation.

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?

Bey BEYONCÉ CRAZY IN LOVE

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

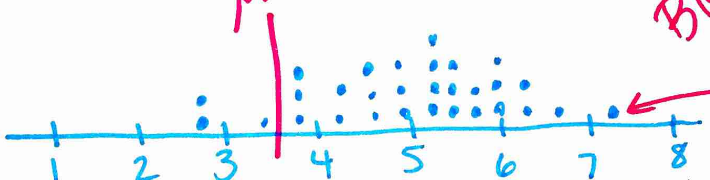
2. What is the average word length of your sample? ________ ← $\bar{x}$

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

Convenience Sample

$\mu = 3.53$

Biased



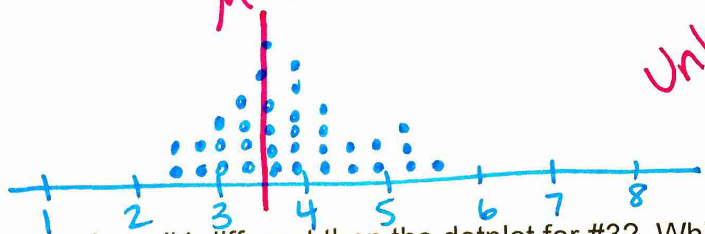
Average word length from 1 sample of 5 words.

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.

Simple Random Sample (SRS)

$\mu = 3.53$

Unbiased



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?

The Dotplot in #4 has less variability. The mean is lower in #4 than in #3.

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

$\mu = 3.53$ → About 3 to 4 letters.

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.

No we don't. The true mean word length is 3.53 letters. This is close to the Destiny's Child mean word length of 3.64. We don't have evidence that she did not write Crazy in Love.

Name: ______________________________ Hour: _____ Date: ______

Lesson 4.1 – Sampling Methods

Important ideas:		
• Population: entire group of individuals we want information about. • Sample: subset of individuals in the population from which we collect data	• Convenience sample - sampling individuals who are easy to reach. • Voluntary response: people choose to be in sample	• Simple Random Sample (SRS) Every group is equally likely to be chosen. To conduct: 1 Label 2 Randomize 3 Select.

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?

   Voluntary response

   b. Explain why this sampling method is biased.

   Only people who are very passionate about the ban will call in. They don't represent the population.

   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?

   Likely greater because people who call in feel strongly that they should be banned. People who don't care wouldn't call.

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?

   convenience sample

   b. Explain why this sampling method is biased.

   The sample doesn't represent the population. Most of the people she talks to are probably students.

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

   They should have done a simple random sample from the population.

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.)?

① ✓✓✓ ③ ✓✗✓ ⑤ ✓✗✗ ⑦ ✗✗✓
② ✓✓✗ ④ ✗✓✓ ⑥ ✗✓✗ ⑧ ✗✗✗

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 makes (.72)(.72)(.72) =.373	0 makes and 3 misses (.28)(.28)(.28) = .022 1 make and 2 misses (.72)(.28)(.28) =.056 × 3 ways	2 makes and 1 miss (.72)(.72)(.28) =0.145 × 3 ways

$3 \times (\text{make})^2 (\text{miss})^1$

${}_3C_2 \cdot .72^2 \cdot .28^1$

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.

${}_nC_k \, p^k (1-p)^{n-k}$

n: Total Shots; k: makes; p: % make; $1-p$: % miss

Louisville	Memphis	Shots Remain.	Probability Memphis Win	Probability Memphis Lose	Probability Overtime
75	73	3	=.373	.022+.169 =.191	=0.435
75	74	2	make make $(.72)^2$=.5184	miss miss $(.28)^2$=.0789	miss make or make miss (.28)(.72)×2 = .4032
75	74	1	0	.28	.72

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

maybe - chance of losing in regulation goes down but they might have to go to overtime and could lose there.

Name: ______________________________ Hour: _____ Date: ______

Lesson 6.3 Day 1– Binomial Random Variables

Important ideas:	
LT #1 Binomial Setting B- Binary I- Independent N- Set Number of trials $n=$ S- same probability $p=$	LT #2 Binomial Prob. $P(X=k)$: must add to total $nC_k\, p^k (1-p)^{n-k}$ must add to 1 total #, # success, Prob success, Prob Failure

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.

Binary: success → Ace, Failure → Not Ace

N: $n=10$

I- Independent because of replacement

S: $p = 4/52$

Yes, binomial.

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

B- Success → over 6 ft. Failure → not over 6ft

Not binomial.

I- Independent? No! No replacement.

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 = success → Red light, failure → Not red

N = $n=10$

I = Independent

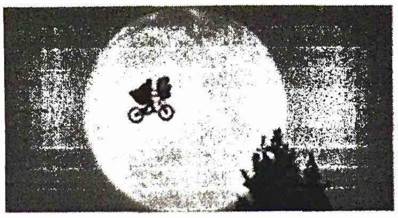
S: $p = .55$

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

$P(Y=7) = {}_{10}C_7 \times .55^7 \times .45^3 = .166$

Name: ______________________________ Hour: _____ Date: ______

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



Reese's pieces Peanut Butter

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.

Review of Chapter 6

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

   B - Binary < Success - orange, Failure - not orange   N - n = 50

   I - 10% condition $50 < 1/10 * 1000$ S - $p = .40$

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.

   X =

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

   0 10 20 30 40

4. What does each dot represent?

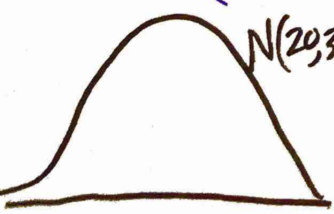
   The number of orange from a sample 50.

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

   $\mu_x = n \times p = 50 \times .4 = 20$ $\sigma_x = \sqrt{n \cdot p(1-p)} = \sqrt{50 \cdot .40 \cdot .60} = 3.46$

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

   Normal because of Large Counts.



N(20, 3.46)

$n \times p$: $50 \times .40 = 20 \geq 10$ ✓ $n \times (1-p)$: $50 \times .60 = 30 \geq 10$ ✓

Name: ______________________________ Hour: _____ Date: ______

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

Divide by 50: $x/50$

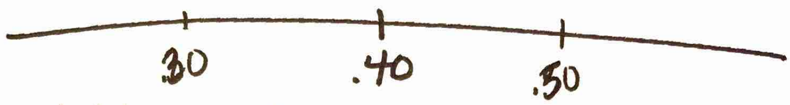
7. Use your samples from #2 and turn each number of orange candies into the **proportion of orange candies** in the sample. Write the proportions below and add them to the dotplot on the board.

$\hat{p} =$

8. Sketch the dotplot below.

Compare to original dotplot.

Sampling Distribution of $\hat{p}$



9. What does each dot represent?

The proportion of orange from a sample of 50

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

$\mu_{\hat{p}} = p \rightarrow$ $\mu_{\hat{p}} = \frac{20}{50} = .40$

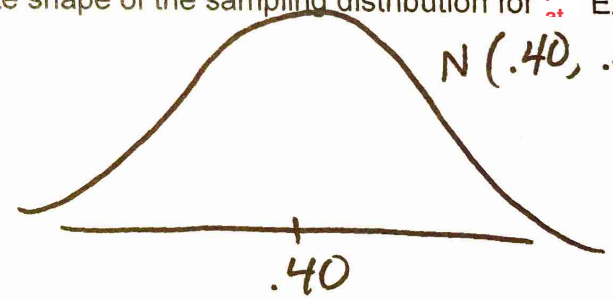
$\sigma_{\hat{p}} = \sqrt{\frac{50 \times .4 \times .6}{50}} \leftarrow \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

$= 0.069$

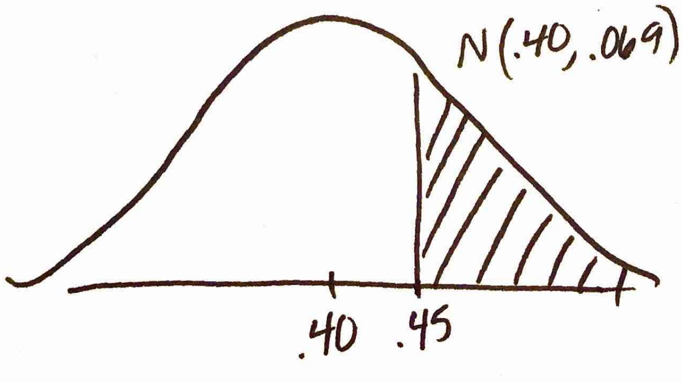
11. What is the approximate shape of the sampling distribution for pn Explain and sketch it below.

Large counts $n \cdot p \geq 10$ $n \cdot (1-p) \geq 10$

Normal because of large counts



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 45% or greater?



$Z = \frac{\hat{p} - p}{\sigma} \leftarrow$ OR $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$

$Z = \frac{.45 - .40}{.069}$

$= .72$

$\rightarrow \boxed{= .2358}$

Name: ______________________________ Hour: _____ Date: ______

Lesson 7.2 – The Idea of a Sampling Distribution

Important ideas: LT#1 mean and SD	LT#2 Normal:	LT#3 Probability
$\mu_{\hat{p}} = p$	Large counts	If sampling dist of $\hat{p}$ is approx Normal:
$\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$	$n \times p \geq 10$	use $z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$
	$n \times (1-p) \geq 10$	

*If the 10% condition is met!

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 $\hat{p}$.

$\mu_{\hat{p}} = 0.75$

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

$\sigma_{\hat{p}} = \sqrt{\dfrac{.75 \times .25}{1000}} = .014$

The proportion of young adults who watch online videos in a sample of 1000 typically varies by .014 from the true proportion of 0.75.

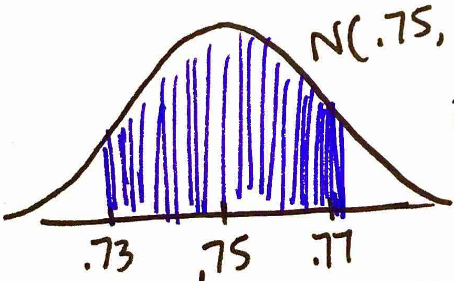
3. Is the sampling distribution of $\hat{p}$ approximately Normal? Check that the Large Counts condition is met.

Yes, it is approx. normal.

$.75 \times 1000 = 750 \geq 10$ ✓

$.25 \times 1000 = 250 \geq 10$ ✓

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



$N(.75, .014)$

$z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$

$z = \dfrac{.73-.75}{.014} = -1.43$

$z = \dfrac{.77-.75}{.014} = 1.43$

$\rightarrow \boxed{.8472}$

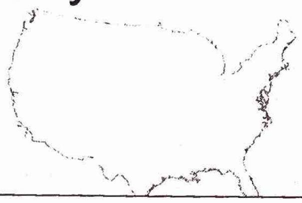
5. If the sample size were 9000 rather than 1000, how would this change the sampling distribution of $\hat{p}$?

The shape would remain the same (approximately Normal). The center would stay the same (p = 0.75). The variability would decrease by a factor of 1/3.

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 high school seniors to estimate a 95% confidence interval for the mean number of states a senior can name in 1 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? Quantitative → means

Categorical → proportions

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: μ = true mean # of states Confidence level: 95%

PLAN: Identify the appropriate inference method and check conditions.

Name of procedure: One sample t interval for μ

Check conditions:

- Random: Assumed ✓
- 10%: $n <$ 1/10 all seniors
- Normal: $n \geq 30$ CLT or Sample shows no strong skew or outliers.

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

General Formula for any confidence interval: Point Estimate ± Margin of Error

Specific Formula for this confidence interval: $\bar{x} \pm t^* \frac{S_x}{\sqrt{n}}$

Plug numbers into the formula:

Answer:

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

Interpret: We are 95% confident that the interval from ___ to ___ captures the true mean # of states a senior can name in 1 min.

TheStatsMedic

Name: ______________________ Hour: _____ Date: ______

Lesson 8.3 Day 2 – The Four Step Process

Important ideas:

LT #1 4 steps (Changes)

State: $\mu \rightarrow$ true mean

Plan: One sample t interval for μ

Normal condition
- Pop. is Normal
- $n \geq 30$ CLT
- sample shows no strong skew or outliers

Do: $\bar{x} \pm t^* \frac{S_x}{\sqrt{n}}$

Conclude: - None

LT #2 Sample Size

$$\text{margin of Error} = t^* \frac{S_x}{\sqrt{n}}$$

Use z^* in place of t^* if it's unknown.

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?

$$30 = 1.645 \times \frac{154}{\sqrt{n}}$$

$$\sqrt{n} = \frac{1.645 \times 154}{30}$$

$$\sqrt{n}^2 = \left(\frac{1.645 \times 154}{30}\right)^2$$

$$n = 71.31 \rightarrow \boxed{72 \text{ students}}$$

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.

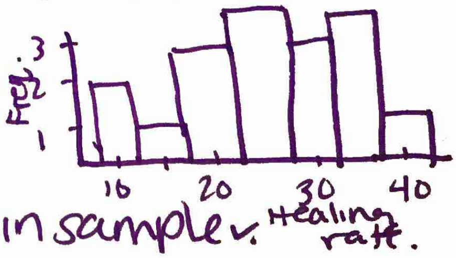
State: Estimate a $\mu \rightarrow$ true mean healing rate at 95% level

Plan: One sample t interval for μ

Random: "random sample of 18" ✓

10%: $18 < \frac{1}{10} \times$ All newts ✓

Normal: No strong skew or outliers in sample ✓.



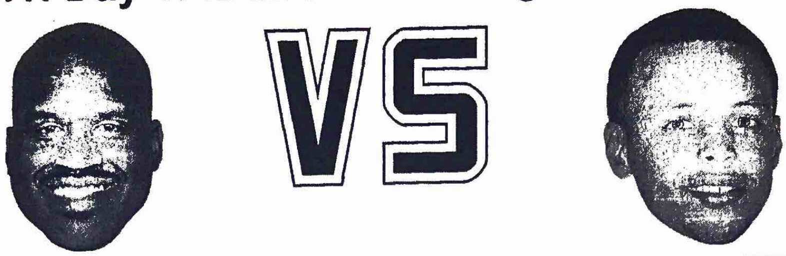
Do: Pt. Est ± margin of error

$$\bar{x} \pm t^* \frac{S_x}{\sqrt{n}} \rightarrow 25.67 \pm 2.110 \frac{8.32}{\sqrt{18}} \rightarrow (21.53, 29.81)$$

Conclude: We are 95% confident that the interval from 21.53 to 29.81 micrometer per hour captures the true mean healing rate.

Name: ______________________________ Hour: _____ Date: _____

Lesson 9.1: Day 1: 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: All free throws shot by Mrs.G Parameter: $p \rightarrow$ true prop. made FT

Sample: 50 free throws Statistic: $\hat{p} = \frac{32}{50} = .64$

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

$H_0: p = .80$ 1.) Mrs. G is an 80% shooter but had an off day..

$H_a: p < .80$ 2.) Mrs. G is a liar.

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. $\hat{p} =$

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? $\hat{p} =$

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.

0.50 0.52 0.54 0.56 0.58 0.60 0.62 0.64 0.66 0.68 0.70 0.72 0.74 0.76 0.78 0.80 0.82 0.84 0.86 0.88 0.90 0.92 0.94 0.96 0.98 1.00

proportion of FT made

Name: ______________________________ Hour: _____ Date: ______

6. What does each dot represent?

The proportion of free throws made from a sample of 50. shot by an 80% shooter.

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.

No, we do not know if Mrs. G is an 80% shooter. The dots represent a proportion of made shots by an 80% shooter.

P-value

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

Interpret this percentage in context.

Assuming Mrs. Gallas is an 80% free throw shooter, there is a _______ probability of getting a sample proportion of .64 or less purely by cha

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?

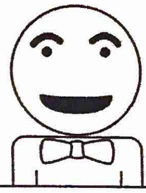
conclusion

Because the p-value of _______ is less/greater than 5% we do/do not have convincing evidence that ms. Gallas is not an 80% shooter.

Name: ______________________________ Hour: _____ Date: ______

Lesson 10.1: Day 1: Is Yawning Contagious?

MYTHBUSTERS



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 → Random Assignment → Group1 (34) → Yawn seed → Compare yawning

Random Assignment → Group2 (16) → No seed → Compare yawning

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. $\hat{p}_1$ = 10/34 = .29

Call p_2 the true proportion of people who given no yawn seed will yawn. $\hat{p}_2$ = 4/16 = .25

What is the difference in proportions $\hat{p}_1 - \hat{p}_2$? 0.29 - 0.25 = .04

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

Yes, people given the yawn seed yawned more often than people not given the yawn seed. (.29 to .25).

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

No, it could have happened that people who got the yawn seed yawned more often purely by chance.

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?

$H_0: P_1 - P_2 = 0$ → The treatment doesn't affect whether or not the person yawns.

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.

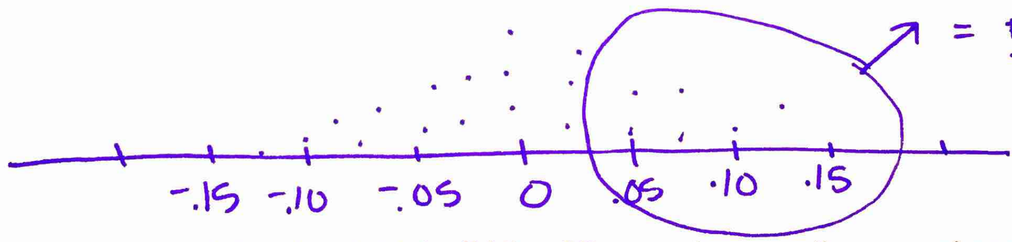
5. Shuffle the 50 cards and put them into two piles, one group of 34 that gets the yawn seed and one group of 16 that does not get the yawn seed. Record the proportion of people who yawned in each group. You will do this three times.

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	$\hat{P}_1 =$	$\hat{P}_2 =$	$\hat{P}_1 - \hat{P}_2 =$
2			
3			

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

Randomization Distribution of $\hat{P}_1 - \hat{P}_2$

P-value $= \frac{\text{\# circled dots}}{\text{total \# of dots}}$



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)?

P-value $= \frac{\text{\# circled}}{\text{\# total}}$

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

We don't have convincing evidence that yawning is contagious.

Name: ______________________________ Hour: _____ Date: ______

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

Important ideas:

LT#1 Shape, center, spread of the sampling distribution of $\hat{p}_1 - \hat{p}_2$.

Shape: Approx. Normal
Large counts
$n_1 \times p_1$
$n_1 \times (1-p_1) \geq 10$
$n_2 \times p_2$
$n_2 \times (1-p_2) \geq 10$

Center:
$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$

Spread:
$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

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.

p_1 p_2 n_1 n_2

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

Large Counts:

$50 \times .25 = 12.5$ $\quad 40 \times .35 = 14$
$50 \times .75 = 37.5$ $\quad 40 \times .65 = 26$

≥ 10 ✓
Approx. Normal

(b) Find the mean of the sampling distribution.

$\mu_{\hat{p}_1 - \hat{p}_2} = .25 - .35 = -.10$

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

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$= \sqrt{\frac{.25 \times .75}{50} + \frac{.35 \times .65}{40}} = .097$$

The difference in sample proportions typically varies by .097 from the true diff. in prop. of -.10.