Stat 100 Exam 1 Spring 2018

Question 1 *(14 pts)* A study in Stat 100 is comparing the learning outcomes of online versus in person students taking the same *exact course* (same lectures, homework, exams, etc), the only difference being that students who choose the online class watch videos of the lectures, while students who choose the in person class watch the lectures in person. Last semester's results of 1600 students showed the online class had a significantly higher final grade average than the in person class.

a) *(2 pts.)* What type of controls were used in this study?
i) Randomized controls **(ii)** Non-randomized controls **iii)** Historical Controls **iv)** No Controls

b) *(2 pts.)* If the online section is called the treatment group, who is the control group?
i) There is no control group. **ii)** All Stat 100 students from previous semesters. **(iii)** Students in the in person section.

c) *(2 pts.)* Can we conclude that the reason the online students did significantly better is because of the online teaching?
i) Yes, the higher grades in the online class had to be due to the online lectures since everything else was *exactly* the same.
(ii) No, since the students chose which class to be in there could be confounders mixing up the results.
iii) Yes, the online class had the advantage of being able to watch the lectures at their convenience and review them multiple times to reinforce what they learned enabling them to score higher on exams and homework.
iv) Yes, because any selection bias introduced by the students choosing which group to be in would average out with such a large sample size making the 2 groups as balanced as possible to begin with, so confounders wouldn't be a problem.

d) *(6 pts.)* Circle whether the following are confounders, causal links, or neither based only on the given information:
i) Access to the videos- The online class had immediate access to the video and the in person section did not. Online students can watch the videos as many times as they want and pause them at anytime to review, allowing them to grasp the concepts better and do better in the class. ***Circle*: a)** confounder **(b)** causal link **c)** neither
ii) Math ability—Students who are good at math are more likely to do better in a class like Stat 100 and are more likely to sign up to take Stat 100 online. ***Circle:*** **(a)** confounder **b)** causal link **c)** neither
iii) Sample size – The in person section is significantly larger than the online section (1100 vs. 500). The online class did better because there were less students, meaning they'll have less F's. ***Circle:*** **a)** confounder **b)** causal link **(c)** neither

e) *(2 pts.)* Suppose we thought that income might be a confounder, how would we stratify to eliminate its confounding effect?
i) At the end of the study we'd divide the students into subgroups based on income (high, medium and low) and compare the percent of in person to online students within each subgroup.
(ii) At the end of the study we'd divide the students into subgroups based on income (high, medium and low) and compare the response (final grades) of the in person vs. online students within each subgroup
iii) At the end of the study we'd divide the students into subgroups based on income (high, medium and low) and then compare the response (final grades) of students between the 3 income groups.

Question 2 *(10 pts.)* Now let's say another university is doing a similar study. There are 500 intro stats students who are ***randomly assigned*** to either take the online class or the in person class. At the end of this study, they found no significant difference between the two groups. The final grade average in the online class was about equal to the final grade average in the in person class

a) *(2 pts.)* Which of the following best describes this study?
i) An observational study with controls
(ii) A randomized controlled experiment
iii) A non-randomized experiment with historical controls

b) *(2 pts.)* Is the study double blind? **i)** Yes **(ii)** No, the students know which class they're in. **iii)** Not enough info

c) *(2 pts.)* Are there potential confounders in this study?
i) Yes, we should always be worried about potential confounders mixing up our results.
ii) Maybe, we have no way of knowing if there are confounders until we repeat the study.
(iii) No, randomization into treatment and control eliminates any systematic differences between the groups.

d) *(2 pts.)* In this study, there are only 40 upperclassmen in the class of 500. The researchers want to make sure the 40 upperclassmen are evenly divided between the treatment and control group. What's the best way to do this?
i) Let the upperclassmen choose which group to be in allowing 20 slots for each.
ii) Put all the upperclassmen in the control group so they won't bias the effects of the treatment.
(iii) Randomly assign half the upperclassmen to treatment and half to control then do the same with the rest of the class.

e) *(2 pts.)* If you chose the best method in (d) would it introduce selection bias? **i)** Yes **(ii) No**

Question 3 *(6 pts.)* Two pharmaceutical companies each tested their own drug for baldness on both men and women. The table below gives the results after 12 weeks of treatment.

	Drug A			Drug B		
	# Improved	# Not-Improved	% Improved	#Improved	# Not-Improved	% Improved
Men	40	160	20%	30	170	15%
Women	85	15	85%	300	100	75%
Total	125	175	41.67%	330	270	55%

a) *(2 pts.)* Based on the table above which drug works better?

i) A since it has a higher improvement rate for both males (20% vs 15%) and females (85% vs 75%).

ii) B since it has the higher total improvement rate (55% vs. 41.67%).

iii) It depends on whether the people are male or female.

b) *(2 pts.)* B had a higher improvement rate for women that it did for men (75% vs. 15%) because more women than men took B (400 women vs. 200 men). **i)** True **ii)** False

c) *(2 pts.)* If A has a higher improvement rate for men and for women when looked at separately, why does A have a lower improvement rate when the groups are combined?

i) It's because A works better on men and women when they're separated into groups by gender but B works better on them when they're all given the same treatment without regard to their gender.

ii) It's because A was tested on only 300 subjects total whereas B was tested on 600 subjects total.

iii) It's because A was tested on proportionately more men (40/125 vs 30/330), and men have a lower improvement rate.

Question 4 pertains to the following study: *(6 pts.)*

The university is doing a study to see if drinking a daily dairy beverage will help improve people's performance on memory based tests. 100 adults aged 25-45 are randomly assigned to either drink a dairy based beverage or a non-dairy based beverage every day for 10 weeks. Both beverages look and taste the same so neither the researchers nor the participants know who's in which group. At the end of the study, the researchers will compare the test scores of each group.

a) *(2 pts.)* What type of bias could be present in this study? ***Choose one:***

i) No systematic bias **ii)** Subject Bias **iii)** Evaluator Bias **iv)** Selection Bias **v)** ii, iii, and iv

b) *(2 pts.)* Which of the following could confound the results? ***Choose one:***

i) Commitment- Those who commit to a 10 week study and remember to drink the beverage on a daily basis are also more likely to do well on memory tests.

ii) Differences in Memory ability -- The people in the 2 groups are bound to be different in their memory ability at the start of the study so those initial differences could easily overshadow whatever effect the drink may have.

iii) Mindfulness and engagement—Studies like these usually require daily food diaries and check-ups which encourage people to improve their health habits which in turn could improve their memory.

iv) All of the above

v) None of the above

c) *(2 pts.)* One of the Stat 100 instructors, Karle Flanagan, is participating in this study but quits drinking the beverage for unknown reasons. She's a non-adherer! A few others do the same. Which comparison is best when analyzing the final data?

i) Compare everyone assigned to treatment to everyone assigned to control to keep the original randomization.

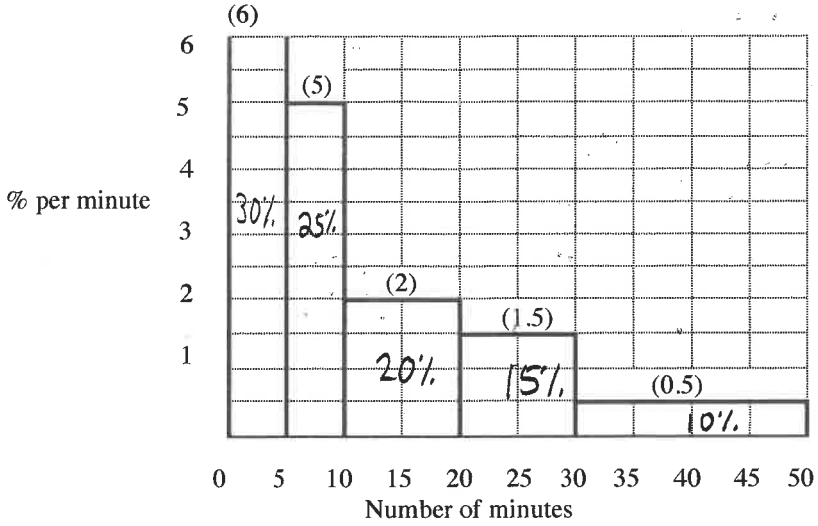
ii) The dairy beverage can only affect those who take it. Compare those who drank the dairy beverage (adherers) in the treatment group to those who drank the non-dairy beverage in the control group.

iii) Compare the people who drank their beverage (adherers) in the treatment group to the people who didn't drink their beverage (non-adherers) in the treatment group. Then do the same with the control group.

iv) If people don't adhere to the study protocol, there is no way to still analyze the results without introducing bias.

Stat 100 Exam 1 Spring 2018

Question 5 *(16 pts.)* The histogram below depicts the number of minutes students reported being distracted (not paying attention) during a 50 minute Stat 100 lecture. Intervals include the left-endpoint but not the right. (For example someone who reported being distract for 10 minutes would fall in the 10-20 block not the 5-10 block.) The height of each block is given in parentheses.
(*Assume even distributions throughout each interval.*)



a) *(3 pts.)* What percentage of the students reported being distracted in the following intervals?

0-5 minutes 30 % 10-20 minutes 20 % 20-30 minutes 15 %

b) *(2 pts.)* The median is closest to **i)** 5 **ii)** 6 **(iii) 9** **iv)** 15 **v)** 20

c) *(2 pts.)* The average is > the median

i) less than **(ii)** greater than **iii)** equal to **iv)** cannot be determined

d) *(2 pts.)* Did more people report 0-10 minutes or 20-50 minutes, or are they the same?

(i) 0-10 **ii)** 20-50 **iii)** Same

e) *(2 pts.)* What percentage reported exactly 4 minutes? (Assume an even distribution through the intervals) 6 %

f) *(2 pts.)* The 90th percentile corresponds to 30 minutes. Fill in the blank with the correct number.

g) *(3 pts.)* If everyone had reported 10 more minutes of distraction, how would the average, median and SD change?

The median would:	**(i)** increase	**ii)** decrease	**iii)** stay the same
The average would:	**(i)** increase	**ii)** decrease	**iii)** stay the same
The SD would:	**i)** increase	**ii)** decrease	**(iii)** stay the same

Question 6 *(9 pts)* Consider this list of numbers: 1,1,0,2,6 0,1,1,2,6

a) *(2 pts.)* The average = 2 **b)** *(2 pts.)* The median = 1

c) *(5 pts.)* Compute the SD.

i) *(2 pts.)* The deviations are: −2, −1, −1, 0, 4 accept any order

ii) *(1 pt.)* The sum of the deviations should = 0. Fill in the blank with the correct number.

iii) *(2 pts)* The SD = _____ Round answer to 2 decimal places. **Show work below.** *(No work, no credit!!)*

square deviations: 4, 1, 1, 0, 16

avg of squared deviations: $\frac{4+1+1+0+16}{5} = 4.4$

square root: $\sqrt{4.4} = 2.10$ or 2.09

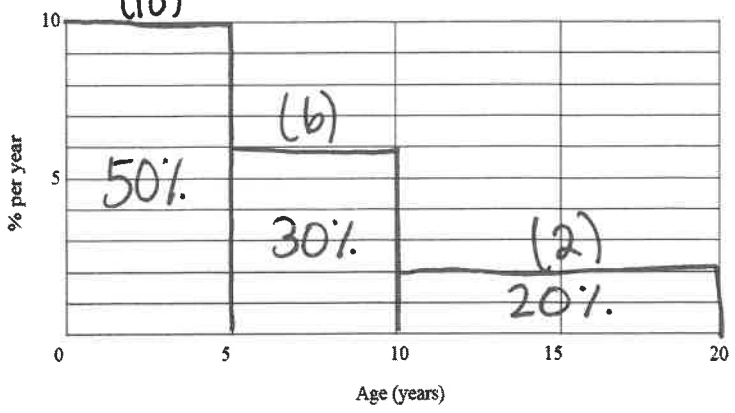
1pt work
1pt answer

Question 7 *(12 pts.)* A distribution table for the ages of passenger cars in the US is given in the table below. Class intervals include the left endpoint but not the right. For example, a 5 year old car would be included in the second row, not the first.

a) *(3 pts.)* Fill in the height column.

Age in years	%	Height (% per year)
0-5	50	10
5-10	30	6
10-20	20	2

b) (3 pts.) Draw the histogram below.



avg > med

c) *(2 pts.)* The median = 5

d) *(2 pts.)* Fill in the blank with **<, > or =** to make the following statement true: The median < the average.

e) *(2 pts.)* If everyone in the US kept their cars 10% longer how would the average, median and SD of the age of cars change? (adding 10% is the same as multiplying by 1.1)

i) The average, median and SD would increase by 10%

ii) The average and median would increase by 10% but the SD would stay the same.

iii) Only the average would increase, the median and SD would stay the same.

Question 8 (2 pts.)A class of 400 students is divided into two sections of 200 each. Both sections are given a common final exam. Section A scored an average = 80, SD=10 and section B scored an average = 60, SD=10. Suppose the 400 scores are combined as one list.

a) *(1 pt.)* The average of the combined list = 70. Fill in the blank with the correct number.

b) *(1 pt.)* The SD of the combined list is > 10. Fill in the blank with **>, <, or =.**

Question 9 *(6 pts)* According to your Survey 1 responses, your mothers' ages follow the normal curve with average=49 yrs and SD=6 yrs. What percent of the class has mothers between the ages of 46 and 55 years? Answer this question in 3 steps.

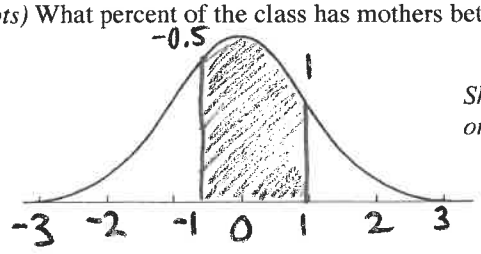
a) *(1 pt.)* The z-score for 46 years = -0.5. $z = \frac{46-49}{6} = -0.5$

(1 pt.) This means that mothers who are 46 years old are 0.5 SDs below average. ½ pt each
(Fill in the first blank with a number and the second blank with "above" or "below")

b) *(1 pt.)* The z-score for 55 years = 1. $z = \frac{55-49}{6} = 1$

(1 pt.) This means that mothers who are 55 years old are 1 SDs above average. ½ pt each
(Fill in the first blank with a number and the second blank with "above" or "below")

c) *(2 pts)* What percent of the class has mothers between the ages of 46 and 55? 53 % 1 pt
accept 53.28



Show work by shading the correct area on the normal curve. 1 pt

$\frac{1}{2}(38) + \frac{1}{2}(68) = 53$

Stat 100 Exam 1 Spring 2018

Rounding instructions for Question 10: You may "round" z scores and percentages to fit the closest line on the normal table and you may round percentages on the table to the nearest whole number.

Question 10 *(18 pts.)*

The histogram of the number of hours of sleep Stat 100 students reported getting each night roughly follows the normal curve with an **average of 6.8 hours and a SD of 1.2 hours.**

z=1

a) *(2 pts.)* Approximately **68%** of the students sleep between 5.6 **hours** and 8 **hours**
(Put the smaller number first)

z=2 or 1.95

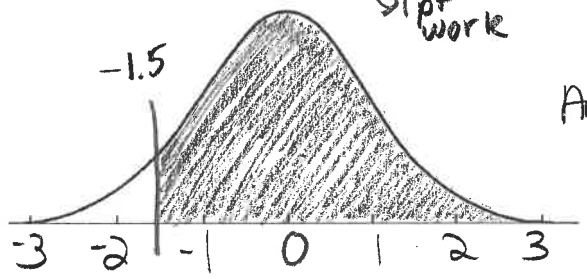
b) *(2 pts.)* Approximately **95%** of the students sleep between 4.4 **hours** and 9.2 **hours**
(Put the smaller number first)
accept 4.46 and 9.14

c) *(4 pts.)* If a student sleeps 5 hours per night what percent of the people sleep more than her?
Show work and write your answers in the blanks provided.

i) (2 pts) **Z score =** -1.5 $Z = \frac{5-6.8}{1.2} = -1.5$ 1pt work, 1pt answer

ii) (2 pts) 93.5 **% sleep more** *(Show work by marking Z on the curve and shading the correct area.)*
or 93.32

cont error from z-score



1pt work

1 pt answer

Area = 87 + 6.5 = 93.5

d) *(2 pts)* What **percentile is she in?** (no work necessary- use your work from part c) ← cont error

6.5th

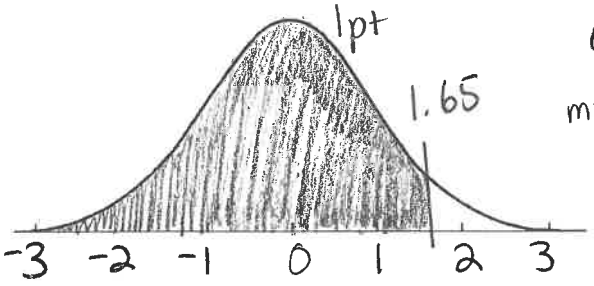
accept any answer that is 100 minus their answer in pt c.

e) *(6 pts)* If a student is in the 95th percentile how much does she sleep? Fill in the 3 blanks below.

95th percentile corresponds to middle area= 90 % (2 pts) → **Z score=** 1.65 (1pt) and she sleeps 8.78 **hours** 1pt answ

Mark the Z score on the curve below.

Show work for how you got the answer for the number of hours. Give hours to 2 decimal places.



cont error from middle area

1pt work

val = avg + (Z)(SD)
= 6.8 + (1.65)(1.2)
= 8.78

f) *(2 pts)* If a student is in the 5th percentile, how much does she sleep? 4.82 hours. (Hint: See part e)
Give answer to 2 decimal places.

val = 6.8 + (-1.65)(1.2)
= 4.82