Project NExT Course (X): Teaching Introductory Statistics in a Data-driven World

MathFest 2019 – Cincinnati, OH

Michael A. Posner, PhD, PStat®
Associate Professor of Statistics
Director, Center for Statistics Education
Villanova University
michael.posner@villanova.edu

Project NExT Course (X):
Teaching Introductory Statistics in a Data-driven World

Session 1: Thursday, August 1, 2019, 3:15-5:15pm
Session 2: Friday, August 2, 2019, 3:15-5:15pm
Convention Center, Room 250

Abstract: This session exposes participants to the big ideas of statistics and the ASA-endorsed Guidelines for Assessment and Instruction in Statistics Education (GAISE) report. It considers ways to engage students in statistical literacy and thinking and contrasts conceptual versus procedural understanding in the first statistics course. Participants will see popular activities that all statistics instructors should know. I will share resources for datasets, best practices, and how to become involved in the statistics education community.

TENTATIVE SCHEDULE

Session 1: Thursday, August 1, 2019, 3:15-5:15pm

3:15 – 3:45	Introductions, Overview, and GAISE
3:45 – 5:15	Conceptual Understanding in Statistics (with Activities)

Session 2: Friday, August 2, 2019, 3:15-5:15pm

3:15 – 3:25	Questions/Review from Session One
3:25 – 4:15	Multivariate Thinking
4:15 – 5:15	Investigative Process

Michael Posner (Sepia, 2006) is an Associate Professor of Statistics at Villanova University and Founding Director of the Center for Statistics Education. He is the Chair-elect of the ASA Section on Statistics Education, the Vice-Chair of the Data Science Subcommittee of AMATYC, a past Chair of SIGMAA Stat Ed, and a recently elected Fellow of the ASA. He has taught professional development courses at many different venues but, as a proud sepia dot, his favorite is clearly to Project NExTers at MathFest!

If you would attain to what you are not yet, you must always be displeased by what you are. For where you are pleased with yourself there you have remained. Keep adding, keep walking, keep advancing.

~Saint Augustine

Why Statistics?

1. Art Benjamin's TED talk (2009) about statistics and probability as the pinnacle of mathematics education: http://www.ted.com/talks/arthur_benjamin_s_formula_for_changing_math_education.html

2. McKinsey Report (2011) on "Big Data: The Next Frontier for Innovation, Competition, and Productivity" (http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation)

   > "By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions"

3. NYTimes article (2009) "For Today's Graduate, Just One Word: Statistics" (http://www.nytimes.com/2009/08/06/technology/06stats.html). Hal Varian, Chief Economist at Google, says "…the sexy job in the next 10 years will be statisticians. And I'm not kidding!"

4. Common Core State Standards push statistics, probability, and data analysis earlier into the curriculum (http://commoncoretools.files.wordpress.com/2012/04/ccss_progression_sp_hs_2012_04_21.pdf)

GAISE: Guidelines for Assessment and Instruction in Statistics Education
http://www.amstat.org/education/gaise/

The GAISE report consists of two reports - PreK-12 and College. The initial report was endorsed by ASA (American Statistical Association) in 2005 to provide guidance for statistical education. The updated report was approved by the ASA in 2016.

The **college** report consists of the following key elements:

Introduction
A brief history of the introductory college course and summarizes previous reports, including George Cobb's 1992 report and updates since the 2005 GAISE guidelines. A discussion of how statistics education has changed in the recent past, the impact of the Common Core (K-12), Data Science (as a related field), and the availability of data.

Recommendations
1-3 are statistical, 4-6 are pedagogical

1. Teach statistical thinking.
    a. Teach statistics as an investigative process of problem-solving and decision-making.
    b. Give students experience with multivariable thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data
6. Use assessments to improve and evaluate student learning.

Goal for Statistics Course

1. Students should become *critical consumers* of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted.
2. Students should be able to recognize questions for which the *investigative process* in statistics would be useful and should be able to answer questions using the investigative process.
3. Students should be able to produce *graphical displays and numerical summaries* and interpret what graphs do and do not reveal.
4. Students should recognize and be able to explain the central role of *variability* in the field of statistics.
5. Students should recognize and be able to explain the central role of *randomness* in designing studies and drawing conclusions.
6. Students should gain experience with how *statistical models*, including multivariable models, are used.
7. Students should demonstrate an understanding of, and ability to use, basic ideas of *statistical inference*, both hypothesis tests and interval estimation, in a variety of settings.
8. Students should be able to interpret and draw conclusions from standard output from *statistical software packages*.
9. Students should demonstrate an awareness of *ethical issues* associated with sound statistical practice.

What Should be Omitted or Coverage Reduced in the Intro Stat Course?

1. Probability Theory
2. Constructing Plots by Hand
3. Basic statistics (which are taught at the high school level) – mean/median, histograms, scatterplots, etc.
4. Drills with tables
5. Advanced training of software programs (should be in later courses)

Appendices (>100 pages)

A: Evolution of Intro Stat and Stat Ed Research
B: Multivariate Thinking
C: Activities, Project, and Datasets
D: Using Technology
E: Assessment Items
F: Learning Environments

For those who are teaching future teachers…

Statistics Education of Teachers (SET) -
http://www.amstat.org/education/SET/SET.pdf

GAISE has a PreK-12 report as well (published in 2007). The focus of PreK-12 is to develop statistical literacy. The premise is that foundational statistical concepts should be introduced and nurtured in the elementary grades and then strengthened and expanded throughout the middle, high school, and postsecondary grades. Emphasis is placed on the role of variability and the importance of context. Process components for the three levels of mastery can be summarized as follows:

Process Component	Level A	Level B	Level C
Formulate Question	Beginning awareness of the *statistics question distinction*	Increased awareness of the *statistics question distinction*	Students can make the *statistics question distinction*
Collect Data	Do not yet *design for differences*	Awareness of *design for differences*	Students make *designs for differences*
Analyze Data	*Use* particular properties of *distributions* in context of specific example	Learn to *use* particular properties of *distributions* as tools of analysis	Understand and *use distributions* in analysis as a global concept
Interpret Results	Do not *look beyond the data*	Acknowledge that *looking beyond the data* is feasible	Able to *look beyond the data* in some contexts

For those who teach K-8 or train K-8 teachers, Bridging the Gap Between Common Core State Standards and Teaching Statistics (http://www.amstat.org/education/btg/index.cfm) will be helpful. It provides 20 data analysis and probability investigations consist with the CCSS.

The report concludes with suggestions for how to make these changes, and includes numerous examples in the appendix to illustrate details of the recommendations.

Conceptual Understanding

From ASA/MAA Joint Committee on Undergraduate Statistics Report:

> Almost any course in statistics can be improved by more emphasis on data and concepts, at the expense of less theory and fewer recipes. To the maximum extent feasible, calculations and graphics should be automated.
>
> Any introductory course should take as its main goal helping students to learn the basics of statistical thinking. [These include] the need for data, the importance of data production, the omnipresence of variability, the quantification and explanation of variability.

Concepts vs. Procedures (Lock Morgan)

- Many (most? all?) introductory courses contain too much material.
- If students don't understand concepts, there's little value in knowing procedures.
- If students do understand concepts, specific new procedures are easy to learn.
- Minimize time spent teaching anything that doesn't help reinforce or deepen conceptual understanding!

Statistical Thinking vs. Mathematical Thinking.
Cobb and Moore: http://www.macalester.edu/~kaplan/statchat/2009-2010-year/2010-01/January-2010.pdf

Statistical Thinking

- **Uncertainty**
- **Bias (Sampling, Missing Data, etc.)**
- **Multivariable Thinking (Confounding, Causation, etc.)**

Conceptual understanding of standard deviation

(formulas are often effective for mathematically-inclined students to develop conceptual understanding, but should be in conjunction with below)

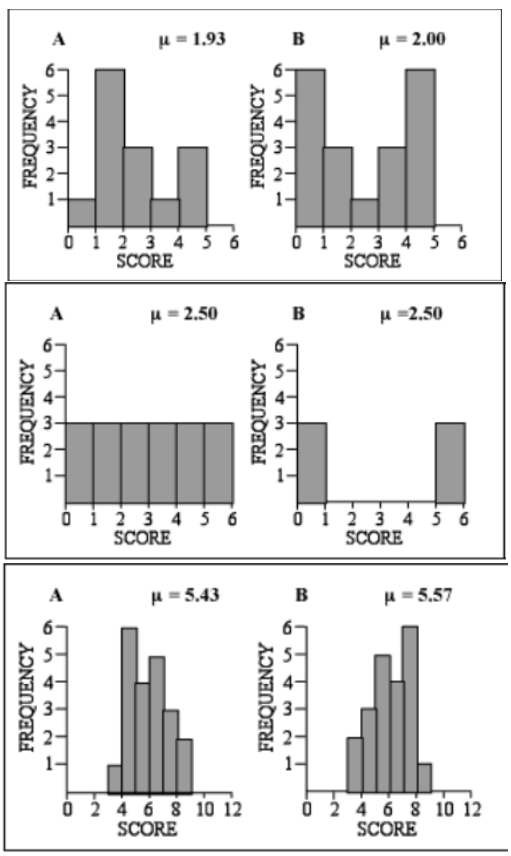
$$\text{Common formula: } s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{n-1}}$$

$$\text{Shortcut formula: } s = \sqrt{\frac{\sum_{i=1}^{N} x_i^{\,2} - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{n}}{n-1}}$$

There is no reason to teach the shortcut formula anymore! There is no need to have students calculate SD (more than once?).

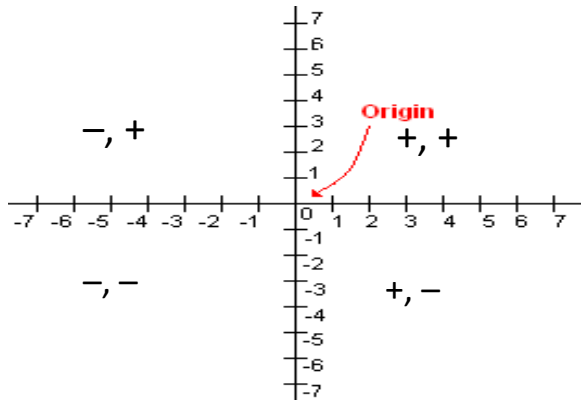
Conceptual question: For each of the following, choose one answer from the list below:

a) A has larger standard deviation than B
b) B has larger standard deviation than A
c) Both have equal standard deviation

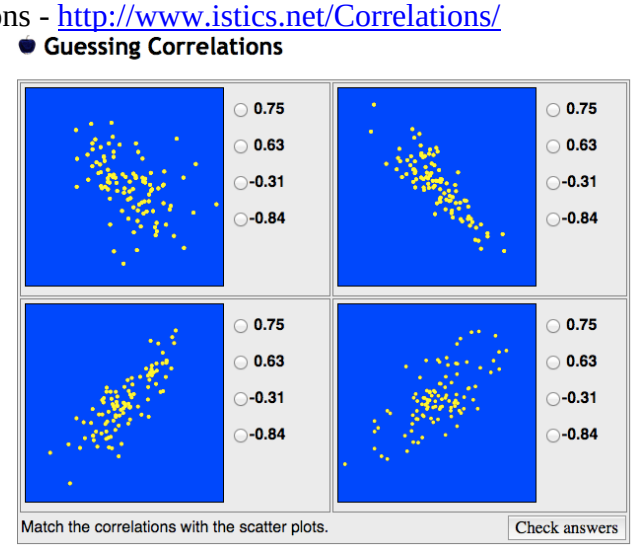


Conceptual understanding of correlation

$$r = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$$ is better than $$r = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum X_i)}{n}}{\sqrt{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}\sqrt{\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}}}$$



Guessing correlations - http://www.istics.net/Correlations/



Conceptual correlation question (from REGRESS (Enders, 2015))
A bank analyst wishes to determine whether there is a relationship between customers' average monthly credit card balance and the income stated on the original credit card application form. Which of the correlation coefficients below would show the strongest relationship between income and average monthly credit balance?

a) 0.85
b) 0.50
c) -1.25
d) -0.95
e) 0.00
f) I don't know

Conceptual Understanding of Confidence Interval

The applet http://www.rossmanchance.com/applets/ConfSim.html, among others, generates confidence intervals for means or proportions. Users set the population parameters, sample size, number of intervals, and confidence level. Click "Sample," and the applet will graph the intervals. Intervals shown in green contain the true population mean or proportion, while intervals in red do not. The true mean or proportion is shown by a blue line. The applet displays the proportion of intervals containing the population parameter for each sample and a running total of all the samples. Users can also click on a particular interval to display the numerical interval or sort the displayed confidence intervals from smallest to largest. This applet is part of a collection designed to accompany the textbooks *Investigating Statistical Concepts, Applications, and Methods* (ISCAM) and *Introduction to Statistical Investigations* (ISI).

Activity: What percent of Reese's Pieces are Orange?

a) Look at one candy. If we only care whether or not this candy is orange, what is the sample space for this "experiment?"
b) Look at your entire sample. How many observations are in your sample?
c) Sample size = n =
d) Let X = "number of orange candies in your sample." What are the possible values for X? Will everyone get the same value for X?
e) What type of variable is X?
f) Does the outcome of one observation in any way influence or affect the outcome of the next observation? What statistical term can we use to describe this relationship?
g) What is the total population size for these observations? Would you consider this more than 10 times the size of your sample? (note: this is about the FPC)
h) Based on your answers to (b)-(g), what type of random variable is X? Can you identify and report values for all of the parameters of the distribution?
i) Let π=proportion of all Reese's that are orange. Do you know the exact value for π?
j) Let p=X/n = proportion of Reese's in your sample that are orange. Do you know the exact value for p?
k) Will everyone in the class obtain the same value for p? Is p a random variable? What do we know about the distribution of this random variable?
l) Does this distribution remind you of any other distribution?
m) When is the "normal approximation to the binomial valid"? Is it valid here? What are the parameters of the normal distribution used to approximate Binomial (n,p)?
n) What else do we know about normal distributions?
o) What should be true about 95% of the p's that we randomly sample from the population of Reese's candies?
p) Do you think your value of p value is equal to the actual value of p? Why or why not? What can you say about the relationship between your p and the unknown value of π? Is this always true?
q) What can we say about the standard deviation of p?

Use an Internet browser to open the "confsim" applet. In this activity, let's assume that the true proportion of Reese's Pieces that are orange is 45%.

In this applet,

- Enter .45 for the value of π. (i.e. 45%)
- Enter 25 for the sample size.
- Enter 1 in the "intervals" box.
- Click the Sample button.

If you click on the resulting interval, you should see the value of p that generated the interval, along with the endpoint values.

a) Did all of us obtain the same interval? Why not?
b) Click the Sample button again. Did you get the same interval? Did the interval capture the "true value of π" (in this case, .45 because we said so)?
c) Click the Sample button a few more times. The interval will be red if .45 is not within the endpoint values. How often do you expect this to happen?
d) Put 100 in the intervals box and click Sample. The applet will count how many of the resulting intervals actually do capture π. What is this percentage?
e) Click the Sample button a few more times, what value does the "running total" percentage approach?
f) Change the sample size to 10 and hit return (or click Sample). How do the intervals change? Why does this make sense? How does the percentage of intervals containing π change?
g) Change the confidence level to 99% and hit return (or click Sample). How do the intervals change? Why does this make sense? How does the percentage of intervals containing π change?
h) Change the value of π to.1 and hit return (or click Sample). How do the intervals change? Why does this make sense? How does the percentage of intervals containing π change?
i) Set the sample size to 5 and hit return (or click Sample). How does the percentage of intervals containing π change? Why does this make sense?

Conceptual understanding of p-values

1. Out of this world – can you correctly guess the color of cards - **Demonstration**
2. Cookie game (or two-headed coin)
3. Simulation-based inference (Tintle article – best JSE article from 2018)

Activity: Hypothesis Testing Students develop the hypothesis testing process.
http://www.amstat.org/publications/jse/v2n1/eckert.html

Materials needed cookies (or other yummy incentive), two new decks of cards
Procedure summarized

- A "new" deck of cards is opened in front of the students and shuffled.
- Students are asked if they would like a chance to win a cookie to queue up.
- Students draw cards one at a time and show the color. If a red (black) card is drawn, the student receives a cookie. If not, the student sits down. Replace the card in the deck, reshuffle, and ask the next student to draw a card. After five or six draws the remaining students typically "give up".
- The deck of cards is put aside and students never know what is in the deck.

The parallel to hypothesis testing

- Hypotheses are written - Prior to the first draw, students expect the status quo that the proportion of red (black) cards is equal to 50%. The alternate (prior to the first draw) is that the proportion of red (black) cards is not equal to 50%.

$$H_o : \pi = 0.5$$
$$H_a : \pi \neq 0.5$$

- Evidence is collected - Students draw a card from the deck, recording the color. (Card is replaced; deck is reshuffled.) A success is drawing the red card and getting the cookie. A failure is drawing a black card and thus not getting a cookie.
- p-value is calculated - The probability of success using the expected proportion of 50% red is p = 0.5, the null hypothesis. The probability of failure is 1-p = 0.5. We observe a series of failures.

one failure	p-value = 0.5
two failures	p-value = 0.5 * 0.5 = 0.25
three failures	p-value = 0.5 * 0.5 * 0.5 = 0.125
four failures	p-value = 0.5 * 0.5 * 0.5 * 0.5 = 0.0625
five failures	p-value = 0.5 * 0.5 * 0.5 * 0.5 * 0.5 = 0.03125
six failures	p-value = 0.5 * 0.5 * 0.5 * 0.5 *0.5 * 0.5 = 0.015625

- Decision is made - We decide that it is unlikely that the deck has 50% red cards based on the observed evidence
- Population parameter is never known - Deck is put away and students don't know if they've made the correct decision.

Activity: Distracted Driving Example (day one!) – HANDS-ON DEMONSTRATION

Scenario: You are an analyst for a cell phone company. A recent report has come to your attention on the safety of cell phones. In this study, 52 people were randomly assigned to one of two groups, each containing 26 people. The experimental group drove a driving simulator while using a cell phone and the control group used the simulator without the cell phone. Part of the task given to them in the simulator was to exit the freeway at a particular exit. Of the 26 people who used a cell phone, 7 missed the exit. Of the 26 people who did not use a cell phone, 2 missed the exit.

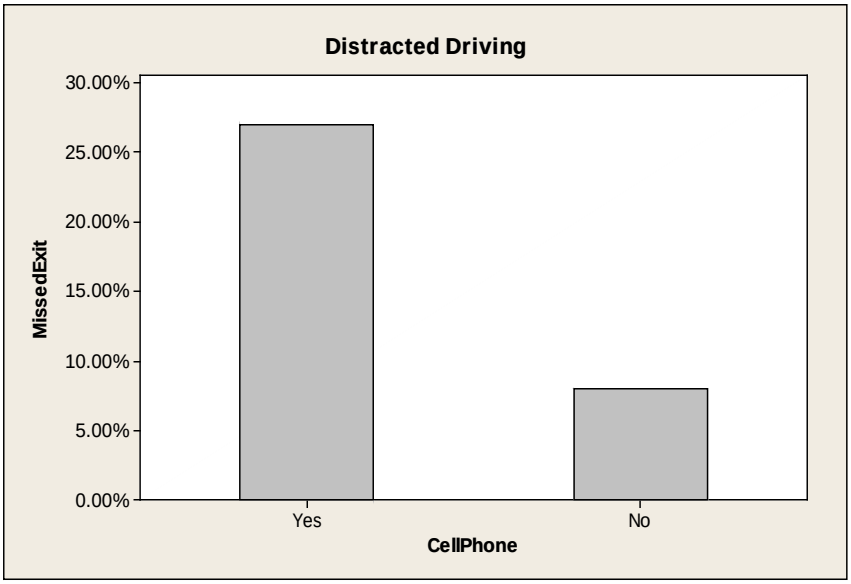
Examine the differences between these groups to determine the impact of using a cell phone on distracted driving.

- How might you summarize these results numerically?
- How might you summarize these results graphically?
- Is there evidence that cell phones cause distracted driving?
  - Simulation-based inference
  - Formal inference

Numerical summary:

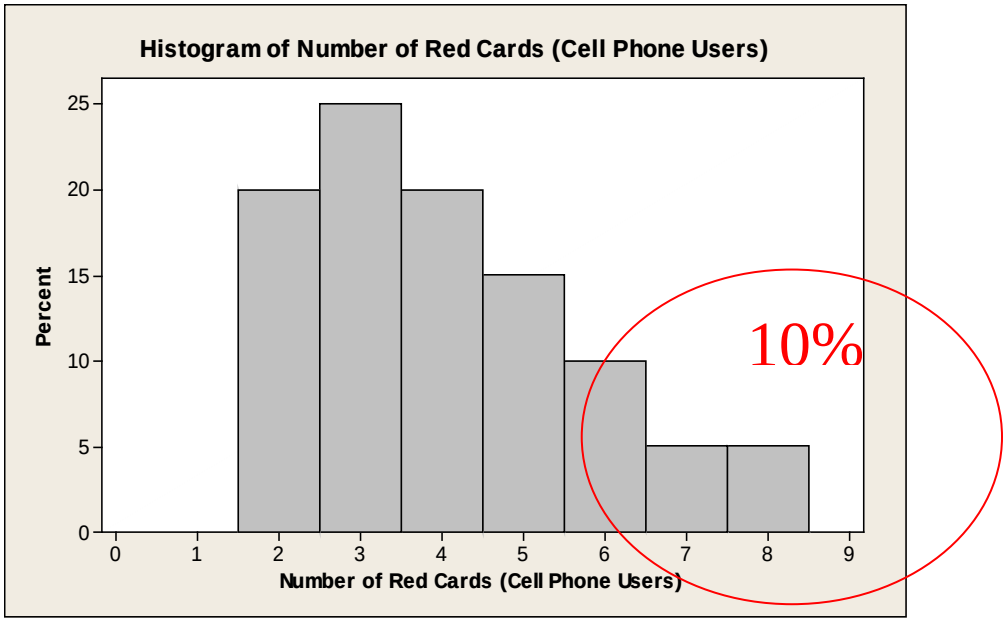
2x2 TABLE	MissedExit	NotMissExit	
CellPhone	7	19	26
NoCellPhone	2	24	26
	9	43	52
PERCENTS	(missed exit)		
CellPhone	7/26	**27%**	
NoCellPhone	2/26	**8%**	

Graphical summary:

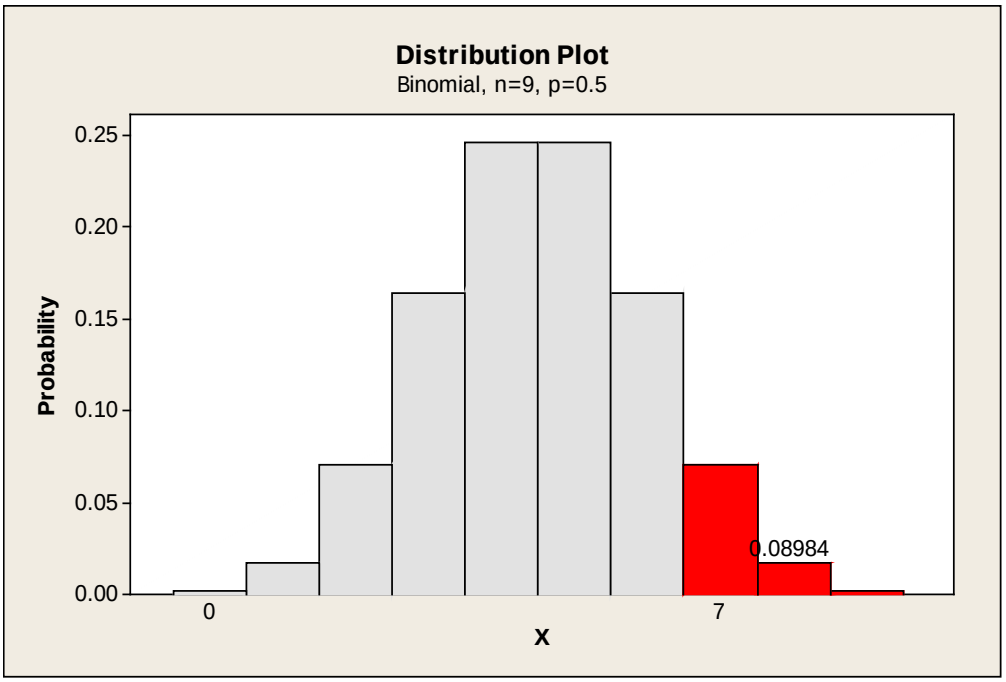


I

Simulation-based Inference: Get into groups of 3-4 students. Each group should have a deck of cards. There are 52 cards in a deck, 26 are red and 26 are black. Imagine that these 52 cards represent the drivers in the simulation. If you know that 9 of them missed their exit, can you determine the chance that 7 of those 9 would have come from the cell phone group, if the selection were done by random chance alone. To do this simulation, consider the red cards to be the cell phone group and the black cards are the no cell phone group. Draw nine cards at random (by shuffling the deck or random selection) and record the number of red cards that you get. Do this 2-3 times for each person in your group. Combining the results for your group, what percent of times did you get at least 7 from the red group? Is this an unlikely event?



Formal Inference:



Other important issues:

- Where do these data come from?
- Is the relationship causal?
- Are there measurement issues?
- What if the data were observational and not a designed experiment?
- Can you control for other factors (potential confounders)?
- What decision should be made about cell phone usage and driving distractions?

Activity: Simulation-based Inference - Smelling Parkinson's Disease (from Doug Tyson)

As reported by the Washington Post (http://tinyurl.com/JoyMilne), 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.

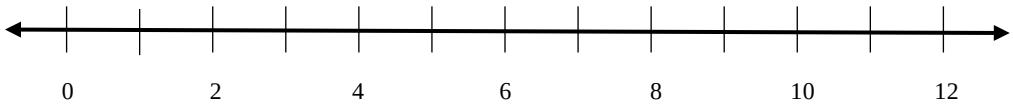
1. Why would it be important to know that someone can smell Parkinson's disease?
2. How many correct decisions (out of 12) would you expect Joy make if she couldn't really smell Parkinson's and was just guessing?
3. How many correct decisions (out of 12) would it take to *convince* you that Joy really could smell Parkinson's?

Simulating the Experiment
Let's test the hypotheses by running a simulation.

4. What are the null and alternative hypotheses for this test of Joy's ability?
5. I will hand you 12 cards (shirts) that have been shuffled into a random order. Don't turn them over yet! The red ones will indicate "Parkinson's" and the black ones will indicate "No Parkinson's." For each card, guess Parkinson's (red) or No Parkinson's (black). 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

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



7. 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? This is the simulated (estimated) *p*-value for our test.

8. How might we improve our estimate of the true probability?

Statistical Inference from the Simulation

9. Use the SPA Applet for One Categorical Variable at https://tinyurl.com/SPAapplets to run this simulation 10000 times. Then use that simulation to get a (likely) better estimate of the *p*-value for 11 or more shirts correctly identified, assuming that this person was just guessing. Is it *possible* that Joy correctly identified 11 shirts just by random chance (guessing)? Is it *likely*?
10. An interesting side note is that Joy's one "mistake" really wasn't a mistake. The shirt was worn by a person who supposedly didn't have Parkinson's even though Joy claimed that she could smell the telltale smell on that shirt. That person called the experimenters 8 months after the experiment and reported that he had just been diagnosed with Parkinson's disease. That meant that Joy correctly identified 12 out of 12 shirts. What is the approximate *p*-value for 12 shirts correctly identified, assuming that this person was just guessing?

Relating to "formal" Inference…

11. Recalculate the p-value using the binomial distribution and the normal approximation to the binomial. Do you think that these methods are appropriate?

Simulation-based Inference - Bonus or Rebate (from Doug Tyson)

Are people more likely to spend money if it is called a bonus or if it is called a rebate? Researchers Nicholas Epley (University of Chicago), Dennis Mak (Harvard), and Lorraine Idson (Harvard) investigated this question with a series of statistical investigations. One of those investigations was an experiment with 47 student volunteers from Harvard. Each student was called to a laboratory and was given $50 with no strings attached. However, 22 of the 47 students were randomly assigned to be told ''you are receiving this *tuition rebate* because our lab has a surplus of funds,'' that ''we will contact you in one week to ask you some questions about your *tuition rebate*,'' and that they should ask the experimenter ''if they have any questions about this *tuition rebate*.'' The other 25 students were given identical instructions, except that the words *tuition rebate* were replaced with *bonus income*. After one week, the students were asked to report how much of the $50 they spent. Many thanks to Nick Epley for providing these data!

Amount spent in dollars (Bonus income)				
0	0	0	0	0
0	0	0	0	10
10	10	20	25	26
30	30	40	50	50
50	50	50	50	50

Mean amount spent: $22.04

Amount spent in dollars (Tuition Rebate)				
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	10	20	30	50
50	50			

Mean amount spent: $9.55

1. Calculate the observed difference in the mean amount spent between these two groups (*bonus – rebate*).

One way to think of the null hypothesis in this case is to imagine that a student was going to spend the amount she spent no matter if she was told it was a rebate or a bonus. If the bonus/rebate wording made no difference in the amount a student would spend, then the difference in the mean amounts (*bonus – rebate*) would be the result of the random shuffling of subjects into their groups.

2. Is it *possible* that a difference in means of $12.49 or more is just the result of the random shuffling of subjects into their groups?
3. Is it *likely* that a difference in means of $12.49 or more is just the result of the random shuffling of subjects into their groups? Actually, that's not a fair question. Why?

Simulating the Experiment

4. Your instructor will give you 47 cards, one for each student. On one side of each card is the dollar amount the student spent. Let's investigate whether the chance of the random assignment is likely to produce a difference in means of $12.49. What are you assuming to be true here?
5. Shuffle your cards and deal them into two piles: 25 cards for the bonus group and 22 cards for the rebate group. Record the difference in means (*bonus – rebate*).
6. Combine your difference in means with that of your classmates to produce a dotplot of your results. Based on this small-scale simulation, does a difference in means as big or bigger than $12.49 seem likely to happen just by chance? Justify your answer.
7. How could you get a better idea of the probability that random chance would produce a difference in means as big or bigger than the observed difference?

Statistical Inference from the Simulation
Use the SPA Applet for One Quantitative Variable at https://tinyurl.com/SPAapplets to run a simulation of at least 10000 repetitions of the random assignment.

8. Does random chance seem like a reasonable explanation for a difference in means of $12.49 or greater? Support your answer with an approximate *p*-value.
9. Do you have strong evidence that the wording (bonus/rebate) *causes* a difference in the amount students will spend?

Extending question: to what population would you feel comfortable generalizing this result? Why?

Some other questions on conceptual understanding

1) Median

The ages of five people were gathered. The values of the first four were 18, 14, 16, 21. The value of the fifth could not be read. What are the possible values of the median of these five observations?
a) 15 b) 17 c) 18 d) somewhere between 14 and 16 e) somewhere between 16 and 18

2) Sampling distributions (from Hooks, et. Al, 2009)

Suppose that you have a very big container with 1000 candies in it; 600 are red and 400 are yellow. The candies are all mixed up in the container. With eyes closed, students draw 10 candies, one at a time. They record whether each candy was red or yellow, and then replace and remix the candies. The teacher asked each student to do this six times. Alex, Beverly, and Chang decided to play a trick on their teacher. Only one of them actually did the experiment, while the other two just made up their data. Below are their reports, where R = red candy and Y = yellow candy.

Alex:	Beverly:	Chang:
RRYRYRRYYR (6 red)	RYYRRRYRRY (6 red)	RRYRYRRRRY (7 red)
YRRYRRRYRY (6 red)	YRYRRYYRRY (3 red)	RYYYRRRRRR (7 red)
RRYRYRRRYY (6 red)	RRYRRRYRRR (8 red)	YRRRYYRRYY (5 red)
RYRRRYRYRY (6 red)	YRYYYYYYYY (1 red)	RRRRRYRRYR (8 red)
YRRYRRYYRR (6 red)	RRRRRRRRYR (9 red)	RYRRYYYRRR (6 red)
RRRYRYYRYR (6 red)	RYYRYYRRYY (4 red)	YRYRYRRRRR (7 red)

Circle the name of the student who you think is most likely to have done the experiment:
Alex Beverly Chang

Explain your reasoning.

3) Resistance to Outliers

The midhinge of a distribution is defined to be the average (mean) of the lower quartile and the upper quartile. The midrange of a distribution is defined to be the average (mean) of the minimum and the maximum. Which of these statistics is/are resistant to outliers?

a) The midrange only
b) The midhinge only
c) Both the midrange and the midhinge
d) Neither the midrange nor the midhinge

4) Sampling Distribution

Which of the following statements is NOT true according to the Central Limit Theorem? Select all that apply.

a) An increase in sample size from $n = 16$ to $n = 25$ will produce a sampling distribution with a smaller standard deviation.
b) The mean of a sampling distribution of sample means is equal to the population mean divided by the square root of the sample size.
c) The larger the sample size, the more the sampling distribution of sample means will resemble the shape of the population.
d) The mean of the sampling distribution of sample means for samples of size $n = 15$ will be the same as the mean of the sampling distribution for samples of size $n = 100$.
e) The larger the sample size, the more the sampling distribution of sample means will resemble a normal distribution.

5) P-values (from RPASS (Lane-Getaz, 2007))

A marketing researcher conducts an experiment on a new incentive strategy for customers and recruits 15 people to participate in her study. She performs the experiment and analyzes the results. She obtains a P-value of 0.17. Which of the following is a reasonable interpretation of her results?

a) She should reject the null hypothesis.
b) This proves that her experimental treatment has no effect on memory.
c) There is evidence of a small effect on memory by her experimental treatment.
d) There could be a treatment effect, but the sample size was too small to detect it

6) What does a p-value mean? A researcher tested the null hypothesis that the population proportion is equal to π_0 (H_0: $\pi = \pi_0$). The z-test for one proportion produced a p-value of 0.01. You should assume that all assumptions of the test have been satisfied. Identify which of the following statements are true.

a) There is a 1% chance of getting a result as extreme or more extreme than the observed one when H_0 is true.
b) There is a 1% chance that the null hypothesis is true.
c) There is a 1% chance that the decision to reject H_0 is wrong.
d) There is a 99% chance that the alternative hypothesis is true, given the observed data.
e) A small *p* value indicates a large effect (large difference from π_0).
f) Failure to reject H_0 means that the population proportion is probably π_0.
g) Rejecting H_0 confirms the quality of the research design.
h) If H_0 is not rejected, the study is a failure.
i) Assuming H_0 is true and the study is repeated many times, about 1% of these results will be even more inconsistent with H_0 than the observed result.

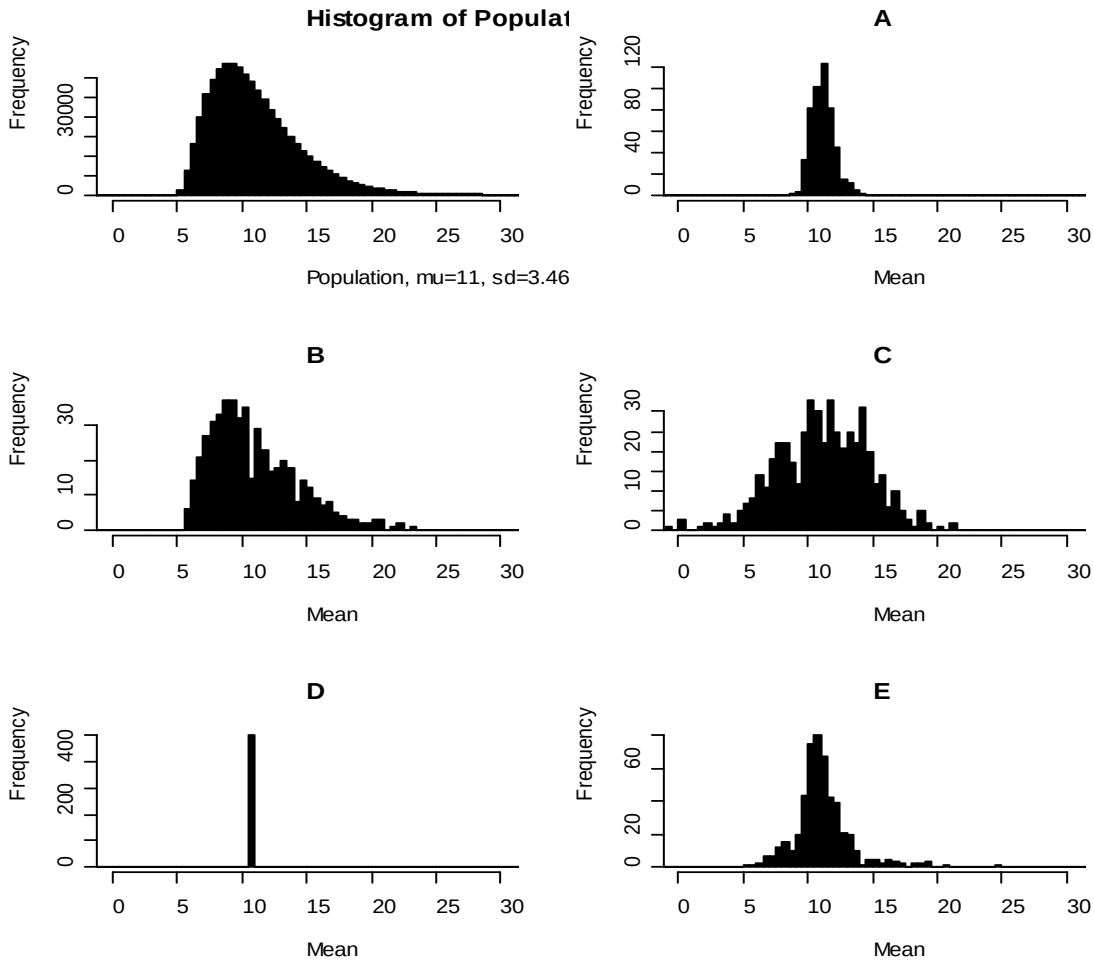
7) IQR and outliers - Identify the Error

Researchers examined the fat content of 48 eggs, measured as the percentage. They desired to create a boxplot to summarize this numeric variable. They found the five number summary to be 6%, 30%, 37%, 43%, and 80%. 3 x IQR was determined to be 39%, so values above 69% and below 4% were flagged as extreme outliers.

8) An email marketer is hiring people to respond to the spam, err, I mean email solicitations sent on his behalf. If the number of response calls each person can take in an hour has a mean of 8 with a standard deviation of 3, and he hires 5 people, then the distribution of the average number of calls these 5 people can take is…

a. $\sim N(8,3^2)$, for any distribution of calls
b. $\sim N(8/5,3^2)$, for any distribution of calls
c. $\sim N(8/5,3^2/5)$, for any distribution of calls
d. $\sim N(8,3^2/5)$, for any distribution of calls
e. $\sim N(8,3^2)$, if the distribution of calls is normally distributed
f. $\sim N(8/5,3^2)$, if the distribution of calls is normally distributed
g. $\sim N(8/5,3^2/5)$, if the distribution of calls is normally distributed
h. $\sim N(8,3^2/5)$, if the distribution of calls is normally distributed

9) Sampling distributions (from Hooks, et. Al, 2009)
The first graph below is a distribution for a population of test scores. Each of the other five graphs, labeled A to E represent possible sampling distributions of sample means for 500 random samples drawn from the population.



1. Which graph represents a sampling distribution of sample means for 500 samples of size 1?

   A B C D E

   Explain your answer:

2. Which graph represents a sampling distribution of sample means for 500 samples of size 16?

   A B C D E

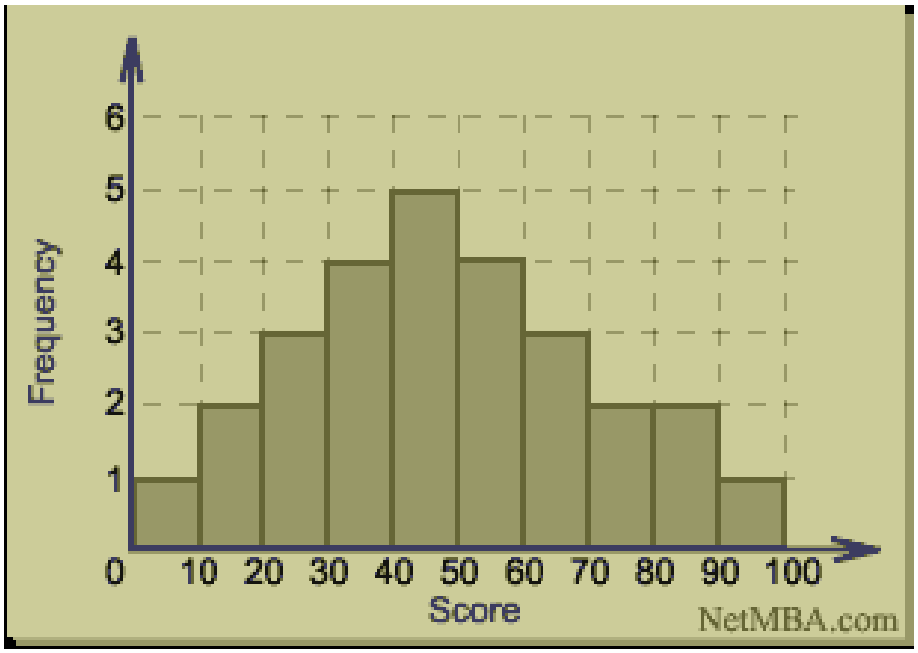
   Explain your answer:

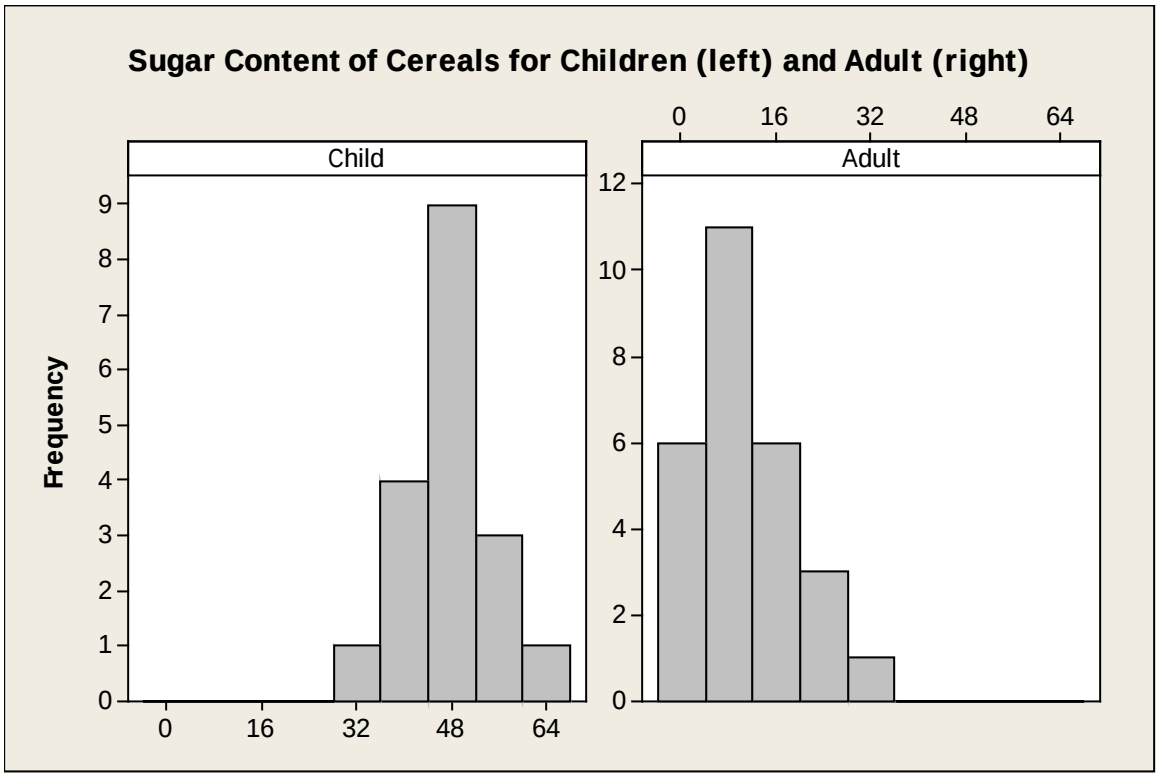
Graphical Displays of Data

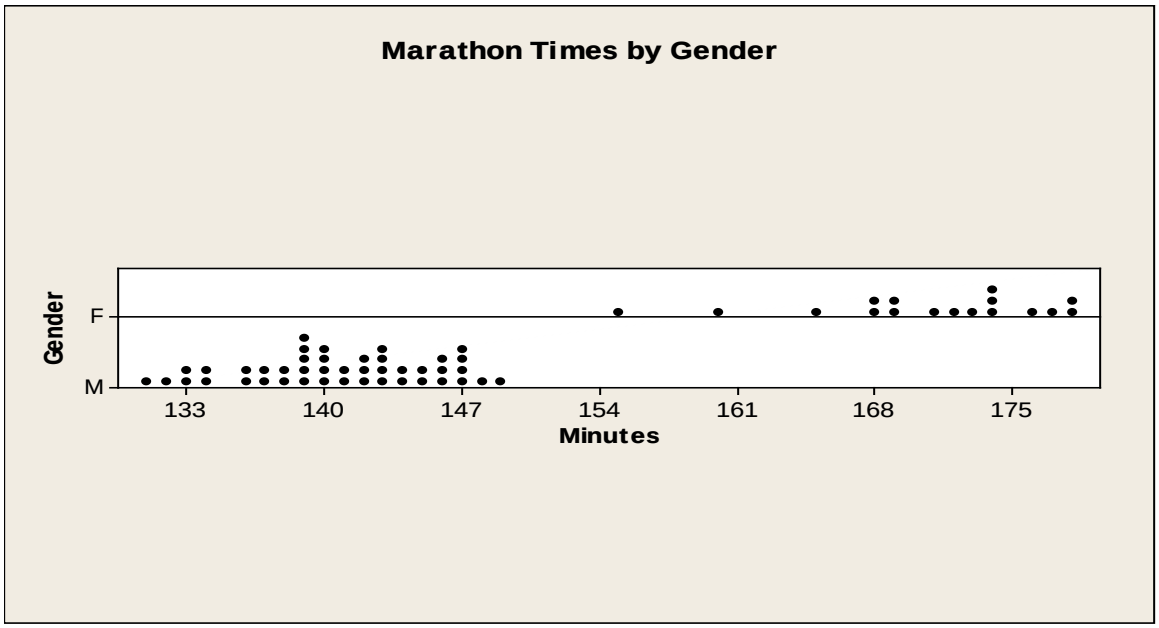
For each of the following, describe the distributions.

Recall SOCS+I+T: Shape, Outliers, Center, Spread, Interesting, Takehome

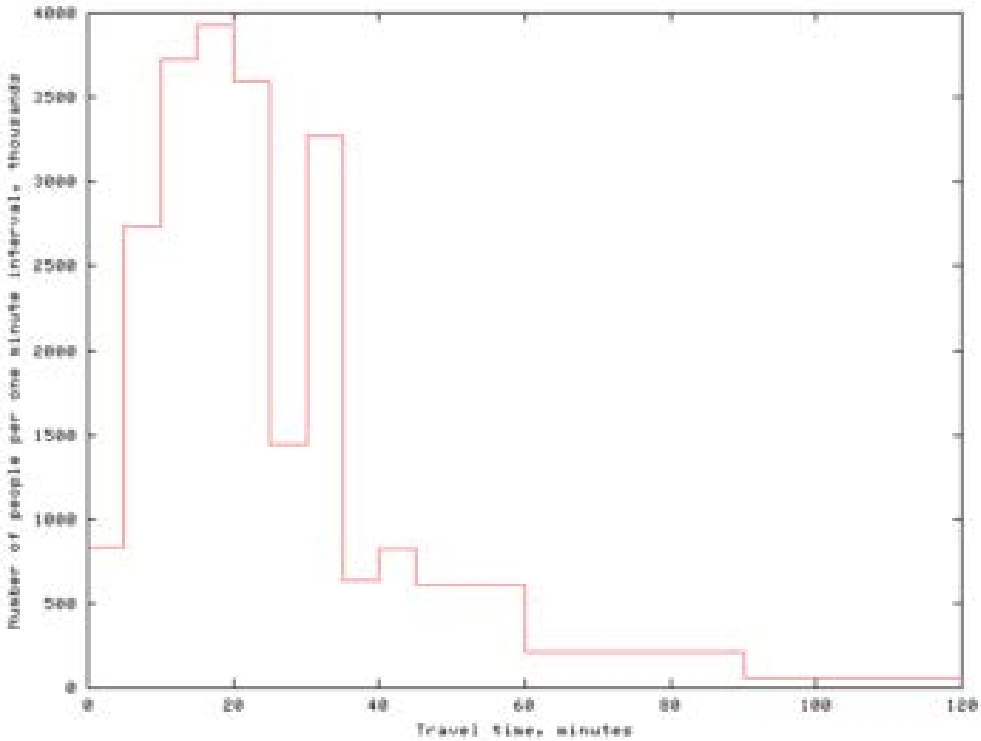
Scores on an exam:





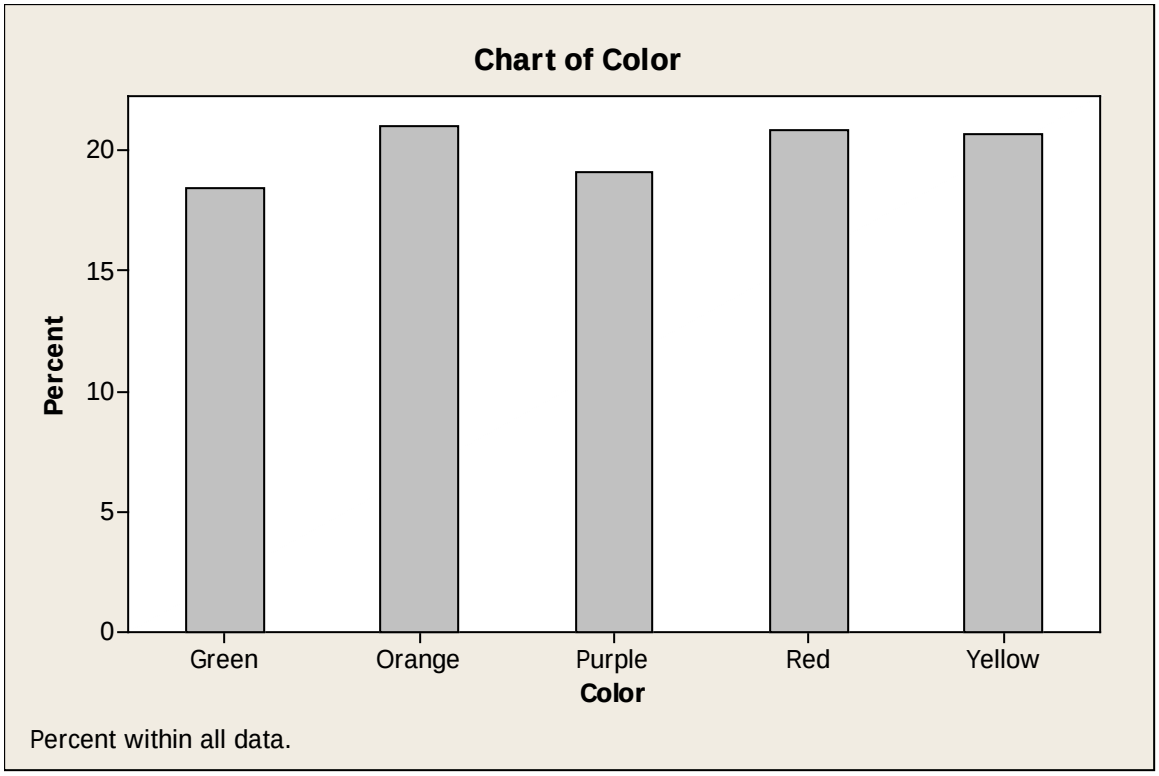


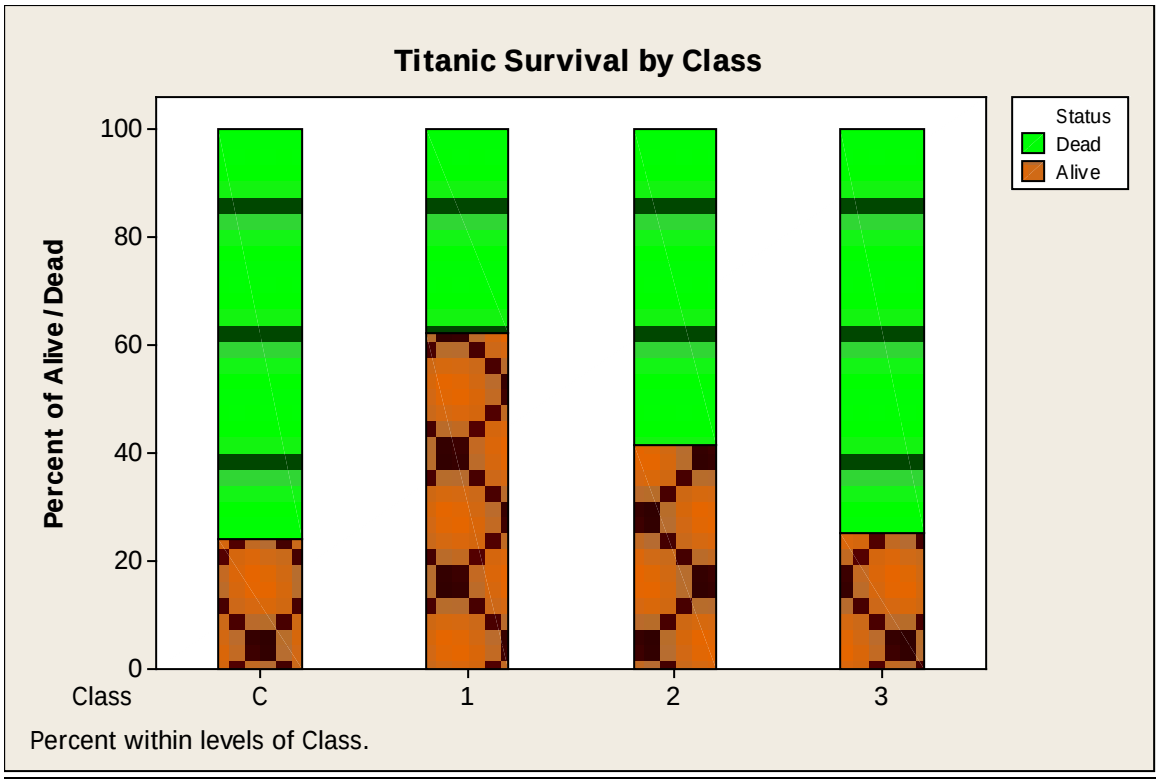
For the 124 million people who work outside their home, how long does it take them to get to work?

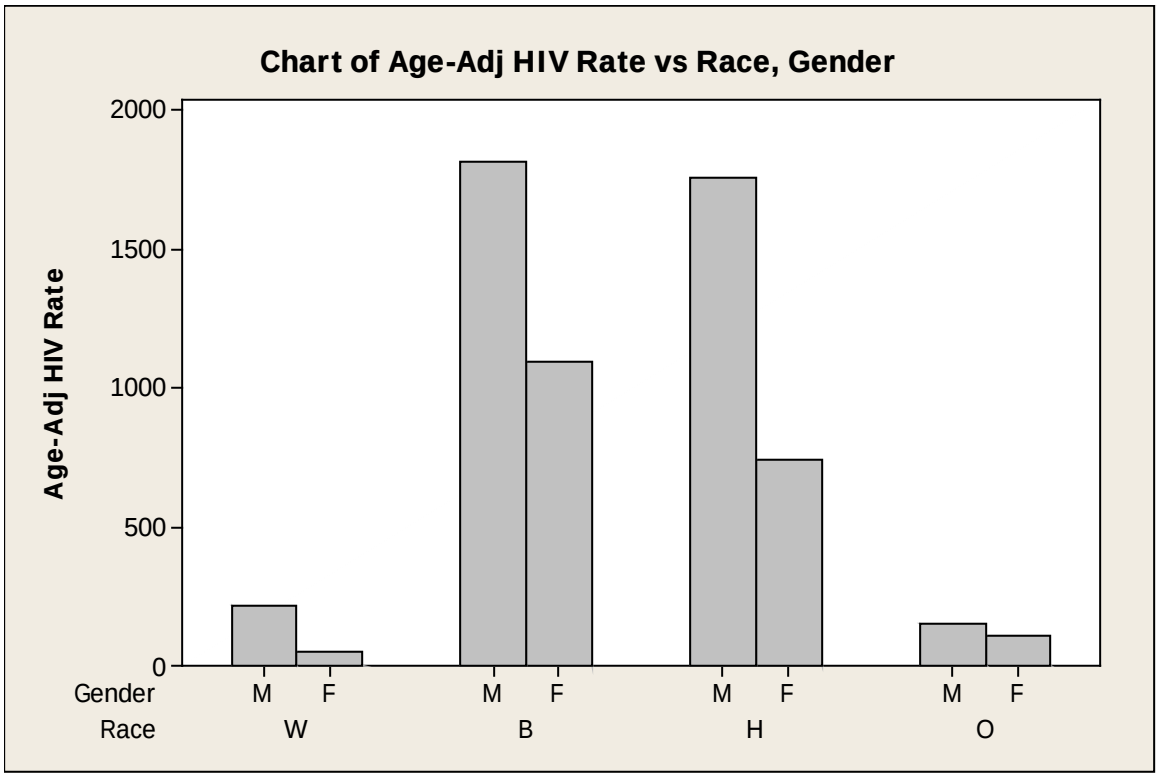


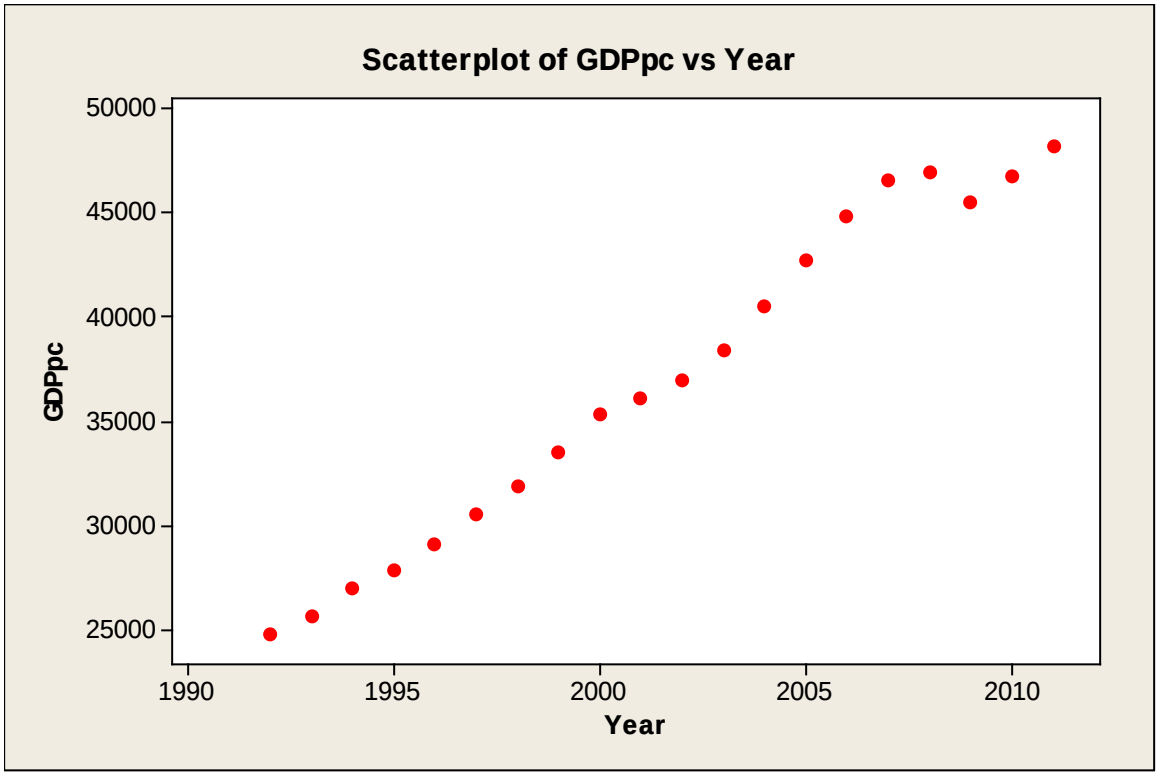
For each of the following, describe what you see.

Color of Skittles:

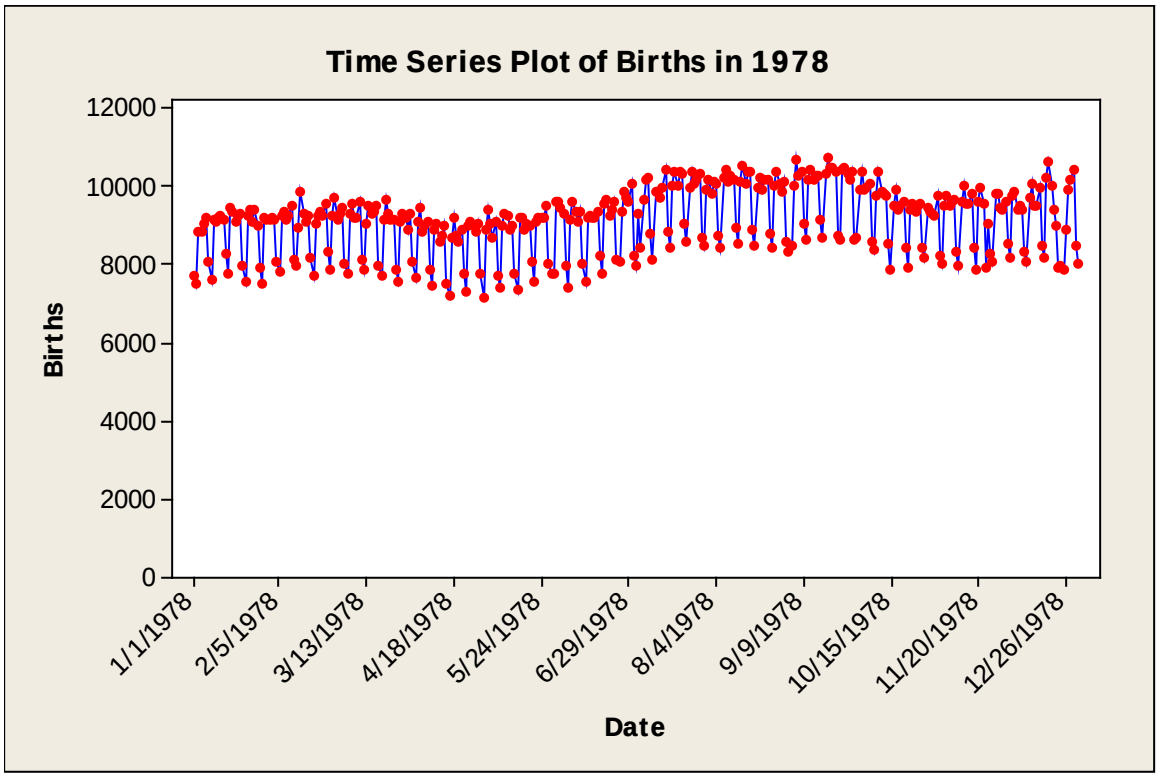




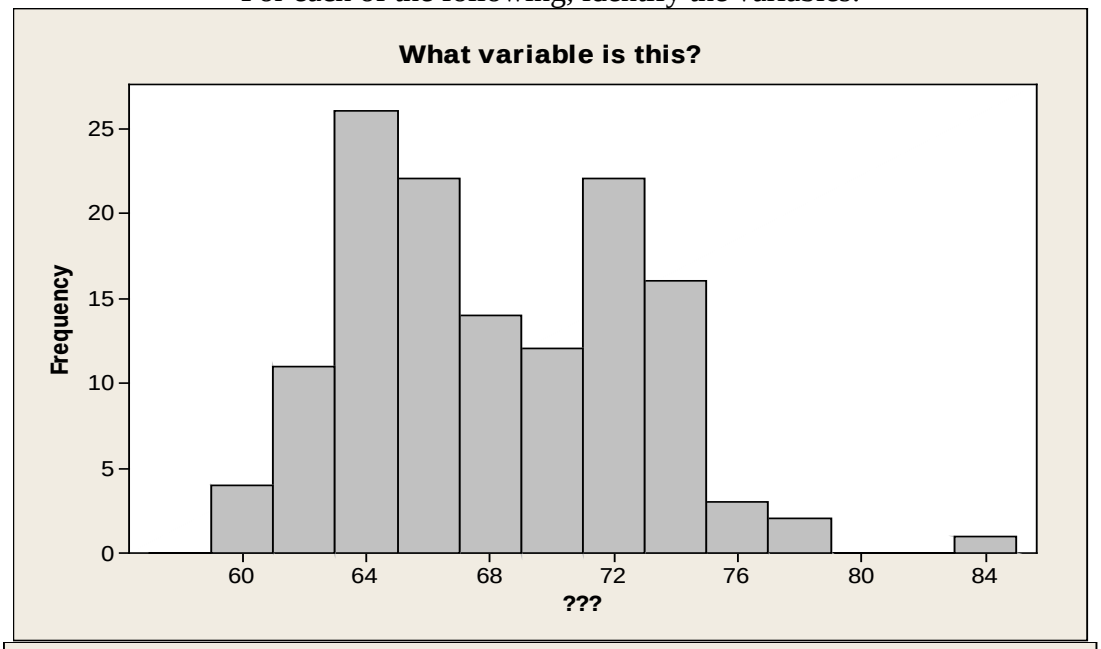


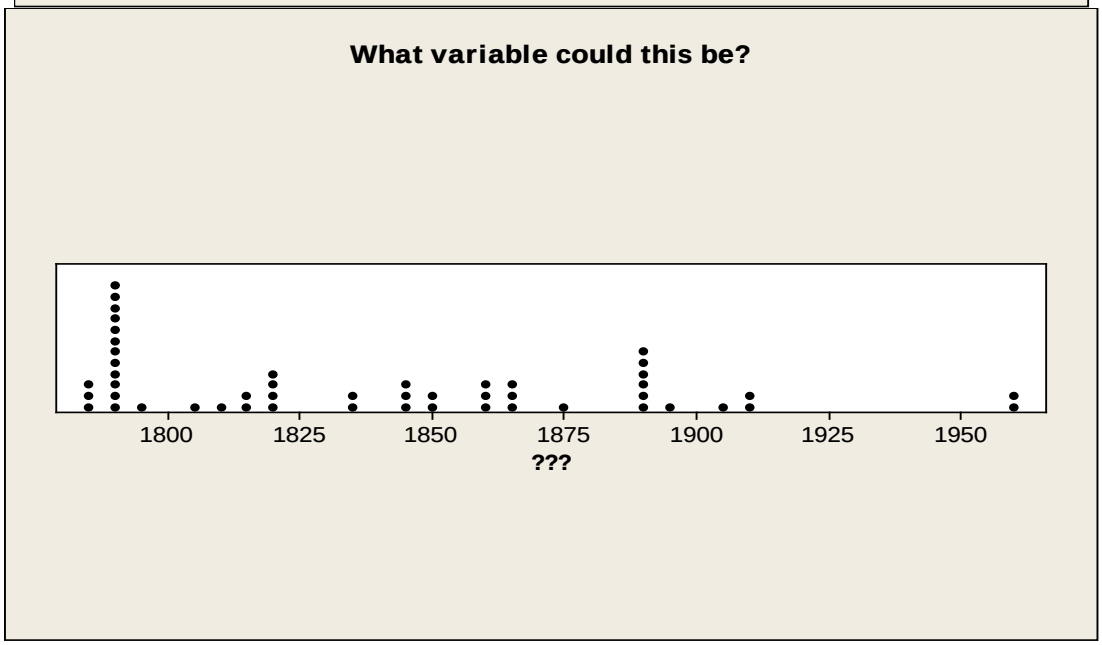


Scatterplot – Linear?, Direction, Strength


Time Series Plot of Births in 1978
Births
12000
10000
8000
6000
4000
2000
0
1/1/1978
2/5/1978
3/13/1978
4/18/1978
5/24/1978
6/29/1978
8/4/1978
9/9/1978
10/15/1978
11/20/1978
12/26/1978
Date

For each of the following, identify the variables:





Stem-and-Leaf Display

```
Stem-and-leaf of Unknown Variable  N  = 31
Leaf Unit = 1.0

 13  0  3333344444444
(6)  0  555788
 12  1  24
 10  1
 10  2
 10  2  69
 8   3  111234
 2   3  8
 1   4  4
```

Random Selection (RS) and Random Assignment (RA) Activities

Activity RS1 – Random Rectangles [details below]
A population of N=100 rectangles of various areas are provided. Students will estimate the average area by guessing, choosing a representative sample (n=5), and random sampling (n=5). They will learn how to select a random sample and show that only the latter method produces unbiased results.

Activity RS2 – Words in Gettysburg Address
See Rossman/Chance

Activity RS3 – Additional Sampling Techniques [details below]
Estimate the number of stars in the sky using Petocz's activity. The use of random sampling and other sampling methods are examined.

Activity RS4 – Pick a Random Number
Random numbers. Ask students to pick a random number from 1 to 10. Graph these. Show that students do not pick numbers at random (3 and 7 are typically favored). Can perform a chi-squared goodness-of-fit test on these data.

Activity RA1 – Fighting Cancer with Raspberries [details below]
Fighting cancer with raspberries: Demonstrating the value of random assignment in experimentation. Understand how random assignment eliminates bias leading to a consistent pattern when repeated.

Activity RA2 – Paper Helicopter Experiment
Paper helicopter experiment. See http://www.public.iastate.edu/~wrstephe/stat101L/ProjectS09.pdf and http://www.aoe.vt.edu/~mason/Mason_f/ASEE98.pdf.

Activity RS or RA - The Island
This activity from Michael Bulmer at the University of Queensland is a great example of experimental design. See http://www.maths.uq.edu.au/~mrb/ for more information.

Random Rectangles (Random Selection Activity 1)
This example is from Activity-Based Statistics. Second edition. By Schaeffer, Watkins, Witmer, and Gnanadesikan.

Random Rectangles Worksheet

Look at the following page for a few seconds. What is your best guess for the average area of the rectangles?

_____

Look at the following page again. Choose five rectangles that you think are representative. Write down their ID#. Calculate the area for each one. Then calculate the average area.

	Rectangle 1	Rectangle 2	Rectangle 3	Rectangle 4	Rectangle 5
ID#					
Area					

_____ = Average Area of the five rectangles

Look at the following page again. Select five random rectangles using the following table of random digits [or using a computer program]. Write down their ID#. Calculate the area for each one. Then calculate the average area.

	Rectangle 1	Rectangle 2	Rectangle 3	Rectangle 4	Rectangle 5
ID#					
Area					

_____ = Average Area of the five rectangles

Join together with three other students to form groups of four. Write down the average of the five random rectangles each group member has calculated. Take the average of these averages to get the average area for 20 rectangles.

	Member 1	Member 2	Member 3	Member 4
Average				

_____ = Average Area of the four group members

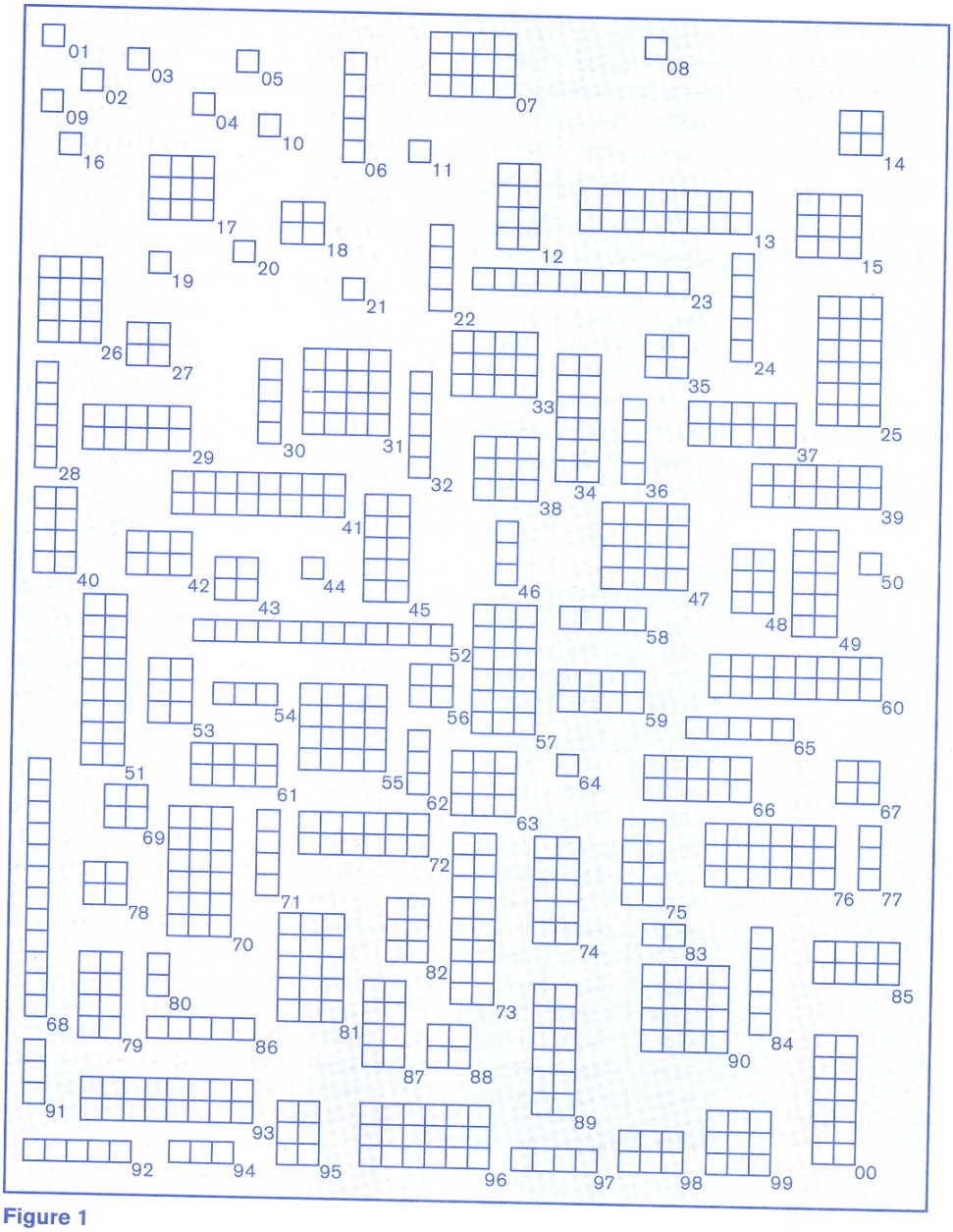
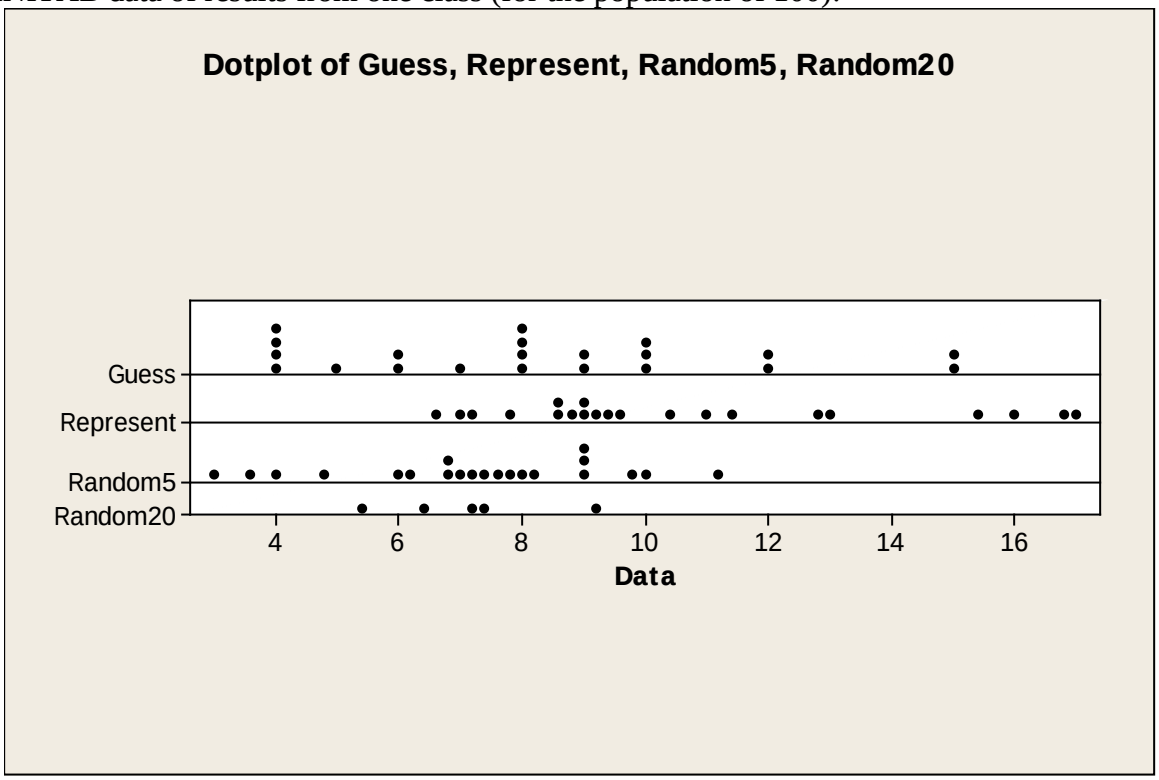


Figure 1

Line	TABLE OF RANDOM DIGITS								
1	79690	86718	33541	82828	12426	51929	14259	60592	55815
2	60945	95508	52743	53401	51154	20627	15127	11701	79304
3	16159	34564	59373	90106	41635	51518	59939	29712	98439
4	54175	85782	37363	12754	93588	92459	40206	42741	24366
5	73157	37067	37005	14058	87126	91385	78266	53500	39901
6	31714	89009	62032	16853	47711	75435	06914	09081	45890

(note: if you get to the end of a line, continue down on the next line)

Sample MINITAB data of results from one class (for the population of 100):



One-Sample T: Guess, Represent, Random5, Random20

```
Test of mu = 7.42 vs not = 7.42


Variable    N    Mean  StDev  SE Mean       95% CI          T      P
Guess      21   8.286  3.349    0.731  (6.761,  9.810)   1.18  0.250
Represent  21  10.690  3.246    0.708  (9.213, 12.168)   4.62  0.000
Random5    21   7.252  2.137    0.466  (6.280,  8.225)  -0.36  0.723
Random20    5   7.124  1.362    0.609  (5.433,  8.815)  -0.49  0.652
```

Summary of Random Rectangles activity:

- Typically, guesses have the greatest variance
- Typically, representative groups are biased, but more precise
- Typically, random samples are unbiased
- Random20 is unbiased but <u>always</u> more precise than random5
- You have to be able to deal with random variation from “typical” occurrences “on the fly” with real data
- The random5 and random20 are not independent tests of each other, since we didn’t choose a new set of 20 observations.

Sampling Stars from the Sky (Random Selection Activity 2, different sampling techniques)

From Petocz (1990), Sampling Space: Practical Experiments for Teaching Sampling (http://www.stat.auckland.ac.nz/~iase/publications/18/BOOK1/A4-9.pdf accessed 11-11-11 at 12:48)

"Star light, star bright, first star I see tonight," …but the first star of how many? How many stars are there in the sky? The figure on the next page is a star map and is assumed to represent a grid of stars observed in the sky at a given time. While you could just count the stars, let's employ statistics to help us out with the task. Use a random sample to determine how many stars there are in the sky. Please answer the following to accomplish this task:

Consider a convenience sample of the first five squares (starting in the upper left hand corner and going across).

a) How many stars does each square have?
b) What is the total number of stars in your sample?
c) Multiply this number by 100/5 (or 20) to estimate the total number of stars in the sky.

Consider a systematic sample of five squares:

d) What is the value of k?
e) Randomly select the first square.
f) Now, choose every k^{th} square from there so a total of five squares are obtained. Which squares did you choose?
g) How many stars does each square have?
h) What is the total number of stars in your sample?
i) Using this number, estimate the total number of stars in the sky.

Consider taking a random sample of squares:

j) Identify five squares at random. How did you choose them? Which squares did you choose?
k) How many stars does each square have?
l) What is the total number of stars in your sample?
m) Using this number, estimate the total number of stars in the sky.

Lastly, consider a stratified random sample of five squares, where the strata are "dense" vs. "sparse" squares (by visual examination only). For simplicity sake, identify 40 squares that look dense (lots of stars) and 60 squares that appear to be sparse (not as many squares).

n) List the 40 squares that you considered to be the most dense.
o) Use Excel to choose dense squares for your sample. Which squares did you choose?
p) How many stars does each square have?
q) Use Excel to choose sparse squares for your sample. Which squares did you choose?
r) How many stars does each square have?
s) What is the total number of stars in your sample?
t) Using this number, estimate the total number of stars in the sky.

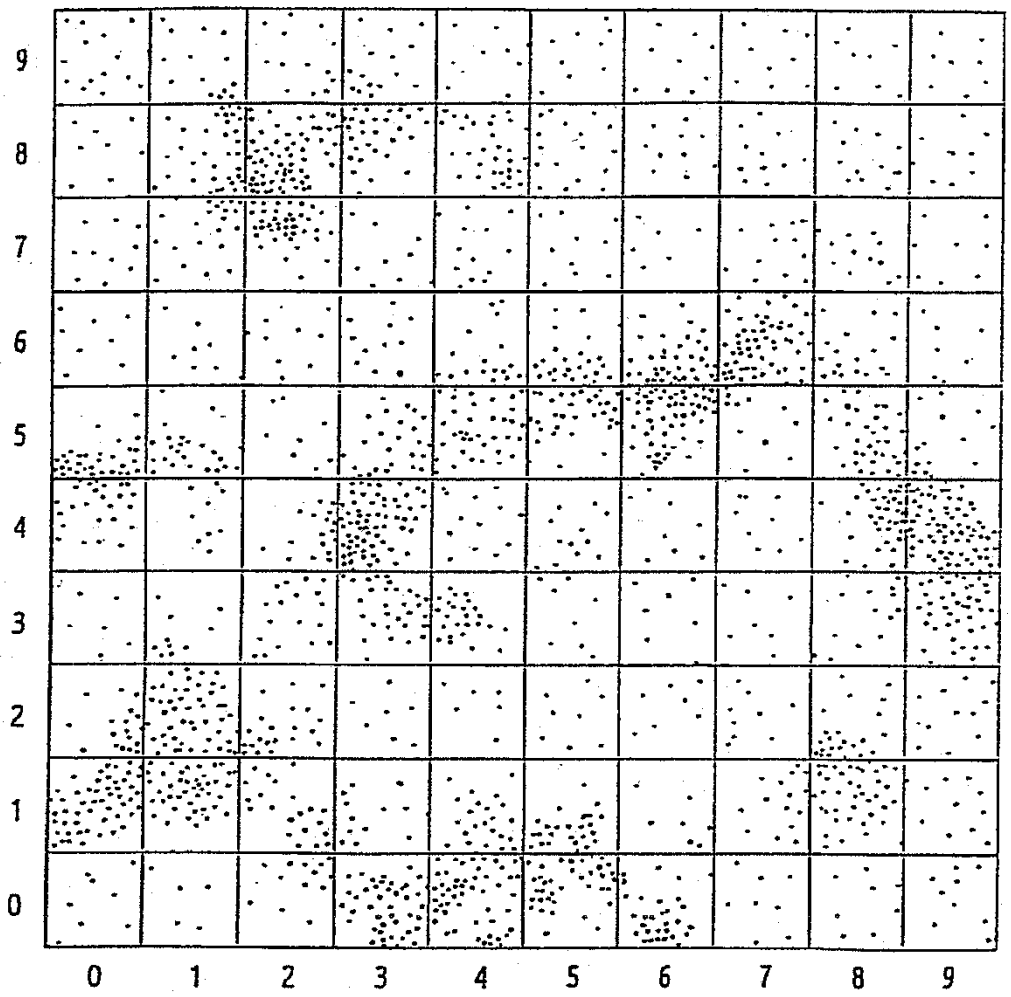
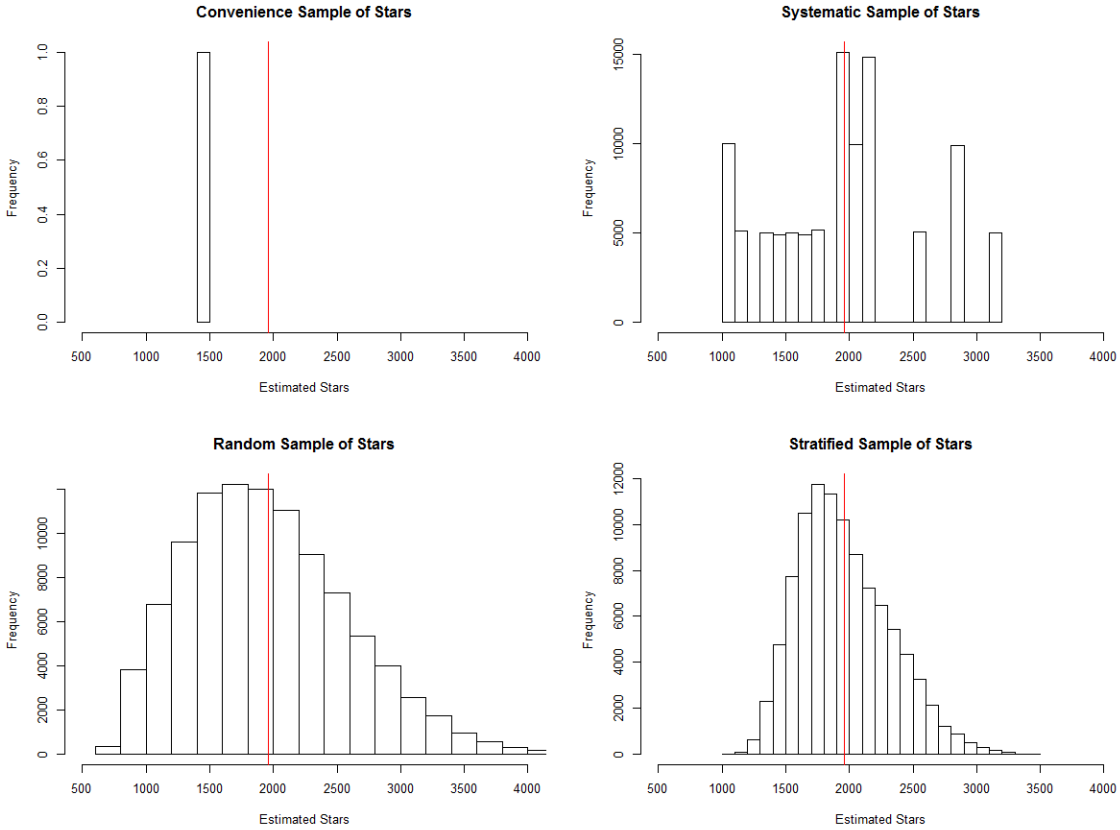


FIGURE 1
Sample space star map

Sampling stars in the sky – simulation results
Note: This is an example where simulation is both more readily available to students and easier to understand



Fighting Cancer with Raspberries (Random Assignment Activity 1)

Fighting Cancer with raspberries: demonstrating the value of random assignment

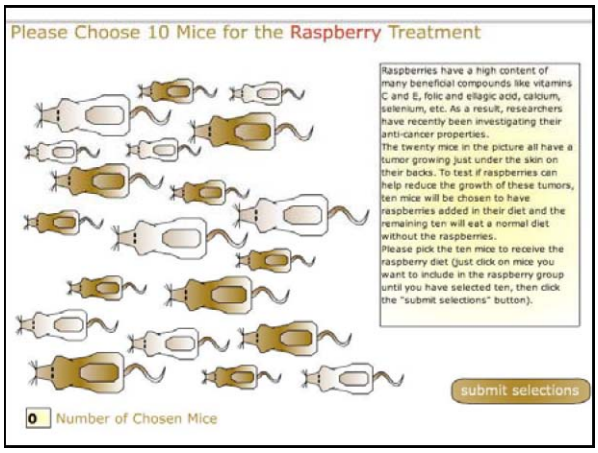
A Flash applet at

www.causeweb.org/mouse_experiment

- Programmed by John Gladden & Kythrie Silva
- Content by Dennis Pearl & Tom Santner
- Funded by Ohio Board of Regents

Learning Objectives

- Haphazard ≠ Random
- Random assignment eliminates bias
- It leads to a consistent pattern of results when repeated
- And thus makes the question of statistical significance interesting and easy to answer



Starting the activity

- Give background
- Give students 10 seconds to pick ten mice for raspberry group

Try it out! Pick your ten mice but do not hit the submit button

…and be careful not to close the two windows associated with the webinar.

Make predictions

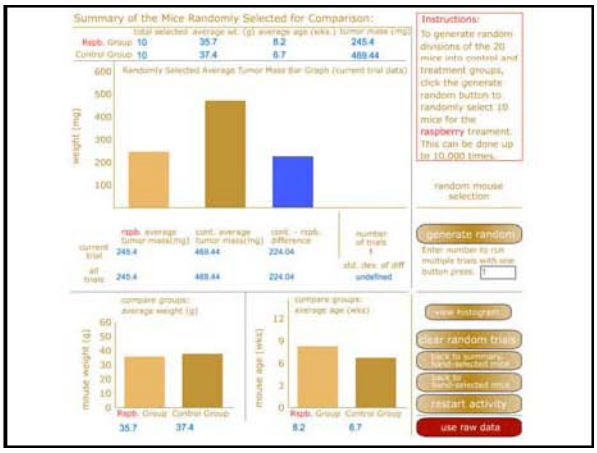
- Ask students if they showed any favoritism in their picks.
- Have students predict how a random sample will behave w.r.t. variables besides treatment such as weight or age of mice.

Compare Predictions to results from haphazard selection

- Which group had larger weights?
- Which group had older mice?
- How did your results differ from your neighbor's?
- How did the class results differ from predicted?
  - (note your results for avg. difference in tumor size)
- Make a new prediction:

Biology says that tumor growth will be faster in older (hence bigger) animals. How will this affect the results?

Let the applet pick the mice randomly and check the results

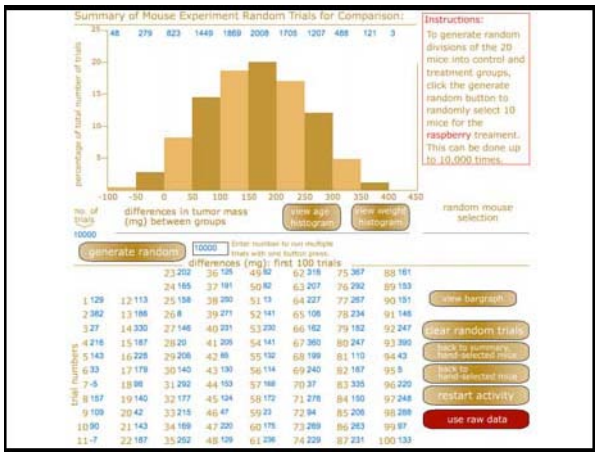


Let the applet pick the mice randomly and check the results

- Which group had larger weights?
- Which group had older mice?
- How did your results differ from your neighbor's?
- How did the class results differ from predicted?
- How did your results for tumor size change?

(Summarize results for individual students and for class as a whole)

Let the applet pick the mice randomly 10,000 times and check the results



Let the applet pick the mice randomly 10,000 times and check the results

- What kinds of differences in weights occurred?
- What kinds of differences in ages occurred?
- What kinds of differences in tumor size occurred?

Compare with graph for ages and weights.

- How did your results differ from your neighbor's?

(Summarize results for individual students and for class as a whole)

Summary Discussion

- Was it harder to see the Effect of Rasapberries when we selected the mice by hand? Why?
  - What "caused" difference in tumor sizes between groups when mice were hand selected?
    (Bias toward older mice
    Luck of the draw
    Effect of raspberries)
  - What "caused" difference in tumor sizes between groups when mice were randomly selected?
    (Luck of the draw
    Effect of raspberries)

Value of applet

- Time saving over classic activities
  - Stones
  - Driftwood
  - Two heads are better than one
- Requires little background
  - Useful for different audiences
  - Useful at different points in the syllabus

Value of applet

- Illustrates "subtle" points about bias
  - Bias can result without intentional favoritism
  - Bias reflects a tendency not a deterministic move
- Allowing individual student input challenges them to beat random allocation
- Comparison to random assignment is direct
- Can quickly compare distributions for both confounders and response variables

Supplemental Material

- Student handout
- Teacher's lesson plan
- Assessment items

Sampling Distributions Activities

Sampling Distribution Activity 1 [included below] – What is a Sampling Distribution?
Students generate sampling distributions based on binomial, uniform, and geometric distributions using coins and dice and software.

Sampling Distribution Activity 2 - Random Pennies - Understanding sampling distributions
Examine the impact of sample size and underlying distribution on shape and precision.

Related Activity - 1962 pennies
Over time, some pennies have been minted with a rim that causes them to land on the heads side when placed on their edge (and "pounded") or spun, substantially less than 50% of the time. However, when flipped the coins land heads about 50% of the time. Find a point estimate for the proportion of times the penny lands heads when it is spun. (Data suggest that the point estimate is approximately 15% for early 60s and 40% for pennies from the late 80s and early 90s).

Sampling Distribution Activity 3 – Happyville
by Kevin Robinson, Millersville University, is an activity similar to random rectangles.
A webinar can be found here: http://www.causeweb.org/webinar/activity/2011-07/
The web site which includes a quiz on sampling distributions is here:
http://www.millersville.edu/~krobinson/happyville

Sampling Distribution Activity 3 – German Tank Problem
German Tanks activity (versions of this activity can be found Mathematics Teacher, 1999 vol. 92 issue 8 and *Activity Based Statistics*) can be used to help students understand that all good estimators are not the mean. A CAUSEweb Webinar was presented by Diane Evans on the German Tank problem. A link to the webinar can be found at http://www.causeweb.org/webinar/activity/2009-09/

Sampling Distribution Activity 1 – What is a Sampling Distribution?

Use a coin and die to begin to develop thinking about sampling distributions and the central limit theorem

Sampling Distributions Worksheet

Flip your coin **one** time, record the proportion of heads: _____ (0 or 1)
Flip your coin **twice**, record the proportion of heads: _____ (0/.5/1)
Flip your coin **ten** times, record the proportion of heads: _____ (0-1)

Roll your die **one** time, record the result ("average"): _____ (1-6)
Roll your die **two** times, record the average: _____ (1-6)
Roll your die **five** times, record average: _____ (1-6)

Roll your die until your first 5 or 6. How many rolls? _____ (1+)
Roll your die until your first 5 or 6, **two times**!
rolls first time? _____ (1+) # rolls second time? _____ (1+)
Average of these two numbers: _____ (1+)

Graph the student responses. Then use technology to generate a large number of samples.

David Little's Plinko applet is also quite useful here. It can also be downloaded, so you don't need to use the internet. http://www.math.psu.edu/dlittle/java/probability/plinko/index.html

Another useful applet is at http://onlinestatbook.com/stat_sim/sampling_dist/
Question for thought – ask students to create the distribution that would take the longest time to converge to a normal distribution when examining the distribution of the sample mean

Sampling Distribution Activity 2 - Random Pennies

adapted from *Activity Based Statistics: Student Guide* see references

Consider the age of pennies in circulation. We expect more new pennies and fewer older pennies.

Someone has collected a large number of pennies from circulation (in 1999?), recorded the mint date and from that and found the age of those pennies. Sketch the graph of the distribution of this population estimating and marking the mean and standard deviation.

Now you take a sample of 50 from this population. Sketch the graph of your sample estimating and marking the mean and standard deviation.

We will use a computer to find the sampling distribution of sample mean age of pennies in circulation. Sketch of graph of this sampling distribution estimating and marking the mean and standard deviation**.**

Use this applet for the demonstration of the sampling distribution:
http://www.rossmanchance.com/applets/OneSample.html (click on Pennies)

Multivariate Thinking

From the GAISE Report:
We live in a complex world in which the answer to a question often depends on many factors. Students will encounter such situations within their own fields of study and everyday lives. We must prepare our students to answer challenging questions that require them to investigate and explore relationships among many variables. Doing so will help them to appreciate the value of statistical thinking and methods.

But…

- Isn't this beyond what we do in the introductory course? Don't students need more sophisticated methods to deal with multivariable situations???
- Well, not really. Much can be done with simple graphical methods and descriptive statistics.
- Students can explore multivariable data to come to understand that things are not always as they seem in a univariate or bivariate world. This is an important understanding.
- And it can be fun!
- Provide students with opportunities to investigate and explore relationships among many variables.
- Recognize that in the intro stat course, can't always cover this in depth, but the 2016 GAISE report says it is important to expose students to these ideas in the first course.

*****Pedagogical note:** Give students many examples of when confounders don't exist as well, lest you leave them looking for the confounder in all situations, when often they don't exist. Subject matter expertise is key here.

Activity – Identifying Confounding Variables

Definition: A **confounding variable** is a variable that is related to the explanatory variable and the response variable and changes the observed relationship between the two. A potential confounder is any variable related to the explanatory and response variables. Non-experiments are very susceptible to confounders.

Kahneman, in *Thinking: Fast and Slow*, states that human nature leads us to quick conclusions, simple explanation, and causal relationship. This is counter to critical thinking skills that we teach.

Confounding Example
Scenario: People who attend church have lower blood pressure than those who stayed home and watched religious services on television.

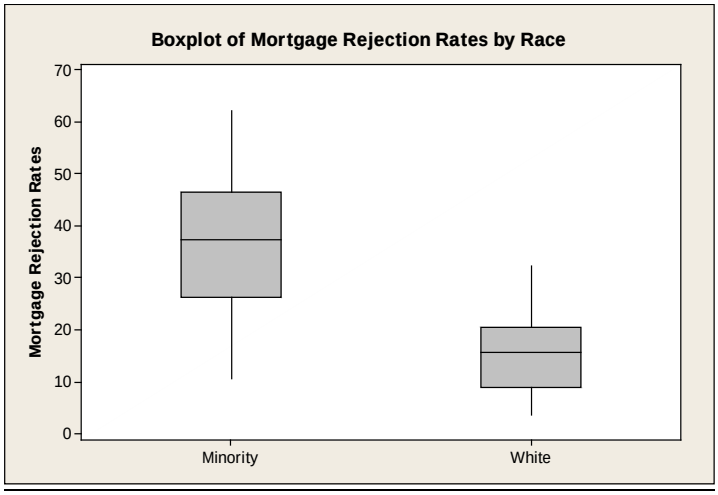
Explanation: Maybe those who attended church had lower blood pressure as a result of a strong social support network rather than from attending church regularly. If this is the case, "amount of social support" is the confounding variable. Two conditions must be met for "amount of social support" to qualify as a confounding variable.

- Amount of social support (confounding variable) affects blood pressure (response variable)
- Amount of social support (confounding variable) is related to attending church regularly (the explanatory variable) because the same people who attend church regularly are likely to have a strong social support network.

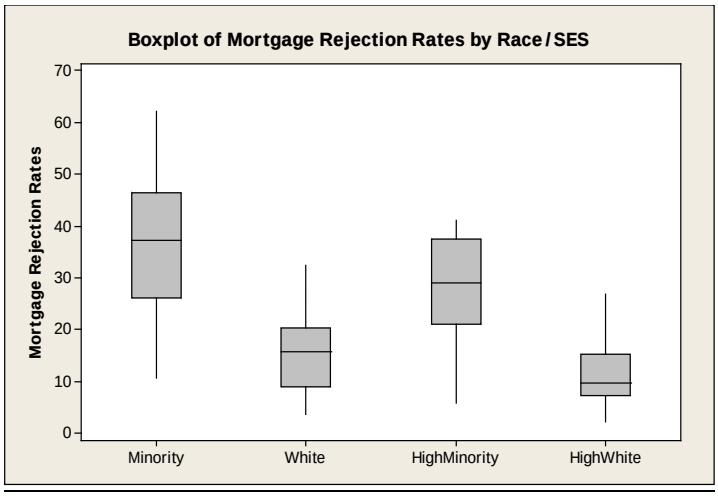
Exercise: For each of the following, identify the confounder:
As crime rates increase, so does ice cream sales, so let's stop selling ice cream!
The number of speeding tickets that a student has is less than that of his/her parents.
Typically, students do worse on exam problems that they spend a lot of time on.
Strong positive relationship between # of TVs per house and life expectancy (by country)
The mortality rate in the US is higher than most other North/South American countries
The more children a woman has, the lower her risk of breast cancer.
Do storks really bring babies?
Bonjour Paris L'ecole (https://www.youtube.com/watch?v=tS55WeYwPfA)

Answers: Temp, Years driving, Difficulty of Problem, Income, Age Distribution, Breast Feeding, Hot Tin Roof

Activity - Mortgage Rejection Rates (an example of multivariate thinking and one method of dealing with confounders – stratification):



List the reasons why these results might differ?



SAT scores and Teacher Salary

Data were gathered at the state level on mean SAT scores and mean teacher salary, among other variables. The full description of the data are below, which can be found at http://www.amstat.org/publications/jse/secure/v7n2/datasets.guber.cfm.
Note – the JSE data archive is a good resource for worked through examples.

```
NAME:  Getting What You Pay For: The Debate Over Equity in Public
       School Expenditures
TYPE:  Census
SIZE:  50 observations, 8 variables

DESCRIPTIVE ABSTRACT:
This dataset contains variables that address the relationship between
public school expenditures and academic performance, as measured by the
SAT.

SOURCE:
The variables in this dataset, all aggregated to the state level, were
extracted from the 1997 _Digest of Education Statistics_, an annual
publication of the U.S. Department of Education.  Data from a number of
different tables were downloaded from the National Center for Education
Statistics (NCES) website (Available at:
http://nces01.ed.gov/pubs/digest97/index.html) and merged into a single
data file.

VARIABLE DESCRIPTIONS:
Columns
 1 - 16     Name of state (in quotation marks)
18 - 22     Current expenditure per pupil in average daily attendance
            in public elementary and secondary schools, 1994-95
            (in thousands of dollars)
24 - 27     Average pupil/teacher ratio in public elementary and
            secondary schools, Fall 1994
29 - 34     Estimated average annual salary of teachers in public
            elementary and secondary schools, 1994-95 (in thousands of
            dollars)
36 - 37     Percentage of all eligible students taking the SAT, 1994-95
39 - 41     Average verbal SAT score, 1994-95
43 - 45     Average math SAT score, 1994-95
47 - 50     Average total score on the SAT, 1994-95

The data are aligned and delimited by blanks.  There are no missing
values.

SPECIAL NOTES:

While an initial scatterplot shows that SAT performance is lower, on
average, in high-spending states than in low-spending states, this
statistical relationship is misleading because of an omitted variable.
Once the percentage of students taking the exam is controlled for, the
relationship between spending and performance reverses to become both
positive and statistically significant.
```

THE STORY BEHIND THE DATA:
I compiled this dataset in response to recent controversy over equity in public school expenditures. While some argue that the prevailing system of financing local schools is unfair, aggregate data reported in the new media seem to suggest, paradoxically, that school spending and academic performance are statistically unrelated. My goal was to create a workable dataset that would introduce students to this continuing debate and allow them the opportunity to build their own conclusions upon a base of solid statistical reasoning.

Additional information about this dataset can be found in the "Datasets and Stories" article "Getting What You Pay For: The Debate Over Equity in Public School Expenditures" in the _Journal of Statistics Education_ (Guber 1999).

PEDAGOGICAL NOTES:

This exercise is ideally suited for classroom discussion in an elementary statistics or research methods course. It allows students to gain experience and confidence in using basic statistical techniques such as scatterplots, measures of association and dispersion, and linear and multiple regression, while giving them the opportunity to examine an issue of substantive importance.

REFERENCES:

Hawkins, J. A. (1994), "Money Matters! But What Difference Does It Make to Schools and Students?" _NASSP Bulletin_, 78, 19-26.

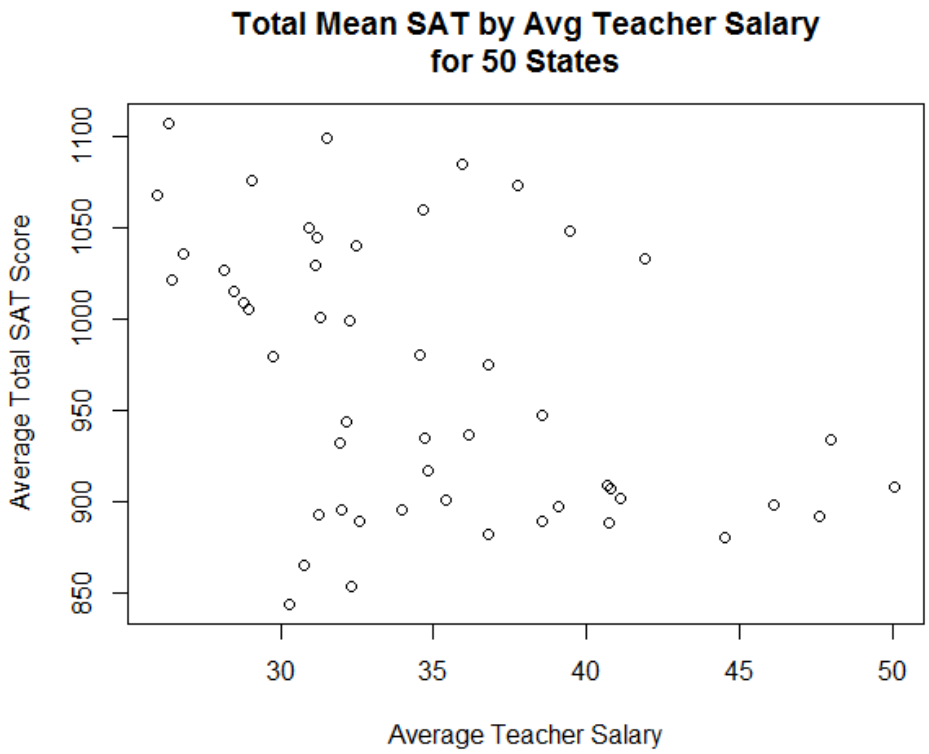
Powell, B., and L. C. Steelman (1984), "Variations in State SAT Performance: Meaningful or Misleading?" _Harvard Educational Review_, 389-410.

Powell, B., and L. C. Steelman (1996), "Bewitched, Bothered and Bewildering: The Use and Misuse of State SAT and ACT Scores," _Harvard Educational Review_, 27-59.

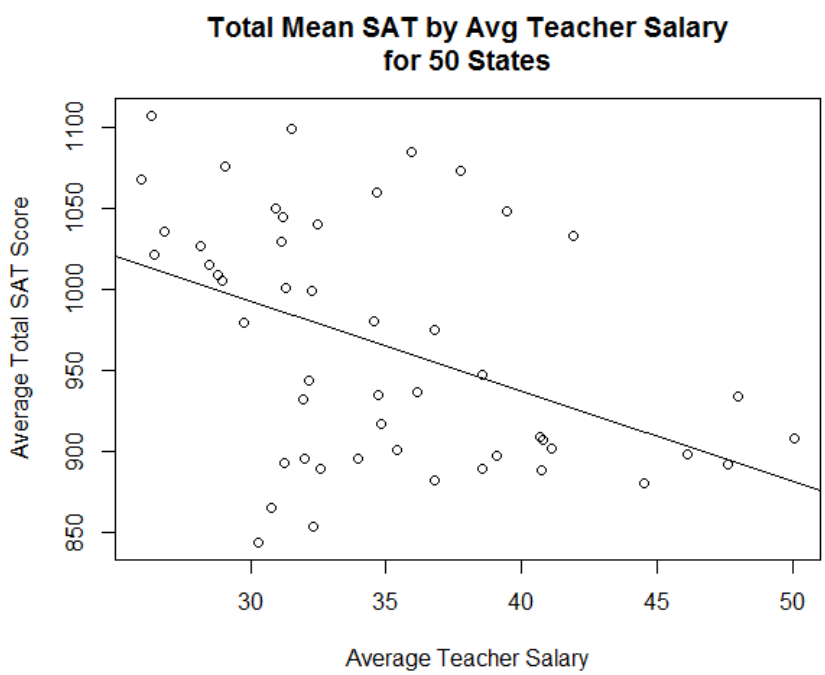
Will, G. F. (September 12, 1993), "Meaningless Money Factor," _The Washington Post_, C7.

SUBMITTED BY:
Deborah Lynn Guber
Assistant Professor
Department of Political Science
The University of Vermont
P.O. Box 54110
Burlington, VT 05405-4110
dguber@zoo.uvm.edu

First, let's examine the linear regression of mean total SAT by mean teacher salary



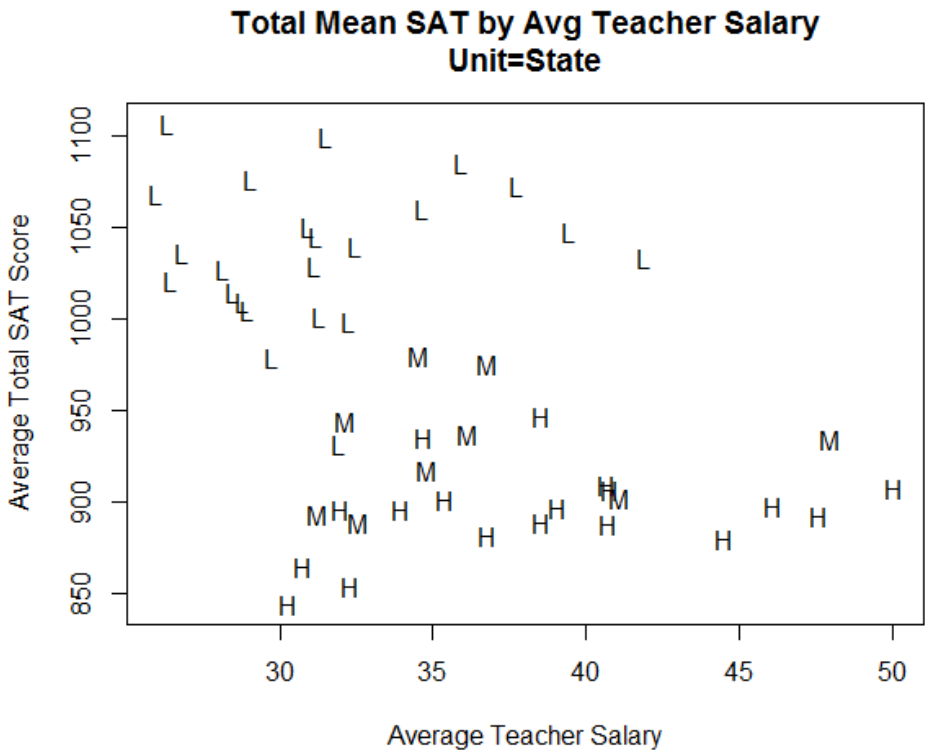
Describe the relationship that you see…



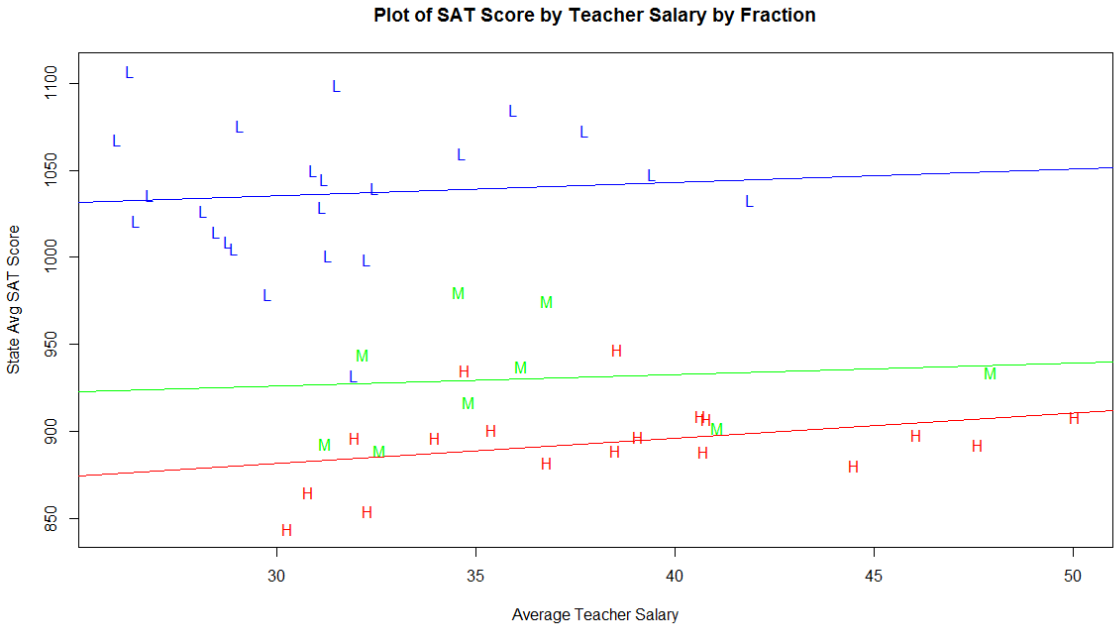
What's going on here? The more you pay teachers, the lower the student's SAT scores? Exercise: identify a potential confounder that exists…

Let's consider the potential confounder by examining these data separately by the fraction of students that actually took the SAT test. For simplicity, I split the data into fraction of students that took the test into three groups:

4% - 21% ("L"), 23% - 48% ("M"), and 51% - 81% ("H")



Look at the individual regressions by fraction taking the SAT (by groups):



Discrimination in Services to the Disabled (Example from the *Journal of Statistics Education*)

Data Set: Random sample of 1,000 clients of California Department of Developmental Services

Variables:

1) Annual expenditure for support to individual and family
2) Gender
3) Ethnicity
4) Age Cohort (based on amount of financial support typically required during a particular life phase) (0 – 5 years, 6 – 12 years, 13 – 17 years, 18 – 21 years, 22 – 50 years, over 50 years)

Is there discrimination by gender or by ethnicity?

Gender	**Average of Expenditures ($)**
Female	$ 18,130
Male	$ 18,001
All Consumers	**$ 18,066**

Table 2. Average Expenditures by Gender

Ethnicity of Consumers	**Average of Expenditures ($)**
American Indian	$ 36,438
Asian	$ 18,392
Black	$ 20,885
Hispanic	$ 11,066
Multi Race	$ 4,457
Native Hawaiian	$ 42,782
Other	$ 3,317
White non-Hispanic	$ 24,698
All Consumers	**$ 18,066**

Table 1. Average Expenditures by Ethnicity

Sample sizes were very small for most groups, so focus on comparing two groups—Hispanic and White non-Hispanic.

Ethnicity	**Average of Expenditures ($)**	**% of Consumers**
Hispanic	$ 11,066	38%
White non-Hispanic	$ 24,698	40%

Table 5.1 Average Expenditures and # of Consumers by Ethnicity

Evidence of discrimination? What else might explain difference in mean expenditure?

Age Cohort	**Hispanic** (avg. of expenditures)	**White non-Hispanic** (avg. of expenditures)
0 – 5	$ 1,393	$1,367
6-12	$ 2,312	$2,052
13-17	$ 3,955	$3,904
18-21	$ 9,960	$10,133
22-50	$ 40,924	$40,188
51 +	$ 55,585	$52,670
All Consumers	**$11,066**	**$24,698**

Table 6. Average Expenditures by Ethnicity and Age Cohort

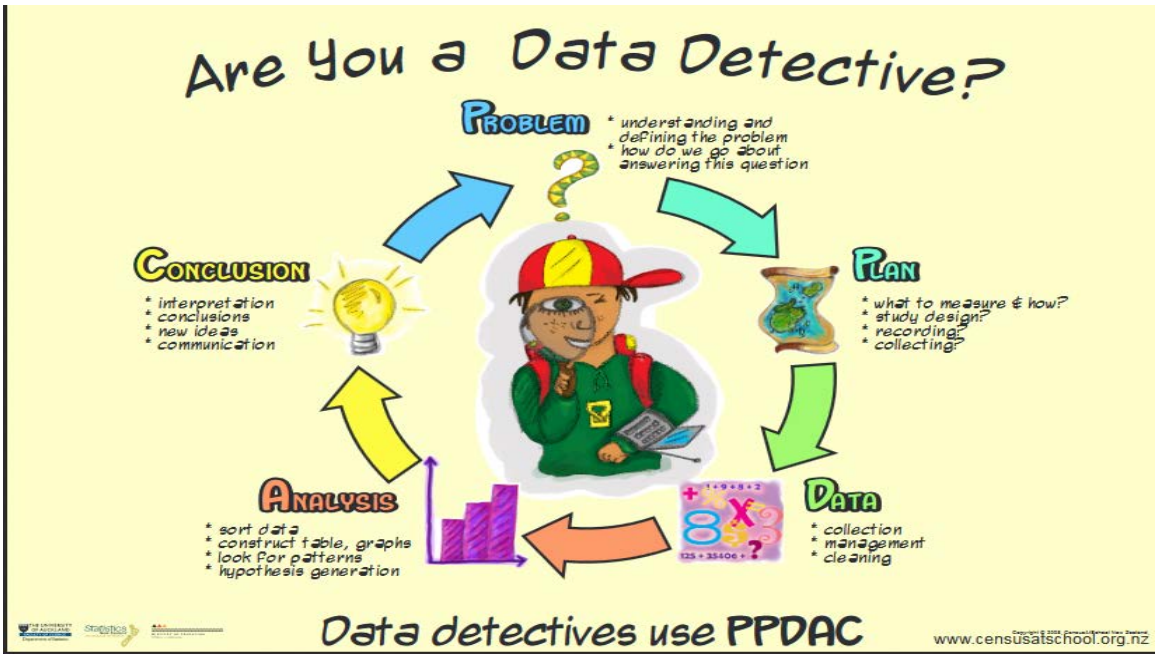
Investigative Process

From GAISE Report:
Students should not leave their introductory statistics course with the mistaken impression that statistics consists of an unrelated collection of formulas and methods. Rather, students should understand that statistics is a problem-solving and decision-making *process* that is fundamental to scientific inquiry and essential for making sound decisions.

The investigative process begins with a question that can be translated into one or more statistical questions – questions that can be investigated using data. While many questions do not have simple yes or no answers, knowing how to obtain or generate data that are relevant to the goals of a study is crucial to providing useful information that supports decision-making in the sciences, business, healthcare, law, the humanities, etc. Understanding and applying the principles of representative sampling for an observational study or designing an experiment is critical to the investigative process. Understanding and, when possible, controlling for the impact of other variables is important.

Once high quality data have been collected, meaningful graphs and numerical summaries (generally created using technology) shed light on the question under study. These summaries help to identify statistical inference procedures that are appropriate to the question. The results of the data analysis, and any limitations, need to be clearly communicated.

What is involved in the investigations process that you might not see in example problems?



Exercise on Bias

Information for group 1) Do you think the population of the Philippines is above 20 million?

Information for group 2) Do you think the population of the Philippines is above 120 million?

Question for both) What is your best guess for the population of the Philippines?

Task) Record and review these, by group

Summary) This is an example of bias!

Mechanics of a problem:

Based on the following data, conduct a two-sample t-test to determine whether elderly people have impaired balance compared to young people (higher numbers indicate worse balance)?

	Elderly	Young
Sample mean	26.3	18.1
Sample st. dev.	9.8	4.1
n	9	8

Solution:

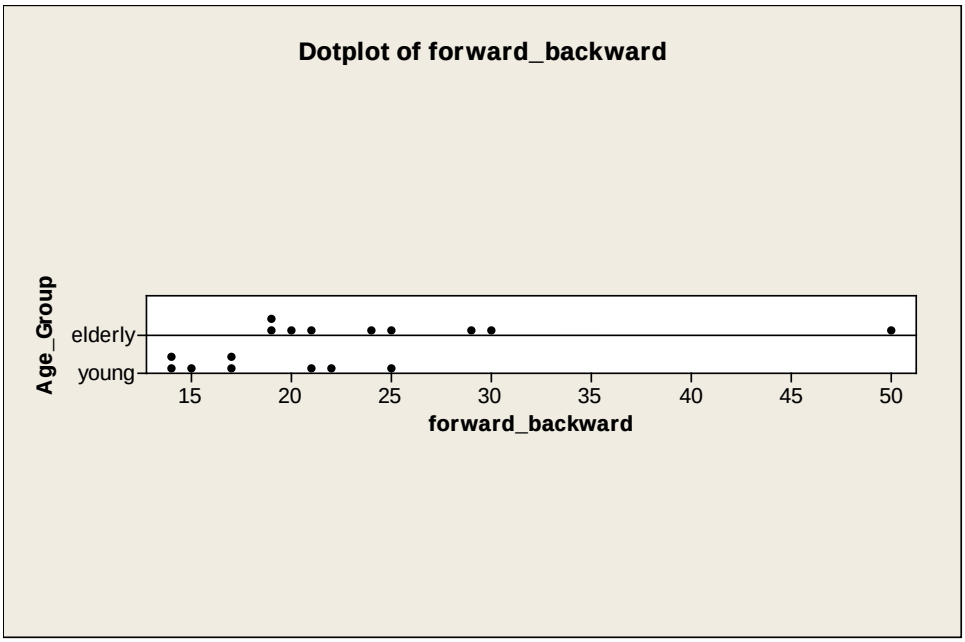
H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$

t_7=(26.3-18.1) / [sqrt((9.8^2+4.1^2)/(9+8-2))/(1/9+1/8)] = 2.294

p-value = 0.028

Reject H_0, since $p<\alpha$, so elderly and balance impaired

Investigative Process for a problem:
Is it harder to maintain your balance while you are concentrating? Nine elderly and eight young men were subjects in this experiment. Each subject stood barefoot on a "force platform" and was asked to maintain a stable upright position and to react as quickly as possible to an unpredictable noise by pressing a hand held button. The noise came randomly and the subject concentrated on reacting as quickly as possible. The platform automatically measured how much each subject swayed in millimeters in both the forward/backward and the side-to-side directions.



Descriptive Statistics: forward_backward

Variable	Age_Group	N	Mean	StDev	Variance	Minimum	Q1	Median
forward_backward	elderly	9	26.33	9.77	95.50	19.00	19.50	24.00
	young	8	18.13	4.09	16.70	14.00	14.25	17.00

Variable	Age_Group	Q3	Maximum	IQR
forward_backward	elderly	29.50	50.00	10.00
	young	21.75	25.00	7.50

1) Identify the research goal
2) Was this an observational or experimental study?
3) Identify the observational unit
4) Identify the variables. For each variable, identify the type of variable and role in the study (explanatory, response, or other).
5) Comment on the study design
6) Identify potential confounders
7) Identify potential sources of bias
8) Describe the distribution of the two variables
9) Which analytic method is appropriate? Explain why.
10) Report your results
11) How could you have designed the study differently?

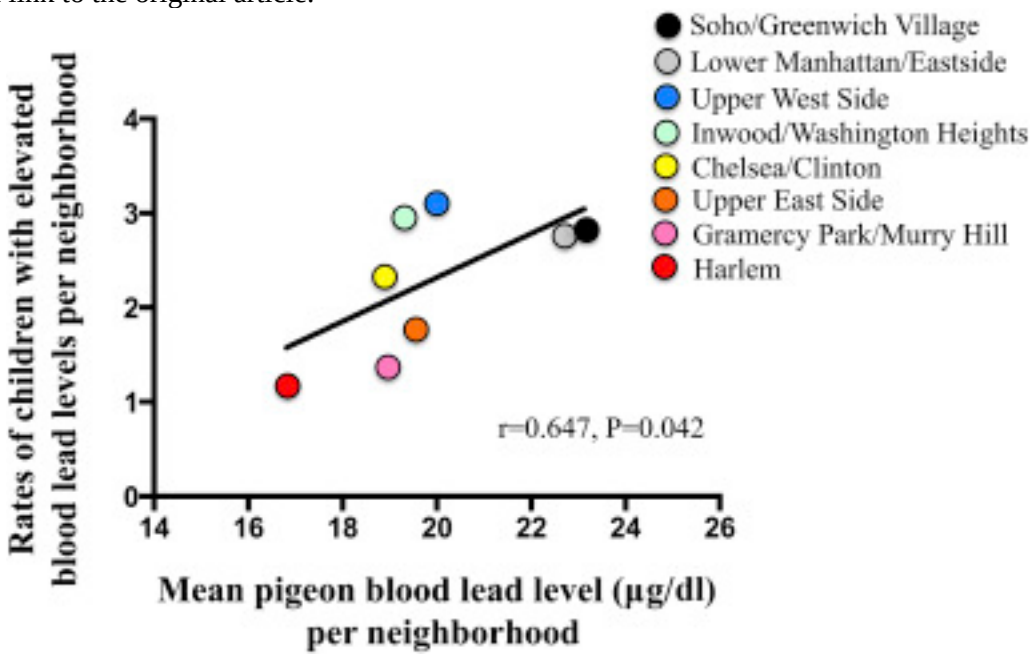
Consider alternative analytic methods. What to do when the assumptions are not met? What happens when you exclude the one outlier for elderly?

Investigative Process - reading vs. designing a study:

Reading a Study	**Designing a Study**
Identify research question/goal	State research question/goal
Observational data vs. Experiment	Decide study design Observational vs. Experiment Type of design
Identify observational unit	Decide on observational unit
Identify variables Name Type of variable (categ/ord/num) Explanatory / Response / Other	Decide on explanatory, response, and other variables
Comment on study design random sampling? assignment?	
Identify potential confounders	Collect data on potential confounders (block?)
Identify potential bias from data collection or processing	Identify potential bias and methods to reduce them
Comment on data visualization (if any)	Use data visualization
Identify analytic method – was it appropriate? Were assumptions met?	Decide on analytic method, check necessary conditions
Comment on reporting	Report results

Studies

- Hans Rosling videos: 200 countries, 200 years, 4 minutes
  https://www.youtube.com/watch?v=jbkSRLYSojo
- Esther Duflo: Social Experiments to Fight Poverty
  http://www.ted.com/talks/esther_duflo_social_experiments_to_fight_poverty
- Who drinks more sugar-sweetened beverages, men or women? In a recent report from 2012 (West, et. al., *Obesity*, 2012), 175 women and 90 men were surveyed. They reported that more women than men (74% vs. 61%) drank sugar-sweetened beverages daily, with the most common one being soda. A two-sample z-test for proportions was done and a two-sided p-value of 0.027 was calculated. Thus, they concluded that women drink more sugar-sweetened beverages daily compared to men.
- Ordering lunch at least an hour before you eat could help you cut calories and avoid unhealthy impulse choices. (http://www.futurity.org/food-orders-calories-1207162-2/). The first study was a secondary data analysis of over 1,000 orders that could be placed any time after 7 am to be picked up between 11 am and 2 pm. Across all three studies, the researchers noted that meals with higher calorie content were ordered and consumed when there were shorter (or no) waiting periods between ordering and eating.
- City pigeons and kids may share lead levels. A new study of pigeons in New York City shows that levels of lead in the birds track with neighborhoods where children show high levels of lead exposure. The summary of the article can be found at http://www.futurity.org/pigeons-lead-1206822-2/, including a link to the original article.



Design a study

1) Tangrams. Using the online tangram site, design a study.
2) Pennies. Consider two ways of generating data from a penny – flipping it or standing it on its edge and banging the table. You will do this experiment, recording the proportion of heads, and then perform a two sample test for proportions to determine whether the proportion of heads is the same using both methods. (Add a second factor of 1962 pennies vs. modern pennies)

Projects in Intro Stat

Confucius says “I hear and I forget. I see and I remember. I do and I understand.”

Why projects?

- Projects I did as an undergrad were among the most formative experiences I had as a statistics major
- Teach students the investigative process (GAISE)
- Understand the *messiness* and *subjectivity* of data…
  - Variable definition / population of interest (ex – likely voter?)
  - Collection – nonresponse, question choice (or order), measurement
  - Cleaning – data errors, outliers, merging data files
  - Processing – creating variables, missing data
- Allows learning of new statistical techniques
- Teach the 90/10 concept of data prep/analysis
- Teach the idea of “building up” your analysis
- Get away from the “one analysis and done” fallacy
- Offer intrinsic motivation due to
  - Application
  - Ownership of work
  - Novel discovery

About Projects

- Six main aspects (not all required)
  - Goal, design study, collect data, process data, analysis, report
- Joint project decisions for the class?
  - None, goal, goal + design, goal + design + data
- Project groups vs. individuals
  - Self-selected vs. assigned groups

Project Timeframe (Fall semester)

- Proposal (Oct)
  - Group members and name
  - Goal of study
  - Variables (3+, 1+ quant, 1+ categ, explan/response)
  - Data collection process
  - Need topic? Meet to talk about hobbies, etc.
- Group meeting with professor (Oct/Nov)
- Submission of Data (before Thanksgiving)
- Project presentation (last week of semester)
- Project report (after feedback from students/me)

Project Presentations/Report

- Goal and hypotheses
- Data collection process
- Graphical and numerical analyses
- Build up – univariate, bivariate, multivariate
- Summary
- Limitations – bias
- How could I have improved the project? What new research questions come up?

Project Assessment

- Individual assessment of group project

All students get the same grade?
All students must present?
Participation grade (4 students = 400 points)

- Peer Evaluation

In one or two brief sentences, summarize the main findings of the presentation
What did the presenter do well?
What could be improved in the presentation or what was unclear?
What other suggestions do you have for the presenter?

- Rubric

Project Rubric

- **Organization (20%) - How well can you understand the sequence of information?**
- 1 – The presentation was disorganized and hard to follow. Transitions were difficult to understand and I wasn't sure what was going on for much of the presentation.
- 3 – At times, I wasn't sure of the logical progression, but the presentation had an order and flow that was I could follow.
- 5 – The presentation was very clear and easy to follow.
- **Statistics (50%) - How well does the presenter demonstrate proficiency in use of the correct statistical tools and measures?**
- 1 – The presenter has minimal, if any, grasp on when to use certain statistical tests or frequently used an incorrect test. They didn't justify their choice of which test was appropriate.
- 3 – The presenter used tests appropriately, but didn't verify assumptions or demonstrate why each one was appropriate. Some interpretations were incorrect or confusing.
- 5 – The presenter has a clear grasp of when each test should be used, how to implement it, and how to interpret the results.
- **Graphics (10%) - Were the graphics helpful or distracting?**
- 1 – The graphs were misleading, mislabeled, or incomplete.
- 3 – The graphs were helpful in understanding the data and results, but distractions persist, including mislabeling or cluttered presentations.
- 5 – The graphs were very helpful in understanding the data and results.
- **Mechanics (10%)**
- 1 – The presentation had numerous grammatical and spelling mistakes that were distracting.
- 3 – The presentation was mostly clear and accurate, with minor errors.
- 5 – The presentation contained no grammatical or spelling errors.

- **Overall Presentation (10%) - Did the presenter mumble or did they speak clearly?**
- 1 – I couldn’t understand most of what was being presented. I was bored by the presentation.
- 3 – I followed most of the talk, but sections were unclear.
- 5 – The presentation was clear and concise. I was excited by the presentation.

Project Resources

- Robin Lock
- Simply Statistics blog
- Duke University
- UC Berkeley Undergrad Research Group
- Austin CC – Mary Parker
- JSE Articles on Projects
    - Implementation of Discovery Projects in Statistics (2013)
    - The Effect of a Student-Designed Data Collection Project on Attitudes Toward Statistics (2008)
    - Teaching Students to Use Summary Statistics and Graphics to Clean and Analyze Data (2005)
    - Learning Statistics at the Farmers Market? A Comparison of Academic Service Learning and Case Studies in an Introductory Statistics Course (2010)
    - Studying student benefits of assigning a Service-Learning project compared to a traditional final project in a Business Statistics Class (2008)
    - Project-driven courses – Love, Dierker (Passion-driven, project-based), Hydorn (Service-learning), etc.

Example of Projects

- The accuracy of real estate price modeling on Zillow. What is it a function of?
- Voter turnout is a function of weather (temperature), whether an incumbent is running, the state of the economy, whether or not it is a presidential election year.
- Housing construction in a township is a function of real interest rates, population growth, tax advantages, increased costs in the city, etc.
- The average attendance per season at Phillies games is a function of how many games won last year, their league standing the previous year, number of new players, number of rainy days, cost-of-living.
- Text Message Use as a function of grade, age, gender, etc.
- Clutch QB performance as a function of time remaining, pressure/non-pressure, distance to endzone, points needed, etc.
- Does modality of instruction impact student learning?
- Examining nearly extinct languages (by country) and how it relates to development, other languages spoken, etc.
- Determining the characteristics and actions of students that lead to success in freshman calculus (resulted in report to Mathematical Sciences Department)
- Expected attendance at Big East Football games at Villanova or other data from Athletics
- Student health data – see email from SHS
- A volleyball team's win percentage is correlates with aces/set. Conferences with a wider variety in skill level produce teams with more variability in aces/set.
- Villanova juniors most prefer getting rid of a philosophy class in the core, followed by ACS, and mathematical sciences. These differ by gender and college.
- 5% of Villanova students get VEMed each year (much fewer than expected)
- GPA does not seem to be correlated with parents marital status, how often you call home, or whether parents track student's grades.
- Meteorologist weather prediction is more variable farther in future (5 days out vs. tomorrow)
- Students who use more alcohol exercise less.
- Villanova students estimate that 30% of all baseball players use steroids. 2% of all Villanova students use steroids.
- Athletes and non-athletes get the same amount of sleep, on average.
- Republicans do slightly more community service at Villanova compared to Democrats, but "others" do the most, by far.
- Villanova (basketball, football, and soccer) teams have a home field advantage.
- Half of all Villanova students are "single." Males tend to date "openly" or "non-exclusively" more than females.
- Students who smoke tend to have a lower GPA.
- Children of divorced parents don't go to school farther from home. Question - Is this different for males and females or based on when the parents got divorced?
- Is there response bias? Students found no difference in responses to questions whether they were asked via interview or anonymous survey. Comment – Need more sensitive question
- Students who spend more time on non-academic computer use had a lower GPA.

Example of Projects (cont'd)

- Student with no sexual partners last semester had a higher GPA than those who seldom (<1/wk), often (1-2/wk), or frequent (>2/wk) had sexual relations. The latter three categories showed no difference. The sample size was too small, though, to make any real inferences.
- Students drink more coffee during finals.
- Students usually get more than the recommended amount of cardiovascular exercise.
- Students with jobs had a higher mean stress level than students wthout jobs.
- The more marijuana students smoke, the lower their GPA.
- GPA was not associated with class time nor study time. Students who took more credits (class hours) did study more, though.
- There was an association found between political affiliation and major.
- Males and females both come late to class (at roughly equal rates).
- Boston University has 6% non-heterosexual females, 8% non-heterosexual males.
- Students with higher GPA are more satisfied with their college education.
- Women and men do not differ in their opposition to the war in Iraq.
- Villanova students who sleep less than 4 hours per day, have a lower GPA, on average. Those that sleep 6-8 hours per day have the highest GPA. (BU study showed no association)
- How religious you are is not associated with whether you go to therapy.

CAUSE runs a project competition, with a category specifically for intro stat students (USCLAP)

https://www.causeweb.org/usproc/

Statistics Education Resources

Meetings

- JMM and Mathfest – SIGMAA Stat Ed
- Joint Statistical Meetings – Section on Statistics and Data Science Education, Section on Teaching Stat in Health Sciences
- AMATYC – Statistics Committee
- US Conference on Teaching Statistics (USCOTS) – TBA, 2021
- electronic Conference on Teaching Statistics (eCOTS) – May 18-22, 2020
- International Conference on Teaching Statistics (ICOTS) – 2022 - Argentina
- International: ISI and IASE Satellite Conference

Teaching Intro Stat

- Qualifications for Teaching Introductory Statistics http://www.amstat.org/education/pdfs/TeachingIntroStats-Qualifications.pdf and Resources for Teaching Introductory Statistics, both from the ASA/MAA Joint Committee on Undergraduate Statistics http://www.amstat.org/education/pdfs/Resources-for-Teaching-an-Intro-Stats-Course.pdf
- Curriculum Guidelines for Undergraduate Programs in Statistical Sciences - http://www.amstat.org/education/pdfs/guidelines2014-11-15.pdf
- CAUSEweb – Consortium for the Advancement of Undergraduate Statistics Education - http://www.causeweb.org/. CAUSE puts on the USCOTS and eCOTS – two very worthwhile conferences!
- Stat101, a series of case studies for teaching Intro Stat from experts in the field - http://community.amstat.org/stats101/home
- ARTIST (Assessment Resource Tools for Improving Statistical Thinking) - https://apps3.cehd.umn.edu/artist/
- Recommendation for second applied course (http://www.amstat.org/education/pdfs/Second-Course-Syllabus.pdf) and math stat course (forthcoming)
- Stat2Labs – Games - http://web.grinnell.edu/individuals/kuipers/stat2labs/

Applets

- Rossman/Chance - http://www.rossmanchance.com/applets/index.html
- StatKey - http://www.lock5stat.com/StatKey/
- Art of Stat - http://www.artofstat.com/
- SPA - https://tinyurl.com/SPAapplets

Discussion Groups

- IsoStat (Isolated Statisticians) - http://www.amstat.org/committees/isostat/isostat.html
- AP Statistics (https://apcommunity.collegeboard.org/web/apstatistics/)
- Simulation-based inference blog - https://www.causeweb.org/sbi/
- Teach Data Science
- Ask Good Questions

Entertaining Videos - This is Statistics, Stats Can Be Cool, Realistic Statistics, When I Run A Test, How Far He'll Go

Statistics Education Journals

- Statistics Education Research Journal
- Journal of Statistics Education
    - Datasets and Stories
- Technology Innovations in Statistics Education
- Teaching Statistics
- Statistics Teacher

Data Resources

- Publishers
- CAUSE – www.CAUSEweb.org
- JSE Data Archive - http://www.amstat.org/publications/jse/jse_data_archive.htm
- TUVA Labs - https://tuvalabs.com/datasets/
- Census at School - http://www.amstat.org/censusatschool/
- Open Data – like http://opendataphilly.org/
- ICPSR (Consortium for Political & Social Research) - http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/index.jsp
- General Social Survey - monitoring social change & the growing complexity of American society, 30+ years - http://sda.berkeley.edu/GSS/
- Publicly available Health-related datasets
    - National Center for Health Statistics
    - NHANES (National Health and Nutrition Evaluation Survey)
    - HHS (Dept of Health and Human Services)
- NAEP (National Assessment of Educational Progress)
- Free Data from the National Climatic Data Center
- FedStats.gov - A link to numerous federal databases
- Futurity.org – daily emails with current scientific studies – not raw data
- Tip – search for .xlsx to help identify raw data instead of summary statistics