Assessment bank

This assessment bank contains unit tests and semester exams. For reading quizzes, calculation-based quizzes, hands-on projects, and other assessments, look in the regular library.

Contents

Name: _____
Date: _____

Unit 1 test – A

1. Jon and Lindsay are trying to decide which restaurant is the best. They've done their research and already agreed upon scores for cost, service, and food quality:

restaurant	**cost** (1 = cheap, 5 = expensive)	**service** (1 = slow, 5 = fast)	**food quality** (1 = terrible, 5 = delicious)	**Jon's final score**	**Lindsay's final score**
Karam	3	2	4		
Dil Se	4	2	3		
Thai Chili Jam	2	4	3		
21st Century pizza	1	5	2		

Jon's formula is $final\ score = service + food\ quality - cost.$
Lindsay's formula is $final\ score = service \times food\ quality - cost.$

A. According to Jon, which is the best restaurant? Which is the best restaurant according to Lindsay?

B. Their final scores are examples of what kind of data?

C. Do their formulas seem reasonable to you? Why or why not?

2. Put "yes" and "no" in the table as appropriate. There are no "maybe"s.

type of data	**categorical**	**ordinal**	**interval**	**ratio**
Allows you to add and subtract data				
Allows you to group data				
Allows you to multiply and divide data				
Allows you to rank data				

3. Provide an example of ordinal data. Describe what kind of research question it might be used to answer. Make up a scenario where ordinal data is sufficient to address the research question.

4. The following data represents cost in dollars for hairdos at a fancy salon in a one-hour period (costs differ because some people do just a cut; others do tinting; give big tips; etc.):

25 26 26 26 30 40 45 51 51 51 53 57 78

   A. Make a histogram of the data on graph paper. Label everything clearly.

   B. Make a box plot on graph paper. Label everything clearly.

   C. What are the lines and boxes in a box plot showing you? Be specific.

   D. Describe the distribution. (*Hint*: SOCS. *Hint 2*: any measure of C is ok. Same for S and O.)

   E. What is something you can notice more easily on the histogram than the box plot? Be specific.

   F. What is something you can notice more easily on the box plot than the histogram? Be specific.

5. Here is some information about Guillermo, a male high school student:

variable	z-score
Height	-2.34
Weight	-1.5
Bicep	+0.9
S.A.T. score	+1.78
G.P.A.	+0.15
Number of friends	+3.25
Number of discipline notices	-1.3

   A. The average number of hours spent per night on homework at this school is 67 minutes, with a standard deviation of 25 minutes. If Guillermo spend 75 minutes, what is his z-score for homework?

   B. If the average G.P.A. at this school is 3.5, and Guillermo's G.P.A. is 3.6, what is the standard deviation of G.P.A.s?

   C. What percentile is Guillermo's number of discipline notices?

   D. Tell a brief story about Guillermo. It should touch on at least five aspects of his life.

6. You want to study high school student stress levels to answer two research questions:

   RQ1: Does a student's stress level ebb and flow during the year or is it consistent?
   RQ2: Is there any relationship between student stress levels and courses?

   A. Write a useful conceptualization of stress. (*Hint*: Is there an option other than self-report?)

B. Come up with a workable operationalization. Don't assume you have infinite money or time, so try to make something realistic but good.

7. Diedrik Stapel curbstoned data sets that appears in 55 articles. True fact! But we might argue that a researcher whose methods produced nothing but bad artifacts—but didn't know it—might be worse. Why would we say that?

8. Mean, and its associated measure of spread standard deviation, are preferred for Normal data, uniform data, and other symmetrical distributions, but not for highly skewed distributions.

   A. Why not for highly skewed data sets?

   B. What summary statistics are preferable to mean and standard deviation for highly skewed data sets? Be specific.

9. A small town grows and shrinks each year depending on how babies are born and how many graduating high school students go on to college that year:

year	2015	2016	2017
population	2,432	2,456	2,414

   A. What is the population percent change from 2015 to 2016?

   B. What is the population percent change from 2016 to 2017?

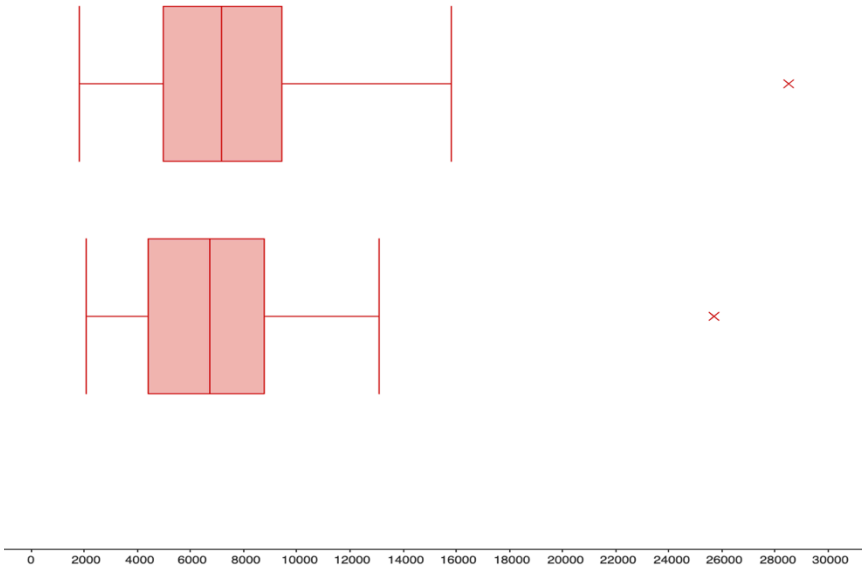
C. What is the mean population and its standard deviation for these years?

10.

A. Explain how changing standards can make comparing results over long time periods problematic.

B. Explain what publication bias is and why it makes us question even trustworthy data.

11. Read the following data display, then answer the following questions. The top box plot is 2021 crimes reported from a sample of major cities in the U.S. The bottom box plot is 2022 crimes from the same cities.[1]



A. What are the cases? Variables? What do the axes mean?

B. Describe the distributions (SOCS).

C. Interpret the meaning of the distributions. What would you conclude?

12. What do you call a group of statisticians? **This is a bonus question and a bad joke.**

[1] https://www.oregon.gov/cjc/CJC%20Document%20Library/2022%20Preliminary%20UCR%20Q4%20Final%20Draft.pdf

Name: _____
Date: _____

Unit 1 test – B

1. The following data represent the number of C.D.s (you know, for music in ye olden times) in various Portland-area school libraries:

0 0 55 62 74 89 101 143 178 197 282 327 459

   A. What is the mode?

   B. Make a box-and-whisker plot on graph paper. Clearly label the key values.

   C. Are any of the data points outliers? Show your work to justify your conclusion.

   D. Make an appropriate histogram on graph paper. Clearly label your axes and key values.

   E. What is something that's easier to notice on the box-and-whisker plot than on the histogram?

   F. What is something that's easier to notice on the histogram than on the box-and-whisker plot?

2. At a fictional college, A., the admission office creates an “academic index” for each student by adding their S.A.T. score (400–1600 points), G.P.A. (0–4 points), and total points for all A.P. exams taken (each one worth 1–5 points). Another fictional college, B., creates its academic index by instead adding each student’s z-scores for their SAT score, z-score for their G.P.A., and z-score for the total A.P. exam points.

   A. Define what an index score is. (Don’t recap the colleges’ formulas; explain the concept of an index score.)

   B. Define what a z-score is (in the second college’s index score).

   C. Given the below table, create index scores for each student at college A and at college B

students	S.A.T. score	G.P.A.	A.P.s (total points)
Student 1	1440	3.8	12
Student 2	1390	4.0	20

   D. Which college’s method makes more sense to you? Why?

   E. One student earned a G.P.A. of 3.4 at a school where $\mu = 2.9$ and $\sigma = 0.3$. What is the z-score of their G.P.A.? Show your work.

   F. What is the percentile at their school of the student in part (E)?

3. To produce statistics on how much violence appears in a T.V. show, some people (coders) must watch the episode and code it.

   A. Explain what coding is.

   B. Why do coders usually have a custom-designed tally sheet to work on?

   C. Fill in the blank with the specific vocabulary. A different researcher decides to use the number of complaints viewers have filed with the F.C.C. about a T.V. program's violence, instead of coding. This is an example of __ data.

   D. What are the pros in favor of coding? Or cons against using viewer complaints?

   E. What are the pros in favor using viewer complaints? Or cons against coding?

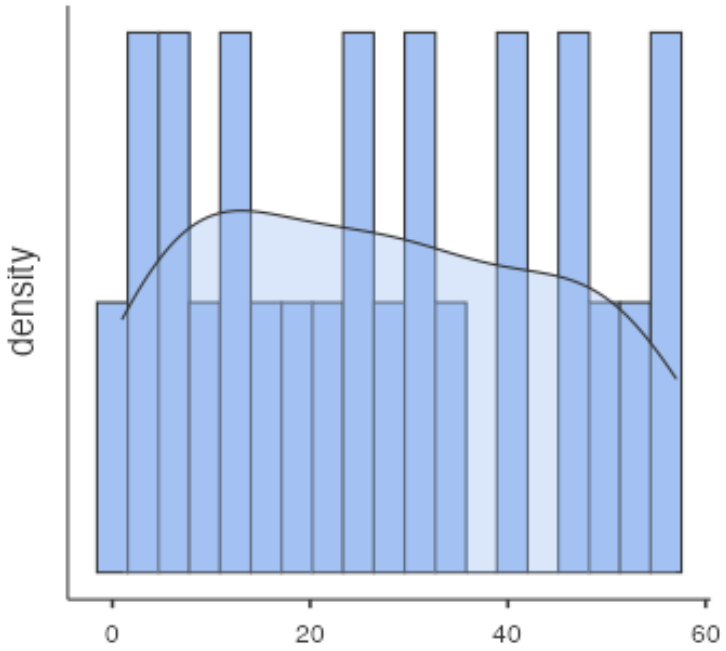
   F. Fill in the blank with the specific vocabulary. Coding and using viewer complaints are two different ways to __ the research method.

4. Two teachers are testing out a new textbook. One teacher says nothing to their students and hears few complaints about the book. The other teacher says, “This is a new book, so let me know if you see any typos.” The students in this class find typos, math mistakes, organization problems, and overall objections to the book.

    A. What is going on? Name the kind of bias and explain.

    B. Maybe the second teacher shouldn’t worry too much! The Law of Best says that most of the errors in the book will be __ , while only a few will be __ .

5. Describe the following distribution using SOCS.



Descriptives

	A
Mean	27.1
Median	26
Standard deviation	18.1
Minimum	1
Maximum	57
25th percentile	11.0
50th percentile	26.0
75th percentile	42.0

6. You're researching whether coffee improves concentration. Your specific research questions are:

   RQ 1: Will coffee improve a person's ability to concentrate intensely for short tasks?
   RQ 2: Will coffee improve a person's ability to concentrate for sustained periods of time?

   A. You realize you need to conceptualize "concentration." Give a conceptualization that works for both research questions—and remember, you're going to have to find a way to measure it!! Your conceptualization must be something realistic.

B. Is the answer that you gave for part (A) a hybrid variable? Explain.

C. Why is a hybrid variable useful (or not useful) for this conceptualization? Explain.

7. Calculate the following. Show your work.

   A. You earn $13 an hour and receive a 10 percent change raise. What's your new wage?

   B. You earn $8 an hour and receive a 12 percent change raise. What's your new wage?

C. Your house cost $500,000 to buy but loses 3 percent change in value. What's the new worth of the house?

D. The bank tells you the late penalties on your mortgage are 17 basis points per day. That doesn't sound too bad? Your loan is for $200,000. What's the daily penalty in dollars?

Name: _____
Date: _____

Unit 2 test – A

1. Define each term, then compare and contrast: questionnaire bias, response bias, and sampling bias.

2. Explain two potential strengths and two potential weaknesses of online polling. Use appropriate statistical terminology in your answer.

   Strength 1:

   Strength 2:

   Weakness 1:

   Weakness 2:

3. Why is a simple random sample almost always superior to an expert-chosen sample?

4. If your sample returns a $\hat{p}$ that is very different from the true proportion for the population, does that mean your sample MUST be biased? Explain using proper statistical terminology.

5. Imagine you are a surveyor measuring the heights of buildings downtown using trigonometry (i.e., using angles). Give a clear explanation of what happened in each of the following scenarios, using the terms accuracy, bias, precision, and variability.

    A. You took very detailed angle measurements, but your instrument was miscalibrated and all angles were too big.

    B. You took very detailed angle measurements with a correctly calibrated instrument.

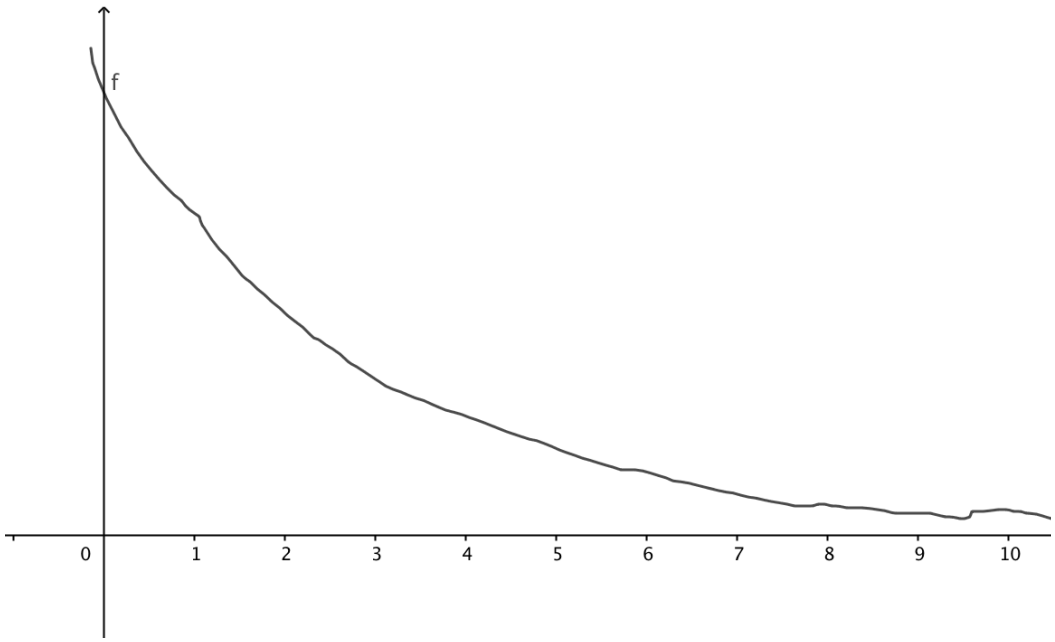
    C. You took sloppy angle measurements (not detailed or careful), and unfortunately, you kept misreading every angle by too much.

    D. You took sloppy angle measurements (not detailed or careful), but they aren't all wrong one way or the other.

    E. Which is the worse mistake: scenario (A) or scenario (D)? Defend your position.

6. When is leading information in a survey acceptable? When isn't it? Give an example of each.

7. The true population distribution of everyone's answer to the question, "On a scale of 1-10, how much do you like Butterfingers candy bars?", is right skewed, as shown on the graph below. A few people like Butterfingers, but most people find that it gets stuck in their teeth. The true mean is about $\mu = 2.2$ and a population standard deviation of 1.5.



A large random sample, $n = 100$, is drawn from this population. Sketch in the shape of the sampling distribution below. Use what you know about standard error to label key points.

8. When might self-report data (from a survey or poll) be more accurate than "official" data?

9. Researchers want to know how many licensed drivers in Portland have ever texted while driving.

   A. Design a simple random sampling method. Explain it below.

   B. Design a good stratified sampling method, considering that teenage drivers will need their parent's permission to answer the survey, whereas adult drivers can answer the survey straight out. Explain it below. ***OR***
   C. Design a good cluster sampling method, considering that you can easily find licensed drivers waiting at the D.M.V. who are bored and happy to talk to someone. Explain it below.

10. Researchers want to know if teenagers have made disparaging comments about other students' bodies. They suspect response bias will be a problem (no one wants to think they're being cruel). Write four indirect questions that get at this latent variable.

    1)

    2)

    3)

    4)

11. How many people should a pollster survey in a town of 100,000 people if they want to know public support for a ballot measure to within $\pm 2\%$? Show your work.

12. Pollsters sometimes use approval questions, where people can approve of multiple candidates (so totals can exceed 100%). Three candidates are running in a race for mayor.

    A. Candidate A gets $\hat{p} = 0.38$ out of 500 people surveyed. Write the confidence interval for this candidate.

    B. Candidate B gets $\hat{p} = 0.25$ on the same survey. Write the confidence interval for this candidate.

    C. Can we definitively say Candidate A has a lead over Candidate B? Explain.

13. Fifty-five people were surveyed about how many times they regret a movie purchase in a year, and the $\bar{x} = 1.7$ with $s = 0.3$. Write the confidence interval for this result.

14. A stratified sample finds that 92% of 903 international students and 46% of 522 domestic students at a college want to refurbish the student union building.

    A. Given that international students are 25% of the student population, and domestic students 75%, find the survey's final $\hat{p}$.

    B. Which group of students was oversampled? Why might this have happened?

Name: _____
Date: _____

Unit 2 test – B

1. Give three SPECIFIC ways that the order of questions in a survey might bias respondents' answers. Give a detailed example for each.

   1)

   2)

   3)

2. Why is "Are you a Democrat or a Republican?" a poorly phrased poll question? Name the type of bias. What is a better alternative question?

3. "Do you support this Congress's inflationary big spending priorities?" is a push poll-type question.

   A. Explain what a push poll is, and why the question is an example of one.

   B. Rewrite the question to fit in a normal poll or survey.

   C. Name the questionnaire bias problem you repaired (yes, it's a push poll. But the problem is…)

4.

A. Explain what social desirability bias is and how it affects polling.

B. Explain how a survey list experiment works.

C. Explain how a survey list experiment addresses social desirability bias.

5. What are the pros and cons of traditional phone polling? List two of each and explain using appropriate statistical terminology.

   Pro 1:

   Pro 2:

   Con 1:

   Con 2:

6. Give an example and explain why response bias doesn't have to be intentional.

7. Explain, in simple terms, why random sampling usually returns a fairly representative sample.

8. Consider the following situation:

   There are two ways to make scissors: symmetrical, where the scissors' fingerholes go straight through, causing some discomfort for both left-handers (about 10% of the population) and right-handers (90%); or handed scissors, where the fingerholes are sculpted, allowing for more control—if the correct hand is used. Teacher A decides to buy 20 symmetrical scissors for their students; teacher B decides to buy 10 lefties and 10 righties.

   A. What did teacher A decide to do? Apply the appropriate terms (accuracy, bias, precision, variability) and explain.

   B. What did teacher B decide to do? Apply the appropriate terms (accuracy, bias, precision, variability) and explain.

9.

   A. Explain: while researchers never know the actual sampling error, they do know the pattern of the sampling distribution.

   B. Explain what statisticians mean by, "We're 95% confident this interval captures the truth."

10. You're tasked with designing a sampling method. Consider the following scenario:

    Your company has been tasked with doing extensive in-person interviews in a very rural county. You have a good list of addresses, but most people on the list don't have phones or email. The county is incredibly large—100 miles by 100 miles—but because it is arid, people live near one of the ten surface water sources (streams, ponds, etc.) in the county.

    Using appropriate statistical terminology, describe an appropriate sampling method you could use in the situation.

11.

    A. A sample of 1002 people found that 37% believe in haunted houses (not the Halloween attraction kind, the "real" kind).[2] Calculate the confidence interval for the true population proportion.

    B. The same survey gave respondents a list of 10 paranormal beliefs (like haunted houses) and asked how many people believed in. The pollsters found $\bar{x} = 2.6$ with $s = 2.5$. Calculate the confidence interval for the true population average.

12. A poll reports that (56.8%,61.2%) of registered voters think the country is on the wrong track.

    A. What is the $\hat{p}$?

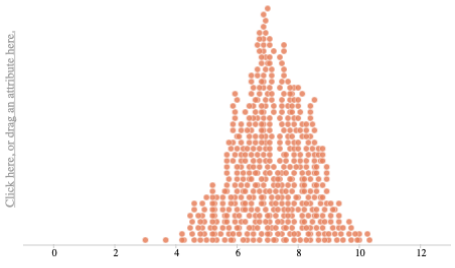
[2] https://news.gallup.com/poll/16915/three-four-americans-believe-paranormal.aspx

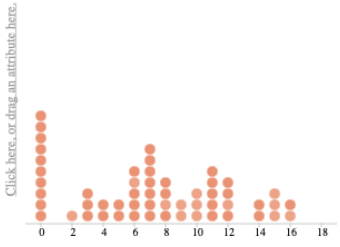
B. What is the standard error?

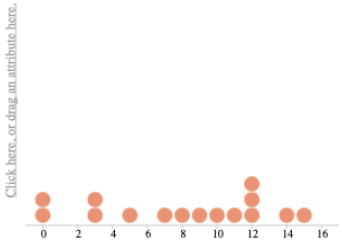
C. What is the n?

13. A stratified poll of high school students finds that 22% of underclasspeople (n=213) care about the lack of parking around campus, while 78% of upperclasspeople (n=108) care. Given that the underclasspeople are 75% of students at the school, while upperclasspeople are 25%, what is the overall $\hat{p}$ that the survey should report out?

14. Which of the following is the sampling distribution, a single sample, and the true population distribution? Justify your answers.

A. 

B. 

C. 

Name: _____
Date: _____

Fall exam – A

Part 1: Vocabulary

Directions: Define each term; give an example; and explain its importance. Each answer should be about 3–4 sentences, and no more than a paragraph.

Curbstoning
Non-response bias
Normal distribution
Operationalization
Oversampling
Simple random sample
Standard deviation
Standard error

Part 2: Calculations

1. A stratified sample of voters asked about support for the new tax cut bill in Congress finds the following results:

Race	Number	$\hat{p}$	Confidence interval
White	292	0.45	
Hispanic	157	0.42	
African American	98	0.39	
Asian American	78	0.47	
Native American	22	0.37	

   A. Fill in the last column with the correct confidence interval for each subgroup.

   B. Can we definitively say that all subgroups have less than 50% support for the bill? Defend your answer.

C. Let's say the demographic weights for the U.S. are as follows: {white = 0.66; Hispanic = 0.16; African American = 0.11; Asian American = 0.05; Native American = 0.02}. What average $\hat{p}$ should the survey report out as the overall result? Show your work.

2. A (made-up) survey of 1,256 Oregonians finds that the average income is $63,246, with a standard deviation of $25,282. What is the confidence interval this survey should report out?

3. You earn a z-score of 1.4 on a test. The average was 72%, and the standard deviation was 5%. What was your score?

4. Your stocks were worth $500 on January 1, 2022, and worth $485 on January 1, 2023. What was the percent change in value over 2022?

5.

A. Your height is a z-score of 2.1 for someone your age. What is your percentile for height? Assume that height is Normally distributed.

B. What percentage of people your age falls between z=-1 and z=2 for height?

6. For the following data:

   Number of granola bars eaten per day: 0, 0, 0, 1, 1, 1, 1, 2, 2, 3, 3, 4, 5, 6, 10, 15, 25

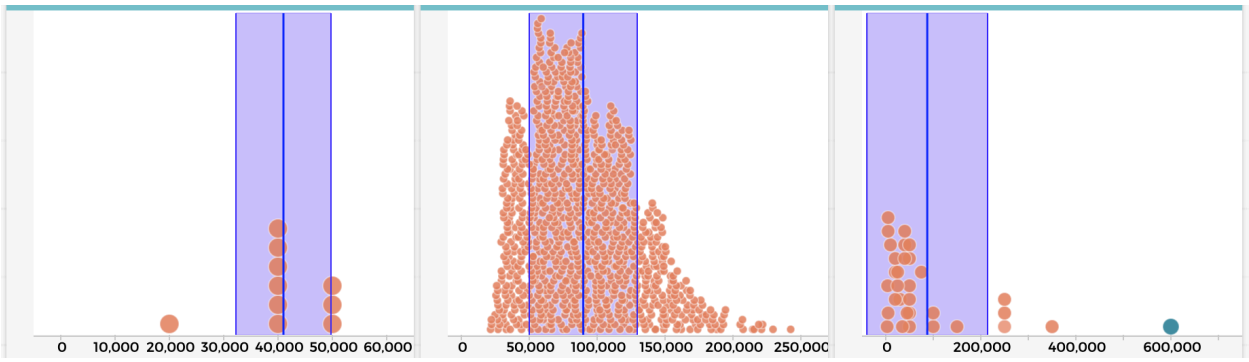
   A. Construct a boxplot.

   B. Describe the distribution (SOCS).

   C. Is someone who eats 9 granola bars per day an outlier or not? Defend your answer.

7. How much would it cost to have Maroon 5 play your prom? A lot, apparently: $600,000. We're going to look at a data set of the cost of hiring 40 different famous and "famous" bands.

   The following distributions are: a population, a sample, and a sampling distribution, though **not necessarily in that order**. Each distribution has its mean (center line) and standard deviation/error (half the width of the shaded area) marked.



   A. What is n, the sample size?

   B. Estimate μ, the population mean.

   C. Estimate $\bar{x}$, the sample mean.

   D. Estimate $\mu_{\bar{x}}$, the mean of the sampling distribution of $\bar{x}$s.

   E. Estimate σ, the population standard deviation.

   F. Estimate s, the sample standard deviation.

   G. Estimate $\sigma_{\bar{x}}$, the standard error of the sampling distribution of $\bar{x}$s.

Part 3: Essays

For each scenario, describe how you would address the problem. Make reasonable assumptions about what the scenario means and how much time/money/effort you could realistically use in solving the problem. Your answer should use appropriate statistical terminology, and it should show sophisticated conceptual thinking. Spend some time planning before you begin to write. You'll lose points if your answer strays into unrelated territory, so keep your answer focused. About 3–5 paragraphs will do.

Essay 1:

You want to sample many K–12 students in Portland about whether there is bullying at school. The sample should include a variety of schools—including elementary, middle, and secondary—as well as different types of public—i.e., neighborhood, charter, and magnet—**and** private schools—i.e., parochial, independent, and special needs. Although your cases are individual students, you want to have enough students in each type and subtype of school to have a decent confidence interval for it. Furthermore, you want to have enough students in each of the different demographic groups—i.e., age, race, sexual orientation, etc.—across the whole city to have decent confidence intervals for each.

Final consideration: parents must sign off on their child taking a survey before you can talk to them in the first place.

A full credit answer will describe, in sufficient but not excessive detail, a **sampling method that is unbiased**. You must consider all sources of bias and determine ways to element each. Don't worry about naming the sampling method; instead focus your energy on crafting a thoughtful method *of your own invention*.

Essay 2:

You want to determine how frequently and in what ways students at your school cheat on tests. There are two problems with asking people directly. First, students might not want to admit to cheating. Second, different students have very different ideas of what constitutes cheating, and so they might consider the strategy they use to be above-board. However, if you say, "Cheating is X, Y, and Z," then you end up with the first problem once again. Consider also that if you ask a generalized question, "Do students in your grade cheat?" then everyone is likely to remember high-profile examples, making it unlikely you will get an accurate sense of the true frequency.

A full credit answer will name and describe the **potential sources of questionnaire bias** in the context of this question and will include excellent questions or online activities that minimize this bias. Your questions or activities will show inventiveness and thoughtful appreciation of the difficulties in getting honest responses.

Name: _____
Date: _____

Fall exam – B

Part 1: Vocabulary

Directions: Define each term; give an example; and explain its importance. Each answer should be about 3–4 sentences, and no more than a paragraph.

Conceptualization
Push poll
Oversampling
Proxy measure
Outlier
Sampling distribution
Cluster sampling
Index data

Part 2: Calculations

1. A stratified sample of people at a school reveals the average number of times they touch their phones every hour (made up):

Role	Number	Average	Std. dev.	Confidence interval
Students	27	45	7	
Faculty/staff	24	72	5	

   A. Fill in the last column with the correct confidence interval for each subgroup.

   B. Can we say definitely faculty/staff touch their phones more often than students do? Defend your answer.

   C. What is the z-score of a teacher who touches their phone 61 times an hour?

D. Faculty/staff are, in fact, about 12% of the people at a school. Students are 88%. What final result should the survey report out?

2. A (made-up) survey of 2,309 Democratic voters, randomly selected, finds $\hat{p} = 0.43$ of support for Bernie Sanders. What is the margin of error of this survey?

3. Your friend height has a z-score of -1.1 for someone of their age and gender. Assume that height is Normally distributed.

   A. What is their percentile?

   B. If their height is 62 inches, and the standard deviation for heights in this group is 2 inches, what is the average height for this group?

4. What is the percent change of someone who earns 75 points on the original version of the test and 95 points on the retake?

5. The following data represent how many candies each child gets on Halloween:

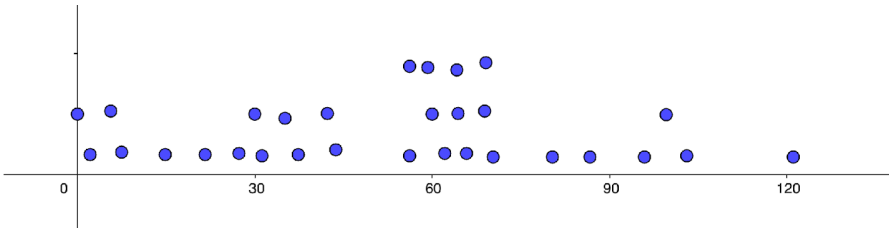
58	59	63	66	67	69	70	70
73	75	75	82	83	84	90	91
91	91	102	122	124	138	141	143
158	170	213	225	240	267	301	305

A. Make a histogram with an appropriate number of bins.

B. Describe the distribution (SOCS).

C. Using the $|z| > 3$ test, do you believe any child is an outlier in terms of how much candy they received? Defend your answer.

6. A random sample of 30 people is drawn from the population of U.S. voters and asked how many minutes of news they consume in a day.



A. Estimate $\bar{x}$ from the dot plot.

B. Estimate s from the dot plot.

C. Draw your best guess as to what the population distribution looks like. Be sure to label μ, the population mean, and σ, the population standard deviation, on the graph you make.

D. If this sample of $n = 30$ were repeated 1,000 times, what would the sampling distribution look like? Draw your best guess. Be sure to label $\mu_{\bar{x}}$, the mean of the sample means, and $\sigma_{\bar{x}}$, the standard error of the sampling distribution.

Part 3: Essays

For each scenario, describe how you would address the problem. Make reasonable assumptions about what the scenario means and how much time/money/effort you could realistically use in solving the problem. Your answer should use appropriate statistical terminology, and it should show sophisticated conceptual thinking. Spend some time planning before you begin to write. You'll lose points if your answer strays into unrelated territory, so keep your answer focused. About 3–5 paragraphs will do.

Essay 1:

After a natural disaster, a relief organization would like to know how many people remain in the affected area (comprising urban, suburban, and rural zones) and what needs they have (the needs may not be the same in different zones). Some areas were affected more than others by the disaster, meaning your mobility getting into these areas will be slow. Most phones are down; electricity is sporadic. You need to go door to door, though many houses are unoccupied—but people who are present often leave their neighborhood to forage for supplies. You do have decent maps of the roads but only a slight sense of which areas were hardest hit. You have the time and staff to sample enough people if you plan well.

A full credit answer will describe, in sufficient but not excessive detail, a **sampling method that is unbiased**. You must consider all sources of bias and determine ways to element each. Don't worry about naming the sampling method; instead focus your energy on crafting a thoughtful method *of your own invention*.

Essay 2:

You are designing the questionnaire for a women's rights group. The group wants to see if sexist attitudes and sexual behavior are correlated. They are going to use a voluntary online survey. Explain why both sexist attitudes and sexual behavior are difficult topics to get straight (no pun intended) answers to. Design the best questions for each topic you can think of, then finally consider any strengths or weaknesses to doing this as an online survey (other than sampling frame bias—assume the women's rights group has a solid sampling frame of email addresses to randomly pick from).

A full credit answer will name and describe the **potential sources of questionnaire bias** in the context of this question and will include excellent questions or online activities that minimize this bias. Your questions or activities will show inventiveness and thoughtful appreciation of the difficulties in getting honest responses.

Name: _____
Date: _____

Unit 3 test – A

1. What is an observational study? Explain the key difference between a poll and an observational study, then explain the key difference between an experiment and an observational study.

2. Correlation doesn't imply causation. Explain why in detail, using proper terminology and at least two examples.

3. You ask practicing lawyers whether they regret going to law schools. Only 10% say "yes." Explain why your method of research has certainly underestimated how many people who went to law school regret that decision, using proper statistical terminology.

4. You're watching a group of extraterrestrials play an alien board game you've never seen before. They can't communicate with you, so you just infer what you think the rules are. After 40 moves of the game, you've written down a list of 25 supposed rules. Explain the mistake using proper statistical terminology.

5. The following table categorizes all 174 people in a drug trial. Calculate an odds ratio, showing your work. Explain the meaning of the odds ratio.

Drug	Side effect	No side effect	Total
Drug A	46	15	61
Drug B	39	74	113

Total	85	89	174

6. Explain this statement, using proper statistical terminology: “The team that won the Super Bowl is a good team, but not necessarily the best team, this year.”

7. Explain the meaning of the following, using clear description or examples: “A good r^2 does not prove that a regression is ‘appropriate’ or good.” Give at least three potential problems.

8. The time that a runner takes to complete a marathon in is best predicted by two variables: their best time on a 15-mile run in the month prior to the marathon, and the temperature on the race day. (High temperatures slow runners down.) Specifically, the model says that $2 \cdot run\,time + 1.5 \cdot temperature = marathon\,time$. Given that the run time can vary from 50 to 100 minutes, and temperatures can vary from 40 to 105 degrees, which variable is more important, by what factor?

9. Calculate an r^2 for the following data and model. Either method is acceptable. Use software to help you—but write down the formula you’re using in case you make a data entry mistake.

x	y	predicted *y*	residual
1	43	44.88	-1.82
3	38	36.18	1.82

4	32	31.83	0.17
4	36	31.83	4.17
5	23	27.48	-4.48
7	19	18.78	0.22

10. Del Tone Car Stereos debuts a new stereo, the XQ, at a high price for a limited release. Fifty people buy it. Del Tone follows up and surveys these buyers. “Louder bass,” they all say. Del Tone upgrades the bass, starts up a large factory which lowers the price substantially, and then launches the product with a major advertising campaign. Five thousand people buy the newly redesigned stereo—yet one thousand want to return the stereo. Their complaint is that the stereo controls are too complex. Explain what has happened to the XQ Stereo in proper statistical terminology.

11. Consider the following correlation table of the predictor variables {A, B, C, D} and the response variable {E}.

	B	**C**	**D**	**E**
A	0.21	0.56	0.08	0.22
B		0.13	0.45	0.34
C			0.76	0.15
D				0.78

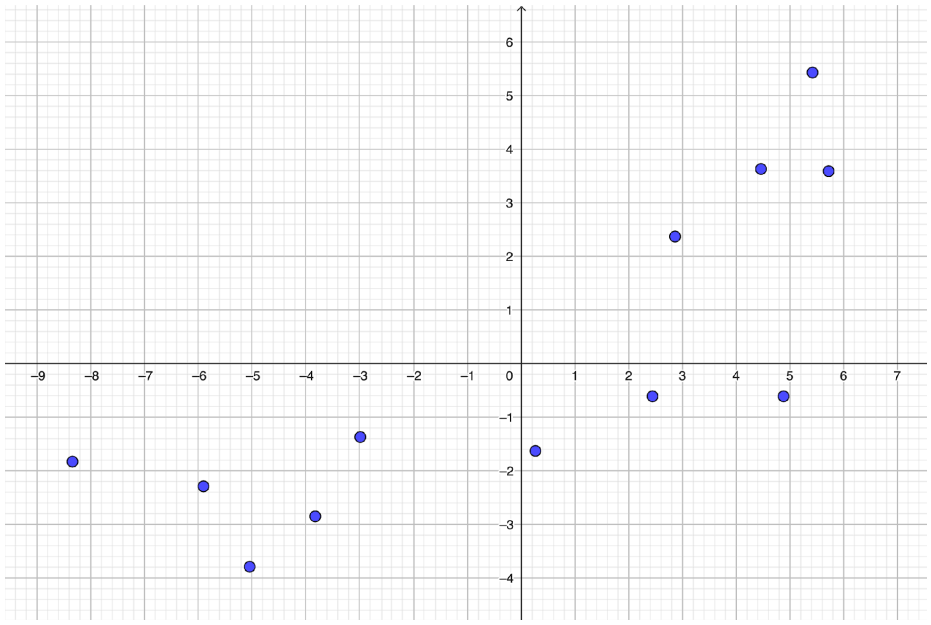
A. Which predictor variables probably shouldn’t be used in combination together? Explain.

B. What models do that leave for you that you would want to test? Explain.

C. If you had to guess, which model do you think will yield the best correlation with E? Explain.

D. If you had to guess, how high do you think that correlation might be? Explain.

12. For the following scatterplot:

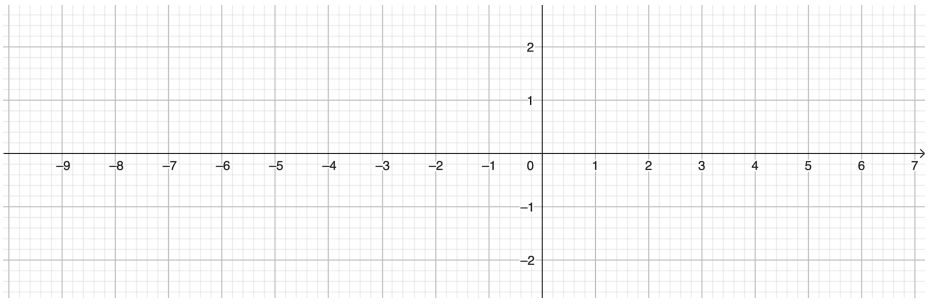


A. Given that $(\bar{x}, \bar{y})$ are the axes pictured, calculate the quadrant count ratio.

B. Describe the data using SODS.

C. Sketch in the best possible line above to fit the data (yes, I said line).

D. Sketch in the residual plot for the linear model you made in part (c) below. Do your best to get the vertical scale accurate.



E. What does the residual plot tell you, specifically about your linear model?

Name: _____
Date: _____

Unit 3 test – B

1. A correlation can defined as: knowing the value of variable A enables a better prediction about the value of variable B than a baseline guess.

    A. Give an example of what a baseline guess might be.

    B. Why is it important that the prediction (knowing A) must be better than the baseline guess?

2. Explain why an observational study isn't appropriate to make causal claims. Then explain why a quasi-experiment is the exception to this rule.

3. Terror Airways had the most accidents of any airline last year.

    A. If you had to guess, what is the most reasonable expectation of how many accidents they will have this year? (Options: the most, a lot, some, a few, the least.) Defend your choice. Why is yours the most reasonable expectation?

    B. What type of spurious correlation is this an example of? Explain what the term means and why it happens.

4. Explain why survivor bias makes many educational studies suspect.

5. You believe studying for more than 3 hours for the final exam might be correlated to a student's chances of passing. Here's a table of all the students participating in your research:

STUDENTS	Pass	Fail
Studying ≥ 3 hrs	35	4
Studying < 3 hrs	12	9

   Calculate the odds ratio, showing your work. Then interpret your results.

6. You are studying cancer rates in a 200-patient data set. You have 199 lifestyle variables for every patient (smoking, drinking, diet, weight, exercise, etc.).

   A. What happens if you use all 199 predictors in your equation? Use the proper statistical term and explain its meaning.

   B. How many predictor variables should have been used at a maximum?

7. Weight and smoking are used to predict heart attacks with the following model:

$$0.25 \cdot weight + 17.5 \cdot daily\ cigarettes = risk\ of\ heart\ attacks$$

For the individuals in the study, weight varied from 120 to 275 pounds, and cigarette consumption varied from 0 to 30. Assess the relative danger of weight compared to cigarette smoking. Show your work.

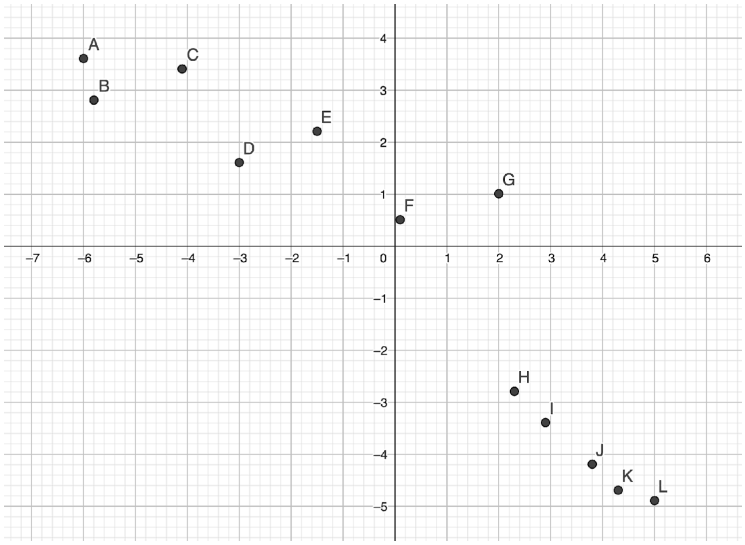
8. In a study that looked at the effect of music programs on high school students' S.A.T. scores, the researchers found the following results:

   $\beta_1 \cdot parental\ educ + \beta_2 \cdot student\ G.P.A. = S.A.T.\ score$ with an $r^2 = 0.20$ and
   $\beta_1 \cdot parental\ educ + \beta_2 \cdot student\ G.P.A. + \beta_3 \cdot hrs\ music\ per\ week = S.A.T.\ score$ with an $r^2 = 0.21$

   A. Using proper statistical vocabulary, describe what the researchers were attempting to do.

   B. What interpretation should the researchers have of these results? Why? Explain clearly.

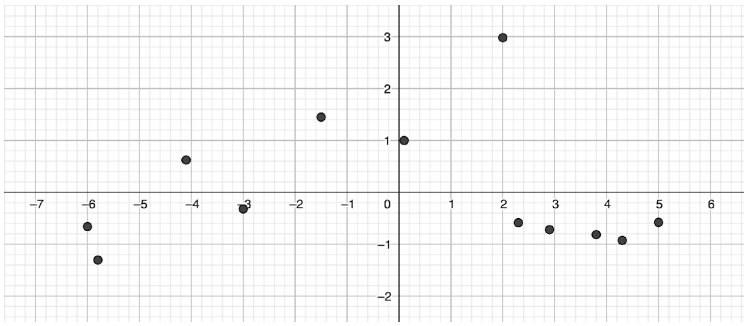
9. Look at the following scatterplot:



   A. Describe this scatterplot using SODS.

B. The axes are $(\bar{x}, \bar{y})$. Calculate the quadrant count ratio, showing your work. Interpret its meaning.

C. For the scatterplot, a linear model is fit to the points. The model produces this residual plot:



(i) Talk about how the residual plot was generated.
(ii) Talk about at least three interesting residuals and the original data points (A, B, …, K, L) that they are linked to.
(iii) Talk about what the residual plot tells us about the linear model used.

D. Calculate r^2. Either method is acceptable. Use software to help you—but write down the formula you're using in case you make a data entry mistake.

x	y	residual
-6.0	3.6	-0.67
-5.8	2.8	-1.314
-4.1	3.4	0.612
-3.0	1.6	-0.33
-1.5	2.2	1.44
0.1	0.5	0.988
2.0	1.0	2.97
2.3	-2.8	-0.596
2.9	-3.4	-0.728
3.8	-4.2	-0.826
4.3	-4.7	-0.936
5.0	-4.9	-0.59

10. The students who participate in an afterschool debate club have the highest G.P.A.s in a school. Your friend argues that debate made the students smarter. You think a spurious correlation is to blame. Explain.

Name: _____
Date: _____

Unit 4 test – A

1. In an experiment about students using a stress-relieving app before tests, researchers decided to have each student go through the experiment twice: once in the condition of not using the app, once in the condition of using the app. After each condition, students reported how nervous they were.

   A. Name the experimental design type.

   B. Explain why the researcher chose to deploy this design type.

2.
   A. In experimental design, what is a “protocol” for clinicians?

   B. Why is having a clear, good protocol important?

   C. Is a protocol an example of controlling, blocking, or randomizing? Explain.

3.
   A. Why is a control group is always necessary in an experiment?

   B. Why can a control group help rule out regression to the mean?

   C. Is a control group always given **no** treatment? Give an example and explain.

4. In a medical experiment, people on the new drug who experience side effects are switched to the old drug. People on the old drug who experience side effects just drop out.

   A. Explain why this is a problem.

   B. What could the experimenters do to repair the problem?

5. Which conceptual model of probability looks at logically derived formulas, such as in games of chance?

6. You receive a mysterious threat in the mail. Your two theories are that your friend is playing a joke on you (hypothesis A) or you have an enemy you didn't realize you had (hypothesis B). You think the odds are 10:1 for A:B.

   The next day at school, your friend asks you if you are sleeping well. If the friend sent you the threatening letter as a joke, there's a 5:1 chance they would ask you this question as part of their ruse. If your friend didn't send you the threatening letter, there's a 1:2 chance they would ask you this question (sometimes, they do genuinely seem to wonder why you're tired).

   Given the information that your friend asked how you're sleeping, update the odds for hypothesis A:B.

7. Consider the following two-way table:

PEOPLE	Has a house	Doesn't have a house
Has a good credit score	138,921	100,776
Has a poor credit score	2,456	329,322

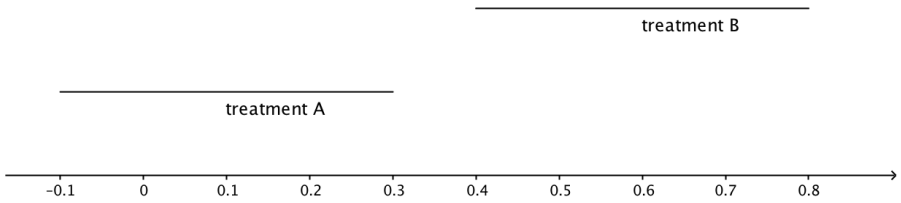
   A. What is $P(has\ good\ credit\ score\ \cap\ has\ house)$?

B. What is $P(has\ good\ credit\ score)$?

C. Are the events of having a house and having a good credit score independent? Show your work.

8. Define what a p-value is the probability of.

9. Look at the following experimental results. The x-axis is standardized effect size.



A. What is the point estimate for A?

B. What can we say definitively about treatment A?

C. What is the point estimate for B?

D. What can we say definitively about treatment B?

E. How do A and B compare?

10. Define statistical significance, and name one thing that could increase the statistical significance of the result of an experiment.

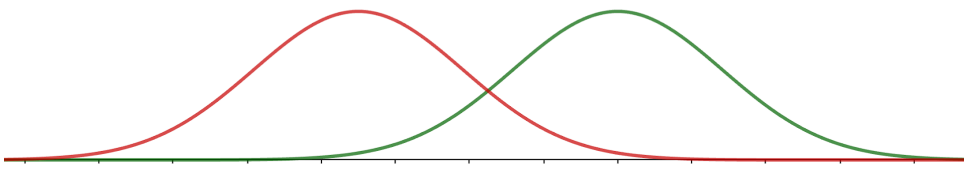
11. A test to identify dyslexia has a 5% false positive rate and a 3% false negative rate. Consider 10,000 students using the test, and assuming 10% have dyslexia.

    A. What is the meaning of a false positive in this context? Which type of error is it?

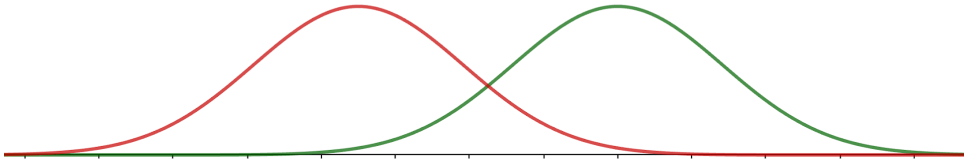
    B. What is the meaning of a false negative in this context? Which type of error is it?

    C. What is this test's false discovery rate? Make the appropriate two-way table.

12. Explain why making α more extreme decreases the Type 1 error rate but increases the Type 2 error rate, with words and a graph.



13. Explain why making α less extreme increases the Type 1 error rate but decreases the Type 2 error rate, with words and a graph.

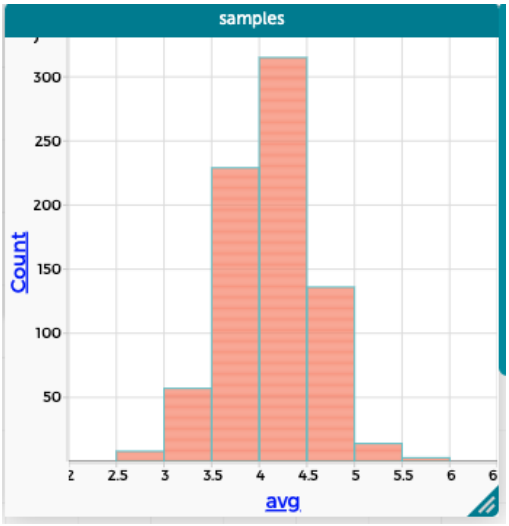


14. Experimenters are testing out whether expert winetasters can pick out the most expensive wine bottle out of a blind taste test. Each winetaster is given three wines (one expensive, one mid-priced, and one cheap) to taste and asked to rank them 1, 2, and 3.

    A. What is the H_0 and H_a?

    B. The experimenters calculated a score for each taster like this: $(rank\ of\ expensive\ wine\ -\ 1)^2 + (rank\ of\ mid\ -\ priced\ wine\ -\ 2)^2 + (rank\ of\ cheap\ wine\ -\ 3)^2$. Someone ranking the expensive wine as 1st, the mid-priced wine as 2nd, and the cheap wine as 3rd would get zero—a perfect score. The experimenters had 40 tasters taste, calculated each taster's score, then calculated an $\bar{x}$ for these scores.

    Here is the distribution of 1,000 possible $\bar{x}$s of 40 scores the experimenter would get under the null hypothesis (made by a simulation).



Let's say that their actual score, averaged over all their experimental subjects, was 3.49 (important statistical point: 3.49 > 0, and remember, 0 is a perfect score). What should be the experimenters' conclusion, do you think, about whether winetasters can tell the price point of a wine just by taste? **Justify your answer**. A correct answer should include your choice about an α the experimenters ought to have set beforehand but didn't.

Name: _____
Date: _____

Unit 4 test – B

1. In an experiment about fertilizers, researchers are testing the chemicals on various fields with different amounts of sunlight, rain, and wind, and different types of soil. Each field is divided by the researchers into two separate plots, A and B, that are right next to each other. A coin flip determines whether A or B receives the fertilizer; the losing plot gets nothing.

   A. Name the experimental design type.

   B. Explain why this design is superior to an alternative design (you chose the alternative).

2. The statistical definition of causality is different in some ways to a normal language definition of causality. One key distinction is the concept of ceteris paribus.

   A. Explain what the statistical definition of causality is.

   B. Explain what ceteris paribus means and its relation to the concept of causality.

   C. How do experiments establish causality in the statistical sense?

3. Blocking is one important tool in designing good experiments.

   A. Explain how blocking works. Pay special attention to how blocking affects random assignment. Give an example.

   B. Blocking can only be used for clearly understood variables. Explain why.

4. In an experiment in the workplace, people must know they're participating in an experiment for ethical reasons.

   A. Explain why people merely knowing they're in an experiment can present a problem.

   B. However, it's a bigger problem if people know which treatment group they're in. Explain why, and suggest a way to fix it.

5. Which conceptual model of probability looks at historical data?

6. Your car won't turn on. You have two theories about what is happening: the battery is dead (hypothesis A) or that the starter motor is dead (hypothesis B). Your initial assumption is that there are 3:1 odds for A:B.

   You notice that you can turn on the radio (although it's quite quiet). If the battery is too dead to turn on the car, there's still a chance it has enough power the radio, specifically 1:5 odds for enough power : too dead. On the other hand, if the starter motor is dead, there's a chance a short circuit prevents the radio from turning on, specifically 50:1 odds for radio will turn on : won't turn on.

   Given the information that the radio will turn on, update the odds for hypothesis A:B.

7. Consider the following two-way table:

STUDENTS	does theater program	doesn't do theater program
does speech & debate team	0	35
doesn't do speech & debate team	28	139

   A. What is $P(does\ theater\ program\ \cup\ does\ speech\ \&\ debate)$?

   B. What is $P(\neg does\ speech\ \&\ debate)$?

C. Are the events of doing theater and doing speech & debate disjoint? Justify your answer.

8. Does a p-value directly show the strength of evidence given by an experimental result? Explain your answer.

9. In an experiment, the treatment group has a standardized effect size confidence interval of (1.2, 2.4). The control group has a standardized effect size confidence interval of (0.1, 1.0).

   A. What is the point estimate for the standardized effect size of the treatment group?

   B. What is the point estimate for the standardized effect size of the control group?

   C. Can we say anything definitive about the treatment in comparison to the control? Explain.

   D. It may be surprising at first blush that the control group is definitively better than 0. But, if you think about it for a moment, it's not surprising at all. Give at least one reason why the control group might be definitively better than 0, and explain!

10. A huge effect size will show up even when n is small; a small effect size will show up when n is huge. Explain why these are both true.

11. Let's say that the T.S.A. (the folks who check your bags at the airport) have a 5% false positive rate and a 1% false negative rate for the computer-powered scanners.

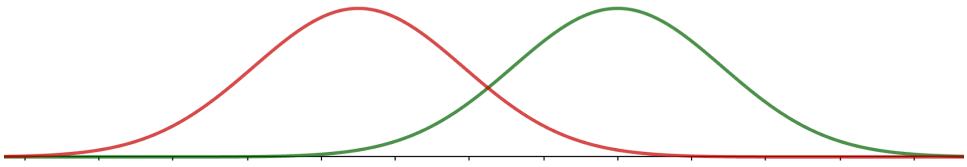
    A. What does a false positive mean in this context? Which type of error is it?

    B. What does a false negative mean in this context? Which type of error is it?

    C. What is the T.S.A.'s positive predictive value (or rate)? Make an appropriate two-way table to justify your answer?

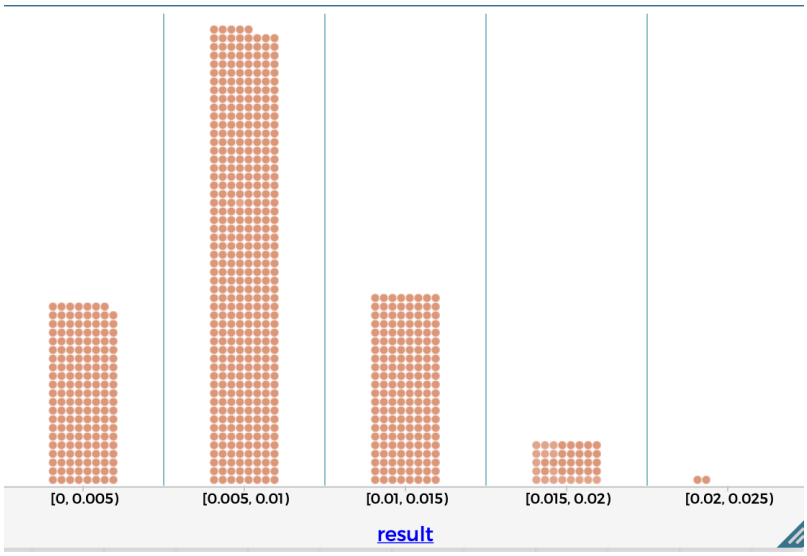
    D. In this context, which is worse: a false positive or a false negative? Why?

12. Explain why a bigger effect size has a lower β, even if nothing is done to change α, with words and a graph.



13. In an experiment to test out whether dead salmon are empathetic listeners, researchers scan the brains of 40 fish in an M.R.I. machine to see whether they have activation in their anterior insular cortex (the part of the brain associated with empathy). Activation for each fish is scored in percent terms from 0 to 1. The researchers get a result of $\bar{x} = 0.014$, as in little brain activation, on average, for their two-score deceased sockeye. But is that statistically significant?

    The M.R.I. machine produces some noise (static), no matter what. The researchers run a model of 1,000 simulations of the M.R.I. with 0 actual brain activity and get the following distribution of potential experimental $\bar{x}$'s:



    What is the conclusion that the researchers ought to draw? A correct answer will connect to a reasonable level of α that the researchers should have set but didn't.

Name: _____
Date: _____

Spring exam – A

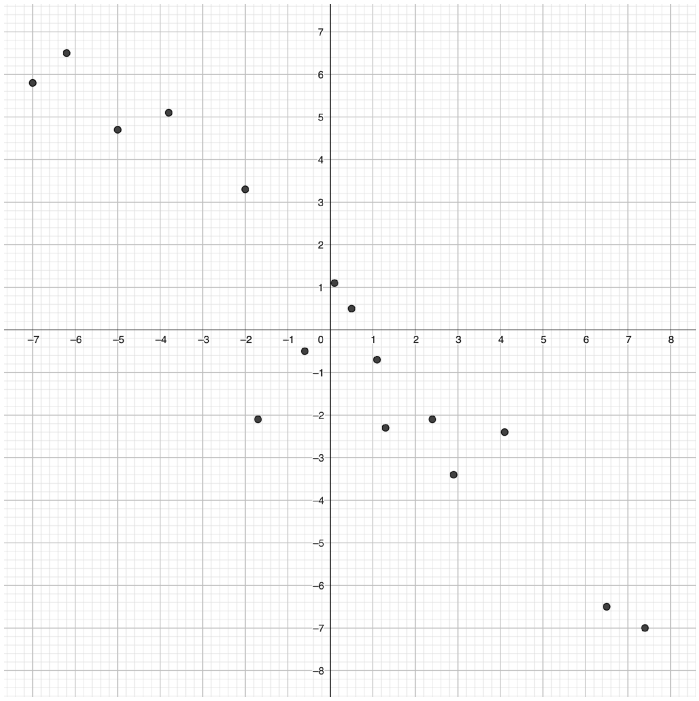
Part 1: Vocabulary

Directions: Define each term; give an example; and explain its importance. Each answer should be about 3–4 sentences, and no more than a paragraph.

Comorbidity
False discovery rate
P-value
Power
Repeated measure design
Berkson's fallacy
Confounding variable
Correlation

Part 2: Calculations

1. For the following data set:



A. Calculate the quadrant count ratio (the axes are $\bar{x}$ and $\bar{y}$).

B. The model that fits this data well is $y = -0.94x$.

x	y	predict y	residual
-7	5.8	6.58	-0.78
-6.2	6.5	5.828	0.672
-5	4.7	4.7	0.00
-3.8	5.1	3.572	1.528
-2	3.3	1.88	1.42
-1.7	-2.1	1.598	-3.698
-0.6	-0.5	0.564	-1.064
0.1	1.1	-0.094	1.194
0.5	0.5	-0.47	0.97
1.1	-0.7	-1.034	0.334
1.3	-2.3	-1.222	-1.078
2.4	-2.1	-2.256	0.156
2.9	-3.4	-2.726	-0.674
4.1	-2.4	-3.854	1.454
6.5	-6.5	-6.11	-0.39
7.4	-7.0	-6.956	-0.044

Calculate the r^2 of this model.

C. Describe this set of data (SODS).

2. Consider the following two-way table about all the computers at a school:

COMPUTERS	have malware	don't have malware
have connected to Internet only through school firewall	25	129
have connected to Internet on other modems or hotspots	38	54

A. What is $P(computer\ has\ malware)$?

B. What is $P(only\ connected\ through\ school)$?

C. What is $P(computer\ doesn't\ have\ malware\ \cap\ only\ connected\ through\ school)$?

D. What is $P(computer\ has\ malware\ \cup\ connected\ to\ other\ modem)$?

E. Are the method of connecting to the Internet and having malware independent events or not? Justify your answer.

3. Consider the following data for on-time flights:

FLIGHTS	on-time	delayed
Delta Airlines	229	1,923
Alaska Airlines	132	1,766

Quantify the odds ratio, then interpret its meaning.

4. Find the false discovery rate for the following setup: An unlikely hypothesis (a 10% chance it's true) about listening to music improving the taste of food is tested. The false positive rate

of this experiment is 5%, and the false negative rate for the design is 1%. If 1,000 experiments are done, what is the false discovery rate? Show your work.

5. An experiment about texting reminders to vote with 500 people in the treatment group and 750 people in the control group finds that $\hat{p} = 0.72$ for the treatment group, that is 72% of them actually end up voting, and $\hat{p} = 0.64$ for the control group.

   A. Calculate the confidence interval for the treatment's true effect. Keep at least three decimal places.

   B. Calculate the confidence interval for the control's true effect. Keep at least three decimal places.

   C. Based on the confidence interval, what can be said with certainty?

   D. The p-value happens to be $p < 0.01$ in this case. What would you conclude based on that?

6. The equation researchers come up with for predicting whether a plant flowers in an agricultural lab is
   $5 \cdot watts\ of\ sunlight\ during\ the\ day + 10 \cdot mL\ of\ water = number\ of\ flowers$, where watts can vary from 15 to 20 and water can vary from 2 to 3. Which predictor variable is more important, and by what factor?

Part 3: Essays

Directions: Use proper statistical vocabulary. Give examples. Show that you understand the concepts thoroughly and deeply by focusing on the most important issues. Answers should be longer than a paragraph, but shorter than five paragraphs.

Essay 1:

You are the new superintendent of K–12 schools in Salem/Keizer, a major school district with thousands of students in an urban/suburban area. Your predecessor launched numerous experiments to help poor children do better in school but didn't insist on good data protocols. It's a mess; you have no idea what really works and what experiments need to be shut down immediately. Experiment A has been tested at a whole-school level at five schools where the principals opted in. Experiment B is an afterschool program that students (or families) elect themselves into. Experiment C is one where student participants were selected by a formula that gave points for teacher recommendations (higher points for responsible students) and points for need (higher points for greater need).

Explain to your data-challenged staff why all these experiments have problematic data, even though all results look positive. Describe to them what a decent method would actually look like. Assume your staff has access to school-wide and individual student test scores, measures of poverty, etc. A full credit answer will name and thoroughly describe the potential sources of bias in A, B, and C. All sources of bias must be addressed in a thoughtful answer of what a good method looks like. You must show your staff what reliable data collection looks like, so future experiments will make good comparisons about how well these experiments are working.

Essay 2:

Experimenters are testing whether a new toothpaste decreases the number of cavities users get compared to the standard toothpaste. They test the new toothpaste on 1000 users and find the standard deviation in the number of cavities is 1.45. The experimenters set an alpha of -0.1 cavities—in other words, if the average improvement for the treatment group is better than 1/10 fewer cavities over the control group, they will call the result statistically significant. Your job is to provide a framework to the researchers about the possible results of the experiment and possible errors they could make.

A full and complete answer will address: what Type 1 and Type 2 errors are; which type of error is worse in this context and why; what the Type 1 error rate the experimenters are willing to accept is (based on what information was given above); and what the Type 2 error rate would be in the specific scenario that, in fact, the new toothpaste decreases the number of cavities on average by 0.15 per person.

Name: _____
Date: _____

Spring exam – B

Part 1: Vocabulary

Directions: Define each term; give an example; and explain its importance. Each answer should be about 3–4 sentences, and no more than a paragraph.

Matched pair design
Multiple regression
Prior odds
Regression to the mean
Residual
Scatterplot
Significance level
Survivor bias

Part 2: Calculations

1. Use this table of athletes at an international sports competition for the following questions:

ATHLETES	athlete is doping	athlete is not doping
positive test	45	12
negative test	3	590

A. What is the false positive rate of this test?

B. What is the false negative rate of this test?

C. What is the false discovery rate of this test?

D. What is the false skepticism rate of this test?

E. Overall, what do you think about this test? Justify your answer.

2. Use this table of students' technology for the following questions:

STUDENTS	has iPad	doesn't have iPad
has cell phone	38	95
doesn't have cell phone	18	45

A. What is $P(has\ cell\ phone\ \cap\ has\ iPad)$?

B. What is $P(has\ cell\ phone\ \cup\ has\ iPad)$?

C. What is $P(has\ iPad)$?

D. Are having a cell phone and having an iPad mutually exclusive events? Justify your answer.

E. Are having a cell phone and having an iPad independent events? Justify your answer.

3. Calculate r^2 from the following data table. Show your work.

y (observed)	residual (from model)
-2.0	-0.2
-1.0	+0.3
0.0	-1.0
1.0	+0.4
3.0	+0.1
5.0	-0.6
6.0	+0.2

4. An experiment to determine whether walking every other day improves resting blood pressure finds that the treatment group has an average improvement of 0.7 effect size over the control group, with a standard deviation of 1.1, for $n = 9$. Another treatment group walks every day and has an average improvement of 2.1 effect size, with a standard deviation of 1.5, for $n = 12$.

   A. Make confidence intervals for both treatment groups, the every other day group and the every day group. Clearly label which is which.

   B. Display both confidence intervals/error bars below. Clearly label which is which.

   C. What conclusions should be drawn from this experiment? Explain why.

5. Consider the following made-up data:

   (*hours of study*, *stress level on 0-5 scale*)
   (0.7, 0.64) (2.38, 0.72) (4.28, 1.22) (5.66, 2.86) (7.76, 4.28)

   A. Plot the following data points below.

   B. Calculate $(\bar{x}, \bar{y})$.

   C. Calculate the quadrant count ratio.

   D. Describe this data set (SODS).

6. A certain heart drug is tested on a group of elderly patients with the following results:

PATIENTS	no adverse event	adverse event
new drug	59	4
old drug	46	7

   Does the new drug give lower odds of adverse events to no adverse events compared to the old drug? Show your work, then interpret the meaning of the odds ratio.

7. A school district uses the following formula to predict how many A.P. course a student will take: $10 \cdot G.P.A. + 5 \cdot visits\ to\ college\ counselor = number\ of\ A.P.\ courses$. Given that G.P.A. ranges from 2 to 4 and visits to visits to the college counselor range from 0 to 3, which predictor is more important, and by what factor?

Part 3: Essays

Directions: Use proper statistical vocabulary. Give examples. Show that you understand the concepts thoroughly and deeply by focusing on the most important issues. Answers should be longer than a paragraph, but shorter than five paragraphs.

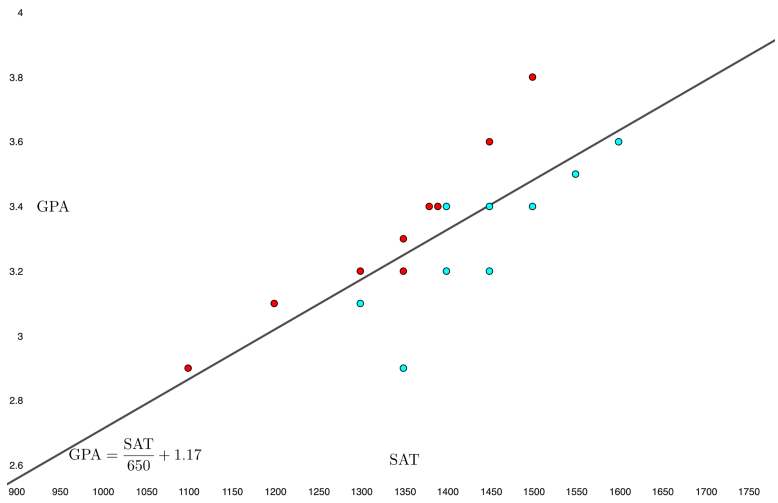
Essay 1

An education company has produced a new textbook for a math support class (it's not a traditional math textbook; it has lots of games, requires careful reading, and aims to teach problem-solving skills). The company wants to test its effectiveness at multiple schools in math support classes. The schools differ in the percentage of students in math support who are also in an English as a Second Language support class. The researchers are worried E.S.L. students might struggle more with the reading level required. Furthermore, students who are in the same math support class might have a different regular math class (e.g., Algebra 1) teacher—and those teachers can be of highly variable quality. Finally, each school has different rules about which students are placed into math support, though none of them seem to have a solid rationale about their rule or care about it very much.

Design a good experimental set up for the company that controls, blocks, or randomizes every confounder listed above—and mention and solve for <u>at least one</u> additional confounder you think of.

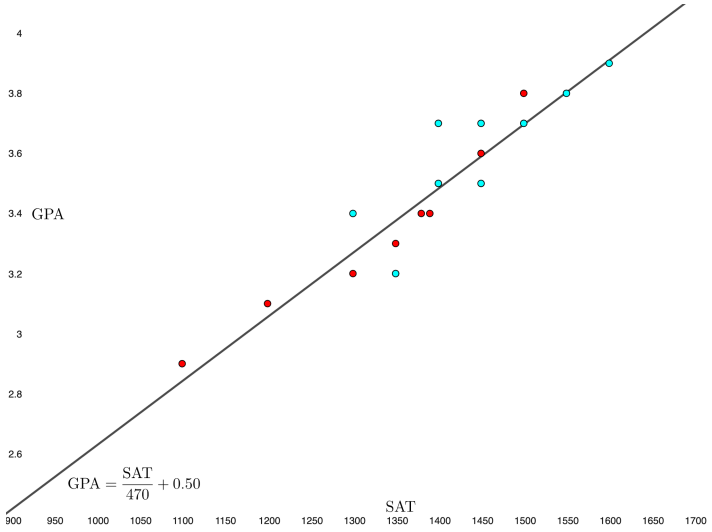
Essay 2

Colleges use S.A.T. scores to predict how well newly admitted students will do in their first-year classes (their G.P.A.). Here's data from a number of students:



The linear model has an $r^2 = 0.61$.

However, some colleges have much worse grade inflation than others. These colleges are pictured in red. The average effect is estimated to be 0.3 points in G.P.A. When 0.3 points are added to the G.P.A.s of students at non-inflationary colleges, the data look like this:



The linear model has an $r^2 = 0.86$.

A full, complete essay will: describe the original model given in the first graph in terms of SODS; name what the researchers who are analyzing the data did with the second model; explain why they did it; what should their conclusion be; and whether this makes sense in the real world.