Stat Page 1

Stat 100 Exam 1 Fall 2014

Question 1 (6 points)

A recent study tested whether the prayers of a large group of strangers could help patients undergoing a critical heart surgery. The subjects were 18,000 patients undergoing coronary artery bypass surgery. The patients were ***randomly*** assigned to 2 groups. One group received prayers from 3000 people in another city and the other group did not. ***Neither the patients nor the doctors treating them knew who was in which group.***

At the end of the study, there were no significant differences in recovery rates between the two groups. They were about the same.

a) Which of the following statements best describes this study? (2 points)

Choose one:

i) This was a randomized controlled experiment, but not double blind.

ii) This was an observational study.

(iii) This was a randomized controlled double blind experiment. ideal design

iv) This was a non-randomized experiment with historical controls.

b) Which conclusion is best supported by this study? (2 points)

Choose one:

i) This study is *strong evidence* that strangers' prayers speed up the recovery rate for patients undergoing heart surgery.

(ii) This study shows that there is *no strong evidence* of any difference in recovery rates between heart surgery patients who receive strangers' prayers and those who do not.

iii) This study only shows that there is an *association* between prayer and faster recovery rates. A confounder may be present.

iv) It's *impossible to conclude anything* from this study since the prayers were from people in another city.

c) Which of the following are possible confounders that may mix up the results? (2 points)

Choose one:

i) Pre-existing health conditions- Those in the prayer group were sicker and may have needed more prayers.

ii) Income- Those in the prayer group were wealthier and may have paid other people to pray for them as well.

iii) Inflated sense of protection- Those in the prayer group may have felt more protected and therefore, they didn't take as good of care of themselves.

iv) All of the above are confounders.

(v) None of the above are confounders. randomization eliminates any confounders

Question 2 pertains to the following study: (4 points)

Suppose I wanted to test whether reading the Freedman textbook before each lecture helped students do better in STAT 100. I decide to randomly assign half my students to the treatment group and half to the control. In the treatment group, the students were given the textbook and sent e-mails before each class reminding them which chapters to read. Those in the control group were not given the book and they were not sent emails about what to read.

a) This study is an example of... (2 points)

Choose one:

i) An observational study.

(ii) A randomized controlled experiment that was not double blind.

iii) A randomized controlled double blind experiment.

iv) A study with no controls.

b) Even though I encouraged everyone in the treatment group to read the book, only about 80% of them actually did. The other 20% didn't bother. What comparison should I make to analyze my results? (2 points) Good experiment gone bad

Choose one:

i) I should compare only those who actually did the reading (80%) to the controls, since reading can only help those who do it.

ii) I should compare those in the treatment group who actually did the reading to those who chose not to, since both groups were given the same opportunities and encouragement to learn.

(iii) I should compare everyone assigned to treatment to everyone assigned to control, otherwise the treatment group might consist of a different type of population (students who study harder in general) than the controls (the slackers) which could confound the results.

Question 3 (11 points)

A recent study done in the United States found that first time parents who give birth to a girl are more likely to get a divorce than first time parents who give birth to a boy. Researchers looked at divorce rates of first time parents and compared the treatment group (first time parents who gave birth to girls) to the control group (first time parents who gave birth to boys) and found that the treatment group had significantly higher divorce rates than the control group.

a) Is this study an observational study or a designed experiment? (i) Observational Study ii) Designed Experiment (1 point)

b) Were there randomized controls? i) YES (ii) NO (1 point)

c) Was the study done double-blind? i) YES (ii) NO (1 point)

d) After this study was done, the newspapers wanted to publish the results. They decided on using this headline: "Oh no, it's a girl! New study finds that daughters cause divorce!" Based only on the results of this study, do you think this headline is the most appropriate? *Choose one:* (2 points)

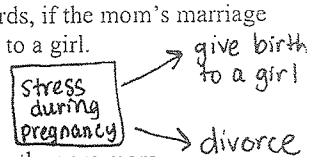
i) Yes, our results show that having a girl first *causes* parents to get divorced.

(ii) No, we cannot conclude definite causation. We can only conclude that having a girl first *is associated with* higher divorce rates; there may be a confounder present.

iii) No, clearly there is no relationship between having a girl first and divorce rate.

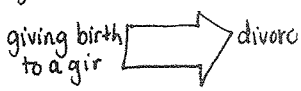
e) Below are either: confounders, causal links, or neither. *Hint: Each term is only used once!* (6 points)

i) Stress During Pregnancy: Stress during pregnancy certainly can lead to divorce after the baby is born. What's less obvious is that female embryos survive stress during pregnancy better than male embryos do. In other words, if the mom's marriage is falling apart while she's pregnant, she's both more likely to get divorced later and to give birth to a girl.



Choose one: i) Causal Link (ii) Confounder iii) Neither

ii) Father-Son Bonding: Fathers tend to bond more with their sons than with their daughters. Hence, they are more willing to stick around and try to make their marriage work if their first born is a boy and more likely to divorce if their first born is a girl. Explains why parents of first born girls are more likely to get divorced



Choose one: (i) Causal Link ii) Confounder iii) Neither

iii) Divorce rates are becoming higher and higher in the U.S. No mention of baby gender

Choose one: i) Causal Link ii) Confounder (iii) Neither

Question 4 (4 points)

Classify the following variables as quantitative-continuous, quantitative-discrete, or qualitative (categorical):

a) College debt when students graduate
(i) Quantitative (Continuous) ii) Quantitative (Discrete) iii) Qualitative

b) Number of A+'s you get in your college career.
i) Quantitative (Continuous) (ii) Quantitative (Discrete) iii) Qualitative

c) UIN (university ID number)
i) Quantitative (Continuous) ii) Quantitative (Discrete) (iii) Qualitative

d) Feelings before taking a Stat 100 Exam
i) Quantitative (Continuous) ii) Quantitative (Discrete) (iii) Qualitative

Question 5 (4 points)
A recent study of 76,000 students in 500 high schools and 225 middle schools nationwide was done to see if testing for drugs at school decreased drug use. The study compared the reported drug use in those schools that had ***chosen*** to implement a drug testing policy to those schools that had not. In each school, students filled out anonymous questionnaires asking them whether or not they used drugs. The study found that there was a difference in drug use rate. The rate was lower in schools that tested for drugs compared to schools that did not test for drugs.

a) This study is an example of.... *Choose one:* (1 point)

(i) Observational Study

ii) Randomized controlled experiment

b) Based solely on the results of this study, which of the following is most appropriate? (1 point)

i) This study ***proves*** that drug testing is effective in reducing student drug use.

ii) This study ***proves*** that drug testing is not effective in reducing student drug use.

(iii) This study ***suggests*** that drug testing may be effective in reducing student drug use.

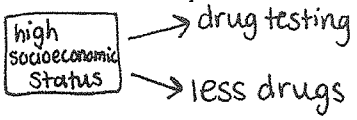
iv) This study ***proves*** that drug testing is not effective in reducing self-reported use of drugs.

c) Which of the following statements could confound (mix up) the results? (2 points)

i) Scare Tactics: Students at schools that perform drug testing may be afraid of getting caught with drugs and therefore, less likely to do them due to fear of failing the drug test. causal link

(ii) High Socioeconomic Status: Wealthier schools may be more likely to afford the drug testing and also may have students that come from higher income homes where parents educate them on the negative effects of drugs and why they shouldn't do them.

iii) Unequal sample sizes: If the sample sizes are unequal, this will mess up our rates and confound the results.



Question 6 (5 pts.)
A study compared the success rate of 2 treatments for wrinkles. Patients were classified as having either deep wrinkles or fine lines. Deep wrinkles are more severe and difficult to treat. The table below gives the results.

	Wrinkle Therapy Factor (WTF)				Lines Off Lotion (LOL)		
	# Successes	# Failures	% Success		# Successes	# Failures	% Pass
Deep Wrinkles	192	71	73%		55	25	69%
Fine Lines	81	6	93%		234	36	87%
Total	273	77	78%		289	61	83%

a) Which treatment has a higher success rate for deep wrinkles? (1 point)
(i) WTF ii) LOL iii) cannot be determined from the information given

b) Which group has a higher success rate for fine lines? (1 point)
(i) WTF ii) LOL iii) cannot be determined from the information given

c) Which treatment has the higher overall success rate (combining those who have deep wrinkles and fine lines)? (1 point)
i) WTF (ii) LOL iii) cannot be determined from the information given

d) What conclusion is best supported by these results? (2 points)

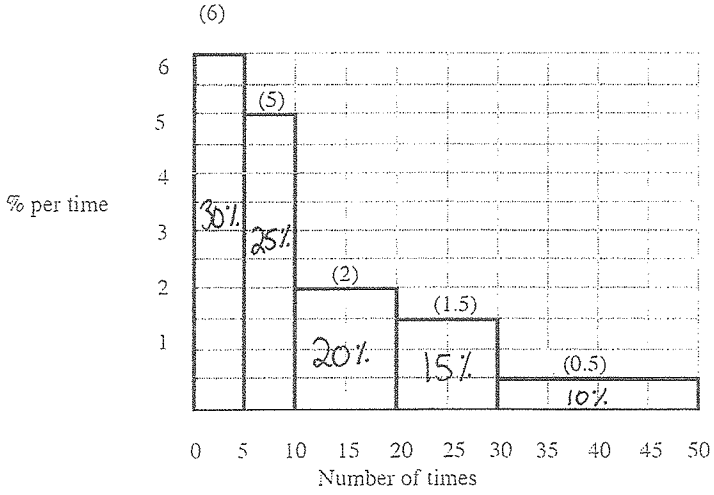
i) If you look at the overall results, clearly LOL is better for people with both types of wrinkles. We want to compare everyone in the treatment group to everyone in the control group. never w/ obs. studies

ii) The success rates of the two treatments depend on the doctor who is giving them.

(iii) If you have deep wrinkles, WTF is the better treatment and if you have fine lines, WTF is also the better treatment. Looking at the overall percentages is misleading because the groups aren't balanced. Simpson's Paradox!

Question 7 (14 points)

The figure below is a histogram for the number of times students use their cell phones to send or receive text messages in Stat 100 class last semester. The height of each block is given in parentheses. (Assume an equal distribution throughout each interval.)



long right hand tail

Area = Height × Width

a) 25% of the students text-messaged 5-10 times. What percentage of the students text messaged... (4 points)

i. 0-5 times 30 % ii. 10-20 times 20 % iii. 20-30 times 15 % iv. 30-50 times 10 %

b) The median number of study hours is closest to ..... (2 points) i) 5 ii) 6 (iii) 9 iv) 15 v) 20

c) The average is > the median. (2 points) i) less than (ii) greater than iii) equal to iv) cannot be determined

d) Did more people text 0-10 times or 20-50 times, or are they the same? (1 point) (i) More 0-10 ii) More 20-50 iii) Same

55% 25%

e) The 90th percentile is (1 point) i) 5 ii) 20 (iii) 30 iv) 40 v) 45 90% is to the left

f) If everyone had text-messaged 10 more times this semester, would the average, median, and SD all increase by 10? (2 points)

i. No, only the average would increase by 10.
ii. No, only the median would increase by 10.
iii. No, only the SD would increase by 10.
(iv.) No, the average and median would increase by 10, but the SD would stay the same.
v. Yes, the average, median, and SD would all increase by 10.

g) If everyone had text messaged 2 times more this semester (their texts were multiplied by 2), would the average, median, and SD all be multiplied by 2? (2 points)

i. No, only the average would be multiplied by 2.
ii. No, only the median would be multiplied by 2.
iii. No, only the SD would be multiplied by 2.
iv. No, the average and median would be multiplied by 2, but the SD would stay the same.
(v.) Yes, the average, median, and SD would all be multiplied by 2.

Question 8 (7 points) This question pertains to the following list of 5 numbers: -2, 1, 0, 5, 6 $\frac{-2+1+0+5+6}{5} = \frac{10}{5} = 2$

a) The average is 2, The median is 1. -2, 0, 1, 5, 6 (2 points)

b) The deviations from the average are -4, -2, -1, 3, 4 (List them in order from smallest to largest). (2 points)

c) The sum of the deviations from the average **should** = 0. *Fill in the blank with a number.* (1 point)

d) Compute the Standard Deviation. (2 points). Round your answer to 2 decimal places. ***Show your work. You may start with the deviations you found in part (b).***

Square deviations: $(-4)^2$ $(-2)^2$ $(-1)^2$ 3^2 4^2

16 4 1 9 16

average of squared deviations $\frac{16+4+1+9+16}{5} = 9.2$

square root: $\sqrt{9.2} = 3.03$

Stat Page 5

Stat 100 Exam 1 Fall 2014

Question 9 (6 points)

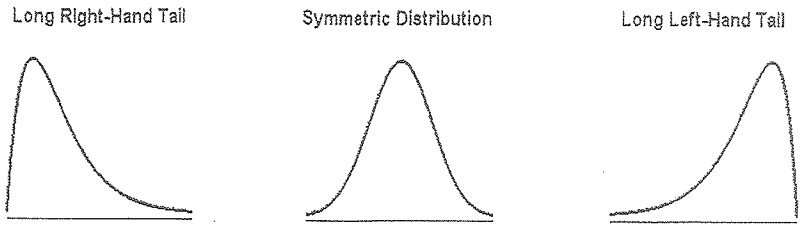
a) For which of the following histograms is it reasonable to use the Normal Approximation?

i. Long Right Hand Tail Only

ii. Long Left Hand Tail Only

(iii.) Symmetric Only

iv. All of them



For the following data sets below, check whether you think the histogram would have a long left-hand tail, long right-hand tail, or be fairly symmetrical. *Next to the Data Set Check the box that best describes its histogram.*

Data Set	Long Right Hand Tail	Fairly Symmetrical	Long Left Hand Tail
b)Exam scores where the median is 90 but the average is only 80. ave< med			✓
c)Height of all female U of I students.		✓	
d)Age at death in the U.S. due to natural causes.			✓
e)Exam scores where the median and average are about the same.		✓	

Question 10 (17 points)

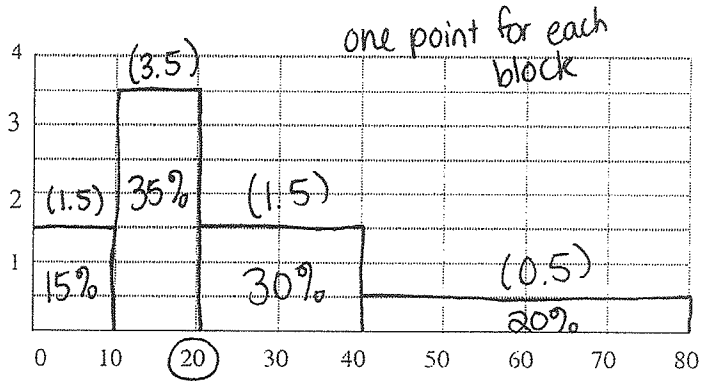
A distribution table for the number of minutes parents spent reading to their pre-school kids each day is shown below.

a) Fill in the column for height and then draw the histogram below. (8 points)

Height = Area / width

Minutes	%	Height (% per minute)
0-10	15	1.5
10-20	35	3.5
20-40	30	1.5
40-80	20	0.5

1 pt. each



b) What is the median number of minutes? 20 (2 points)

c) Is the average greater than, less than, or equal to the median?
Choose one: i) less than (ii) greater than iii) the same as (2 points) long right hand tail

d) Assuming an equal distribution throughout the interval, the percent of parents who reported spending exactly 42minutes reading to their child is closest to *Choose one:* (i) .5% ii) 1.5% iii) 20% iv) 40% (2 points)

e) What percent of parents spent 10-20 minutes reading to their kids? 35 % (2 points)

f) How many minutes corresponds to the 80th percentile? In other words 80% of the parents said their average time reading to their children is less than 40 minutes long. **(Fill in the blank with a number.)** (1 point) 80% to the left

(2 pts.)

Stat 100 Exam 1

Question 11 (11 points)

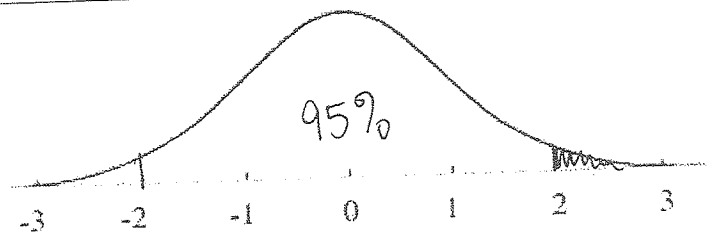
I challenged everyone in this class to do the ALS ice bucket challenge during the first week of school! Of the people who donated, I took data on how much money they donated after doing the challenge. I found that the amount of money they donated to the pretend ALS association followed the normal curve with an **average of $10 and an SD of $2.**

A) What percentage of people donated *more than $14* to the ALS association? $Z = \frac{\text{val-ave}}{SD} = \frac{14-10}{2} = \frac{4}{2} = 2$

i) First, convert $14 to a Z-score. **(Remember: average =$10 and SD =$2) Show work for full credit.** (2 points)

Z-score = 2

ii) Mark your Z-score accurately on the curve and divide the curve into a middle area and two tails. **Write the % of the middle area inside the middle area. Shade the area which corresponds to *over* 14 dollars.** (2 points)



1 pt. for correctly following ALL directions given in (ii) above

2.5 % of people who donated over $14.

Show work. 1 pt. for calculating correct %

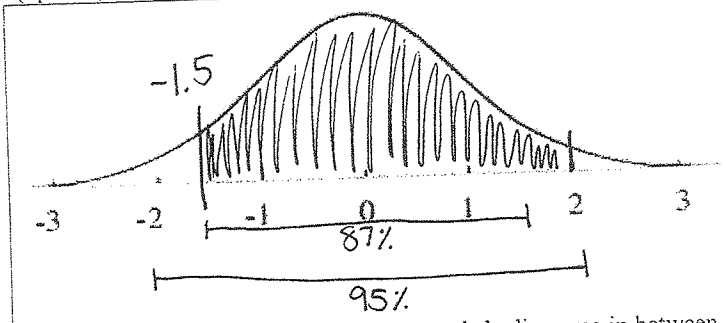
$\frac{100-95}{2} = 2.5$

B) What percent of people donated *between $7 and $14* to the ALS association?

i) Convert $7 to a Z-score. (You've already converted $14 to a Z score above) **Show work for full credit.** (2 points)

Z-score for 7 dollars = -1.5 $Z = \frac{\text{val-ave}}{SD} = \frac{7-10}{2} = -1.5$

ii) Mark both Z-scores on the curve & shade the area that corresponds to donating between $7 & $14. (2points)



1 pt for correctly marking both Z scores and shading area in between

91 % of people donated between $7 and $14.

Show work. 1 pt. for calculating correct %

$87 + \frac{95-87}{2} = 87 + 4 = 91$

C) If you're ***above*** average in donating to the ALS association, is your Z score positive or negative? (1 point)

Choose one: (i) positive ii) negative iii) not enough information given

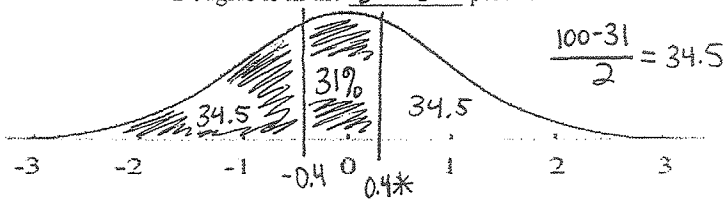
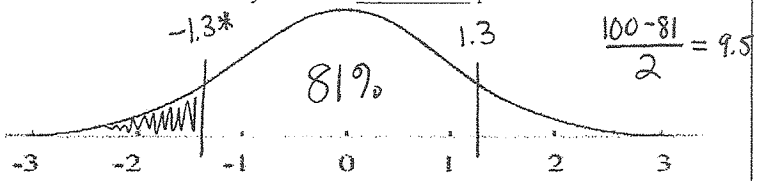
D) If you're exactly at the 50th percentile in donating then your Z score = 0 & you donated 10 dollars to the ALS association. (Fill in the two blanks with numbers.) (2 points)

Question 12 (11 points)

IQ scores among US adults follow the normal curve quite closely with an **average = 100 and SD= 15.** In the table below, you're either given an IQ score, the Z-score, or the percentile for 4 people. Fill in the remaining two blanks.

***DIRECTIONS* FOR PERCENTILE column** , mark the Z-score on the histogram, divide the curve into a middle area and two tails. Then mark the area in each of the 3 sections and shade the percentile.

(Round the middle area given in the table to the nearest whole number.)

IQ Score	Z-Score	Percentile (% of people who scored less)
Douglas had an IQ score of 106. $Z = \frac{val-ave}{SD} = \frac{106-100}{15} = 0.4$	Z = 0.4 (1pt) (show work)	Douglas is in the 65.5 percentile  1 pt for correctly following ALL starred ***directions*** above, 1 pt correct percentile. 34.5 + 31 = 65.5
Chauncey's IQ score was 80.5 (1pt) (show work, do not round) value = ave + z(SD) = 100 + (-1.3)(15) = 80.5	Z = -1.3	Chauncey is in the 9.5 percentile  1 pt for correctly following ALL starred ***directions*** above, 1 pt for correct percentile
Jacki's IQ score was 120.25 (1pt) (show work, do not round) value = ave + z(SD) = 100 + (1.35)(15) = 120.25	Z = 1.35 (1pt) positive (no work necessary)	Jacki is in the **91st percentile** (91% of the people scored less than her). What middle area should you look up on the normal table to find the correct Z score? 82 % (1pt)  1 pt for correctly following ALL starred ***directions*** above. know right hand tail = 9% therefore left hand tail = 9% middle area = 100 - 18 = 82%
Richard's IQ score was 79.75 likes (1pt) (show work, do not round) value = 100 + (-1.35)(15) = 79.75	Z = -1.35 (1pt) negative (no work necessary)	Richard is in the 9th percentile. HINT: No work is necessary for this problem (Just look your histogram above.) 9th percentile means there is 9% to the left of the Z-score