Stat 100 Exam 1 Spring 2017

Question 1 pertains to the following study: *(14 pts.)*

(Hypothetical) A nation-wide study examining cholesterol levels in women found that women with 4 or more children had significantly higher levels of cholesterol than women with only 2 or fewer children.

a) *(2 pts.)*Is this study an observational study or a designed experiment? [✓ **i)** Observational Study] **ii)** Designed Experiment

b) *(2 pts.)* Does the study show that having more children raises cholesterol levels? Circle answer.

i) Yes, it shows definite causation although the exact causal mechanism isn't explained.
[✓ **ii)** No, it only shows that there is an association between more children and higher cholesterol. It does not show that one causes the other.]

c) *(8 pts.)* Since the treatment and control groups chose themselves (researchers can't assign someone to have 4 children) there may be differences between the 2 groups that could **confound** the results. Which of the following describe possible confounders? *Circle "Yes" if possible confounder given the description, "No" if not possible confounder given the description.*

2pts **i.** Stress- More children cause extra stress that could lead to general poor health including raised cholesterol levels. (For example, women with 4 children may have less time and energy to prepare healthy food and to exercise) **a)** Yes [✓ **b)** No]

2pts **ii.** Genetic Predisposition- Cholesterol levels are strongly affected by heredity, so the more children you have the more likely they will be to exhibit a propensity towards high cholesterol. **a)** Yes [✓ **b)** No]

2pts **iii.** Age- Women with more children tend to be older (on the average), and cholesterol rises with age. [✓ **a)** Yes] **b)** No

2pts **iv.** Income- Women with more children tend to be poorer (poorer to begin with, not as a result of more children), and cholesterol leve tend to be higher among the poor. [✓ **a)** Yes] **b)** No

d) *(2 pts.)* Suppose we think that geographical region is a confounder since both family size and cholesterol levels are strongly influenced by region (for example, South Carolina has both high levels of cholesterol and high birth rates while Colorado has both low levels of cholesterol and low birth rates.) How can we minimize the possible confounding effects of geographical region?

i) Break the population into subgroups by geographical region and compare the percentage of women with 4 or more children to the percentage with 2 or less children within each region.
ii) Break the population into subgroups by geographical region and compare the percentage of women with high cholesterol within each region.
[✓ **iii)** Break the population into subgroups by geographical region and compare the cholesterol levels of women who have 4 or more children to the cholesterol levels of women with 2 or less children within each region.]
iv) Break the population into subgroups by geographical region and compare the cholesterol levels of women who have 4 children in one region (say South Carolina) to the cholesterol levels of women who have 2 or less children in another region (say Colorado).

Question 2 *(4 pts.)* - 1 pt for each blank

A doctor measures the temperatures of 10 people with the flu. Their average is 100.8° F (Fahrenheit) with a SD of 1.6° F.

a) Suppose the thermometer was mis-calibrated and reading 1°F too low. To correct the situation the doctor added 1°F to each of the 10 temperatures making the new average 101.8 °F and the new SD 1.6 °F.

b) If the **original** temperatures were converted to another type of units by multiplying each temperature by 3.5, the new average would be 352.8 ° and the new SD would be 5.6 °

Question 3 (8 pts.) pertains to the following list of numbers: 2, 3, 0, 4, 6 0, 2, 3, 4, 6

a) (2 pts.) The average is 3 and the median is 3.

b) (4 pts.) The deviations from the average are: -1, 0, -3, 1, 3 and their sum must = 0. (Check that it does)

c) (2 pts.) Compute the SD. Show work starting from the results you got in part (b). Circle answer.

$$\sqrt{\frac{1+0+9+1+9}{5}} = \boxed{2}$$

1pt work → 1 pt answer

✓

Question 4 *(4 pts.)*
Do students learn better in Stat 100 L (in person sections) or in Stat 100 ONL (online sections)?
Last semester we compared the grade distributions of the two groups and there were no significant differences.

a)Can we conclude that it doesn't matter which section students choose to enroll in, they'll do equally well in either one?
Choose one.

i) Yes, since everything is exactly the same between the two sections (same homework, same exams, etc.) except for the treatment (whether you're watching the lectures in class or on video), there are no confounders.

ii) Yes, as long as everyone in the in-person sections attended class regularly the conclusion is valid. But not everyone did, so the results are likely to be biased against the in class section.

2pts

[✓ **iii)**] No since students themselves chose which section to enroll in there may be other differences between the 2 groups that are confounding the results. If the 2 groups are unbalanced to begin with, balanced results at the end are not conclusive.

b) We plan to do an experiment to help us decide which method helps Stat 100 students learn the best. We randomly select 40 students from next semester's combined Stat 100 rosters to participate.

Then we randomly assign 20 students to attend a short stats lecture given by Karle and 20 to watch the same lecture on video. Two days later, everyone will take the same quiz and we'll compare results.

2 pts

But after we do the randomization we notice that just by the luck of the draw, the in person group ended up with many more girls than the online group. What should we do? *Choose one.*

i) Move the extra girls to the other group so that both have the same percentage of girls.

ii) Keep the randomized groups, there's bound to be more girls in one of the groups because there's more girls in Stat 100!

[✓ **iii)**] Redo the randomization but this time randomize separately for girls and boys. Randomly select half the girls for the in person group and half for the online. Do the same with the boys.

iv) Randomization doesn't work with small samples sizes, it's better to try to match the groups as much as possible by choosing the groups.

Question 5 *(6 pts.)*
Suppose Homer Simpson and Prof Simpson each teach AP Stats. At the end of each year all their students take the AP stats exam. If they pass the test (score 3 or above) they'll earn college credit. If they fail the exam (score below 3) they won't get college credit. All students are classified into 2 groups by math background: advanced (have taken calculus) and regular (haven't taken calculus). The advanced students generally do better than the regular students on the AP stats exam. The table below gives the results for both teachers (for the past 10 years). The school is concerned that the overall percentage passing in Homer Simpson's class is lower than in Professor Simpson's class. (Percents are rounded to the nearest integer.)

	Professor Simpson				Homer Simpson		
	# Pass	# Fail	% Pass		# Pass	# Fail	% Pass
Advanced	300	100	75		85	15	85
Regular	30	170	15		50	150	25
Total	330	270	55		135	165	45

a) (2 pts.)Which **2 percentages** on the table are the most relevant for **advanced** students to compare in deciding which class gives them the best chance of passing? 75 % vs 85 % order doesn't matter

Which teacher should they choose? **i)** Prof Simpson [✓ **ii)**]Homer Simpson

b) (2 pts.)Which **2 percentages** on the table are the most relevant for **regular** students to compare in deciding which class gives them the best chance of passing? 15 % vs 25 % order doesn't matter

Which teacher should they choose ? **i)** Prof Simpson [✓ **ii)**] Homer Simpson

c) (2 pts.) If you want to improve your chances of passing the AP stats class and you're either advanced or regular, which teacher should you choose? **i)** Prof Simpson [✓ **ii)**] Homer Simpson **iii)** It depends on whether you're advanced or regular

Question 6 pertains to the following study: *(6 pts.)*
A study was done to see whether telephone counseling could help low-income people quit smoking. The subjects were 500 adult smokers with Medicaid managed care insurance. Half the subjects were randomly assigned to treatment and half to control. In the treatment group, the subjects were given a physician's examination (including advice and pamphlets on how to quit smoking), along with monthly telephone counseling sessions by trained nurses. The control group was given the same treatment without the monthly telephone counseling sessions. All the subjects were followed for 12 months and their smoking rates were compared.

a) Based only on the information above, this study is an example of .... *Choose one:*

i) Observational Study

2 pts [✓ **ii)**] Randomized Controlled Experiment

iii) Non-Randomized Controlled Experiment

b) Which of the following could confound the results? *Choose one:*

i) Unbalanced groups--The control group was not given the same opportunity to receive monthly telephone counseling, putting them at a severe disadvantage through no fault of their own.

2 pts **ii)** Differential Willpower--Physicians and counselors can only do so much, whether someone quits smoking or not depends a lot on will power. Differences in will power are likely to confound the results.

iii) Income—Low-income people are both more likely to receive Medicaid and more likely to be smokers.

iv) All of the above.

[✓ **v)**] None of the above

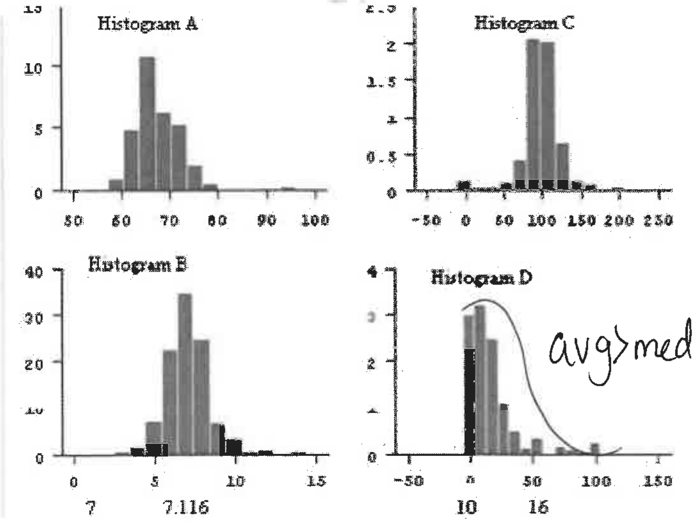
c) Even though the nurses tried to call everyone in the treatment group, only about 2/3 of the subject actually answered the phone and listened to the counseling. The other 1/3 either never answered the phone or refused the counseling. Should the researchers compare the smoking rate of everyone in the treatment group to the controls? Or should they just compare those who accepted the telephone counseling to the controls? *Choose one:*

2 pts **i)** They should just compare those who accepted telephone counseling to the controls, since the counseling cannot help those who refuse to hear it.

[✓ **ii)**] They should compare everyone assigned to treatment to everyone assigned to control, otherwise the treatment group will consist of a self-selected group (those who answered the phone and listened) which could confound the results.

iii) They should compare those in the treatment group who accepted counseling to those who refused counseling, since both groups were given the same opportunity for counseling.

Question 7: *(8 pts)* Below are 4 histograms representing 4 variables from our class survey: fastest speed ever driven in mph, height in inches, # hours of sleep per night, and # times you check the Facebook per week.



a) Which histogram represents height? A 2 pts

b) Which histogram represents # hours of sleep? B 2 pts

c) Below Histograms B and D are 2 numbers. One is the average and the other is the median. .

i) What is the median of Histogram B? 2 pts
Circle answer: [✓ 7] 7.116

ii) What is the average of Histogram D? 2 pts
Circle answer: 10 [✓ 16]

Question 8 *(12 pts.)* pertains to the table of blood pressure for subjects in a study. The table gives the systolic blood pressure interval and the height of the block of the histogram over each interval. The first row says that 4% of the people had blood pressures between 90-100.

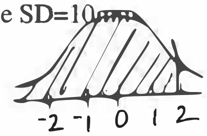
Interval (mm)	Height(% per mm)	Area (%)
90-100	0.4	4
100-105	1.0	5
105-110	1.4	7
110-115	3.0	15
115-120	3.4	17
120-125	2.6	13
125-130	2.4	12
130-135	2.0	10
135-140	1.6	8
140-150	0.6	6
150-160	0.3	3

a) (6 pts.) Fill in the 6 missing blanks in the table. 1pt for each blank

b) (2 pts.) Use the table to determine what percentage of people had blood pressures above 140? 9 %

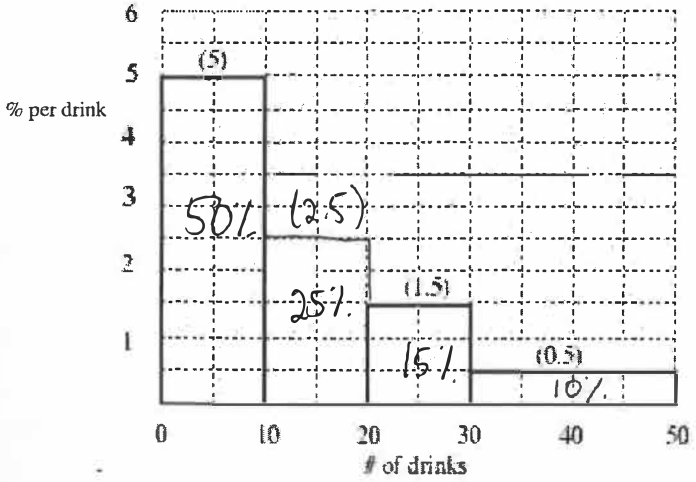
c) (2 pts.) Use the normal approximation (the normal curve) to estimate what percentage of the people had blood pressures below 140. Assume the average=120mm and the SD=10mm

97.5 % $Z = \frac{140-120}{10} = 2$



d) (2 pts.)The table and the normal approximation give very different percentages. Which correctly describes the people in this study? Circle one:
[✓ i) Table]   ii) Normal Approximation

Question 9 (18 pts.)
The figure below is a histogram for the number of alcoholic drinks consumed per week by 500 Stat 100 students (roughly based on a past semester's survey data). Class intervals include the left-endpoint but not the right. (For example someone who drinks 30 alcoholic beverages per week would fall in the 30-50 block not the 20-30 block.) The height of each block is given in parentheses. The block over the 10-20 drink interval is missing.



a) (3 pts.) What percentage of the subjects fell in the in the following intervals?
0-10 drinks 50 % 20-30 drinks 15 % 30-50 drinks 10 %

b) (2 pts.) The block over the 10-20 drink interval is missing. How tall must it be? 2.5 % per drink

c) (2 pts.)What is the median number of drinks? 10 drinks

d) (2 pts.)Which is smaller-- the average or the median? Or are they the same? ***Choose one:***
i) average [✓ **ii)** median] **iii)** the same **iv)** cannot be determined

e) (2 pts.) Did less people answer 0-10 drinks or 20-50 drinks, or are they the same? ***Choose one:***
i) 0-10 [✓ **ii)** 20-50] **iii)** Same

f) (2 pts.)What percentage of the subjects reported drinking 25 drinks per week? 1.5 % (Assume an even distribution throughout the intervals.)

g) (3 pts.) The maximum possible answer on the survey was 50 drinks. But suppose some of the students who answered 50 would have answered higher (all the way up to 70) if given the option. If those students could change their answers to numbers past 50, then how would that affect the average, the median, and the SD?

i) The average would ….. **circle one:** [✓ increase] decrease stay the same

ii) The median would….. **circle one:** increase decrease [✓ stay the same]

iii) The SD would…… **circle one:** [✓ increase] decrease stay the same

h) (2 pts.) How many drinks would a student have to drink to be in the 25th percentile? 5

Question 10 (19 pts.) According to our survey data, the histogram for the weights of the 526 women in this class is close to the normal curve with an average of 136 lbs. and a SD of 24 lbs. (You may round z scores to fit the closest line on the table and you may round percentages on the table to the nearest whole number.)

a) (2 pts.) About 68% of the women are between 112 pounds and 160 pounds.
(Fill in the blanks with weights, NOT z scores)

b) (2 pts.) If a student is 0.5 SD's below average, what is their z-score? −0.5

c) (5 pts.) Approximately, what percent of the females in the class are between 100 and 172 lbs?

i) (2 pts.) Translate interval into Z scores −1.5 **to** 1.5

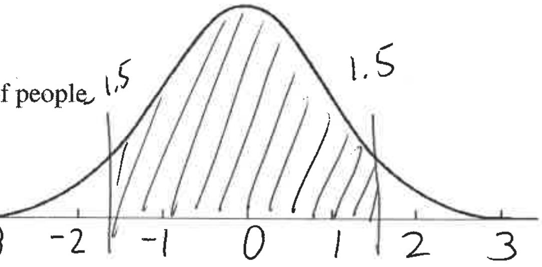
$z = \frac{100-136}{24} = -1.5$ $\frac{172-136}{24} = 1.5$

ii) (1 pt.) Mark the Z score correctly on the curve

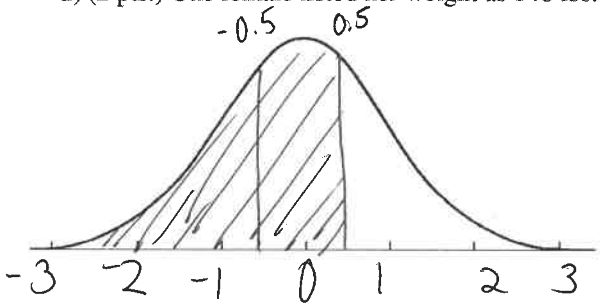
(1 pt.) Shade the region representing the percent of people who are between 100 and 172 lbs. (1 pt)

(1 pt.) Calculate the percent 87 %

accept between 86+87



d) (2 pts.) One female listed her weight as 148 lbs. What percentile is she in? (Show work below.)



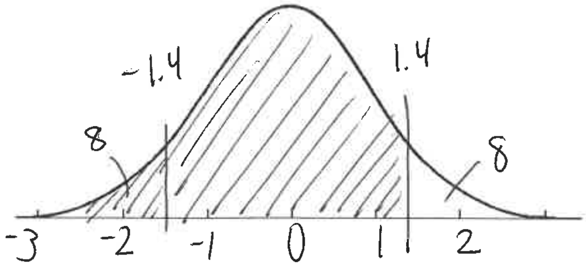
Z score= 0.5 and percentile = 69

$z = \frac{148-136}{24} = 0.5$ 38+31

e) (4 pts.) One female is in the 92nd percentile. How much does she weigh? Show work below.
Mark the 92nd percentile on the curve below, find the corresponding *z*-score, and the corresponding weight. ***Show work.***

92nd percentile corresponds to middle area= 84 %, → **Z score=** 1.4 and **weight =** 169.6 lbs.
(3 pts. + 1 for shading)

val= 136+(1.4)(24)



f) (4 pts.) If a student is in the 8th percentile, how much does she weigh? (Hint: You don't need to re-draw the curve from part (e), no need to show work.)

Z score= −1.4 **weight =** 102.4 lbs.

2pts 2 pts

val= 136−1.4(24)