

**Questions 1-3 (4 points total)**

A recent study compared two different treatments for repairing a torn knee ligament. The subjects were 32 active, young adult volunteers who had acute knee ligament injuries. They were *randomly* divided into two groups: Group A received physical therapy and surgery, while Group B received only physical therapy (with the option to later have surgery). *No group received a fake surgery.* Evaluators who *were aware* of which patients were in which group rated the subjects on knee strength, stability, flexibility, etc, over a two year period and found that both groups said that they felt better, but there were no significant differences on any measure between the two groups.

- 1) (1 point) Which of the following statements best describes this study?
  - ☒ a) It's a randomized controlled experiment without a placebo and without "blind" evaluators.
  - ☐ b) It's an observational study.
  - ☐ c) It's a randomized controlled double-blind experiment.
  - ☐ d) It's a randomized controlled experiment with a placebo.
- 2) (2 points) Since there were only 32 subjects in this study, after the random division, out of the 16 people in group A, only 5 of them were female. What's the best method that the researcher could use to prevent this?
  - ☐ a) There is nothing that the researcher could do. They're just going to have to use the groups that they got.
  - ☐ b) The researcher could allow the subjects to decide if they want to be in Group A or Group B.
  - ☒ c) The researcher could "block" the subjects based on gender first, then randomly assign half of the males to Group A and half to Group B. They would then do the same thing with the females.
  - ☐ d) The researcher could hand pick the groups to have an equal amount of males and females.
- 3) (1 point) Out of the following, which represents the best improvement for this study?
  - ☐ a) Do not provide physical therapy to everyone in Group A since surgery is not given to everyone in Group B.
  - ☐ b) Allow the subjects, in consultation with their doctors, to choose whether to join Group A or Group B.
  - ☒ c) Make sure that the evaluators of the study are not aware of which group the subjects are in.
  - ☐ d) Give everyone in Group A and Group B a fake surgery.

**Questions 4-5 (4 points total)**

A study published in the Daily Illini last semester looked at whether or not using a laptop to take notes helped students do better on exams. The study looked at 327 student participants from Princeton University and UCLA. Students were asked to watch a lecture on a screen and take notes how they prefer to when given the option of using either a notebook and a pencil or a laptop. The students were then tested on both factual and conceptual information 30 minutes after the lecture. **Results:** Researchers found that using laptops for note-taking can have negative effects on educational assessments. The students using laptops for notes did worse when answering both types of questions.

- 4) (2 points) Which of the following could be a potential causal link?
  - ☐ a) Laziness- Lazy students are more likely to want to use a laptop to take notes since typing is faster than writing and lazy students are probably not going to do well when answering questions after a lecture.
  - ☒ b) Access to Internet- Students who take notes with laptops have access to the internet and therefore are more likely to visit websites such as Facebook in class. Hence, they get distracted, miss some information, and then score poorly on the questions.
  - ☐ c) Timing- Perhaps the lecture was too long & the students lost focus and ended up scoring poorly on the questions.
  - ☐ d) Non-adherers- Although everyone was told to take notes, some people chose not to. These people are known as non-adherers and this is why they answered poorly on the comprehension questions after the lecture.
- 5) (2 points) Suppose I think that a confounder may be present in this study about note-taking. I think a potential confounder could be wealth. What is the best way to check if this is actually a confounder?
  - ☒ a) Break (stratify) the subjects into subgroups based on wealth: students who are not wealthy, moderately wealthy students, and very wealthy students. Look at these groups separately and see if the difference in test scores goes away. If so, you've found the confounder.
  - ☐ b) Break (stratify) the subjects into subgroups based on how they did on answering the questions. See if the students who did better are wealthier. If so, you've found the confounder.
  - ☐ c) There is no way that this study can have a confounder, it was randomized.
  - ☐ d) See if Simpson's Paradox is occurring and make sure to block the subjects at the beginning of the study.

**Questions 6-8 pertain to the following situation:** Two professors at Hogwarts School of Witchcraft and Wizardry both taught the same 2 Defense Against the Dark Arts & Potions courses. The table below gives the results of how their students evaluated them on a scale of 1-5 with 5 being the highest.

	Professor Snape		Professor Slughorn	
	# Times Course Taught	Average Rating	# Times Course Taught	Average Rating
Potions	9	4.5	1	5.0
Dark Arts Defense	1	1.5	9	4.0
Total	10	4.2	10	4.1

6. (1 point) Which professor had the higher average rating in the Dark Arts Defense Course? a) Snape b) Slughorn
7. (1 point) Which professor had a higher average rating in the Potions Course? a) Snape b) Slughorn
8. (2 points) Which teacher seems to do better on student evaluations? *Choose one.*
- a) Clearly, it's Professor Snape. He has a higher overall rating than Professor Slughorn. We want to look at overall ratings to keep the original randomization.
  - b) Here, we cannot make any conclusion because we don't know if the students actually went to class.
  - c) There must have been some type of mistake. There's no way that we can see data like this that seems like a paradox.
  - d) If you look at the two courses separately, we see that Professor Slughorn did better on the student evaluations. His overall rating must be lower because he taught the more difficult course (Defense Against the Dark Arts) 9 out of 10 times. In observational studies like this, we never want to look at overall ratings because there could be a confounder present.

**Questions 9-11 (5 points total)**

Let's say that the "Stat 100 drug" experiment that I talked about on the first day of class was real! I came up with a pill that I think will help students do better in Stat 100 and now I want to do an experiment to see if it works. Suppose I randomly assign half of my students to the treatment group and half to the control group. The treatment group gets the Stat 100 pill and the control group gets a sugar pill. Neither I nor the students know who is in each group. All students attended the same lectures, had the same exams, etc, and at the end of the semester, I compared the overall averages of the 2 groups. I ended up finding no significant difference in averages.

9. (1 point) This study is an example of? *Choose one.*
- a) A randomized controlled experiment without a placebo.
  - b) An observational study.
  - c) A randomized controlled double-blind experiment.
  - d) A non-randomized controlled experiment with a placebo.
10. (2 points) Which of the following could confound this study? *Choose one.*
- a) Previous math experience- Students who have taken a lot of math/stats classes, may do better than those who have not.
  - b) Gender- Some people believe that males are better at math/stats than females.
  - c) Health Problems- Some students may have health problems that will affect their performance in Stat 100 and they may think they need to take the Stat 100 drug.
  - d) Year in School- Seniors may be more likely to do better in Stat 100 than freshmen since they have more experience taking college classes. Also, seniors may be more likely to want to take a drug.
  - e) None of the above are confounders.
11. (2 points) Of course, not everyone assigned to both groups actually took their pills faithfully. People who chose not to take their pills are called "non-adherers." Which comparison should be used to best answer the question of whether or not my Stat 100 drug actually works in improving students' grades? *Choose one.*
- a) You should compare only those who actually took the drug to the control group since the drug can only help those who take it.
  - b) You should compare everyone assigned to treatment to everyone assigned to control. Otherwise, the treatment and control group may consist of two different types of populations which could confound our results.
  - c) You should compare the people in the treatment group who took the pill to those in the treatment group who didn't since both were given the option of taking the pill.
  - d) You should compare only the students in the class from both groups that you think would actually take the pill when you tell them.

Questions 12-17 (7 points total)

A study reported in Time Magazine claimed that people who abstain from drinking alcohol die sooner than those who drink moderately to heavily. The study tracked 1,824 subjects aged 55-65 for 20 years and found that those who didn't drink any alcohol at all had the highest death rate (69%), compared to only 41% for moderate drinkers and 60% for heavy drinkers.

12. (1 point) Based only on the information above, this study is an example of... **Choose one.**
- ☒ a) An observational study
  - b) A randomized controlled experiment without a placebo
  - c) A randomized controlled double-blind experiment
  - d) A non-randomized controlled experiment
13. (2 points) Based only on the information above, which statement is best? **Choose one.**
- a) This study is strong evidence that abstaining from alcohol **causes** people to die sooner.
  - b) This study shows that abstaining from alcohol is **associated with but definitely does not cause** people to die sooner.
  - ☒ c) This study only shows that abstaining from alcohol is **associated with** dying sooner; it doesn't show whether or not there's a causal relationship.
  - d) This study shows that there is no association between abstaining from alcohol and dying sooner.

Below are either confounders that mix up the study, causal links that explain the conclusion, or neither. **Circle which is which.**

14. (1 point) Health Problems- The non-drinking group may have included more people who are sick and can't drink for medical reasons as well as former alcoholics. People who are sick are more likely to abstain from drinking and die sooner.

- ☒ a) Confounder      b) Causal Link      c) Neither

15. (1 point) Cardiovascular Benefits- Alcohol consumption increases the good kind of cholesterol (HDL), which lowers ones risk for heart disease and causes people to live longer.

- a) Confounder      ☒ b) Causal Link      c) Neither

16. (1 point) Age- Older people are more likely to die sooner than younger people.

- a) Confounder      b) Causal Link      ☒ c) Neither

17. (1 point) Religion- The non-drinking group may have included more people whose faith prevents them from drinking.

- a) Confounder      b) Causal Link      ☒ c) Neither

Question 18 (a) - (e) pertain to the following: (13 points total)

Consider the following list of 6 numbers: 10, 8, 10, 8, 8, 4 4, 8, 8, 8, 10, 10

- a) (2 points) The average is 8.
- b) (2 points) The median is 8.
- c) (6 points) The deviations are -4, 0, 0, 0, 2, 2.
- d) (1 point) The sum of the deviations should **always** = 0 (Check that the deviations above sum to what they should)
- e) (2 points) The standard deviation is 2.

Show work. You may start by using the deviations you got in (c) above.

$$\frac{16 + 0 + 0 + 0 + 4 + 4}{6} = 4 \quad \sqrt{4} = 2$$

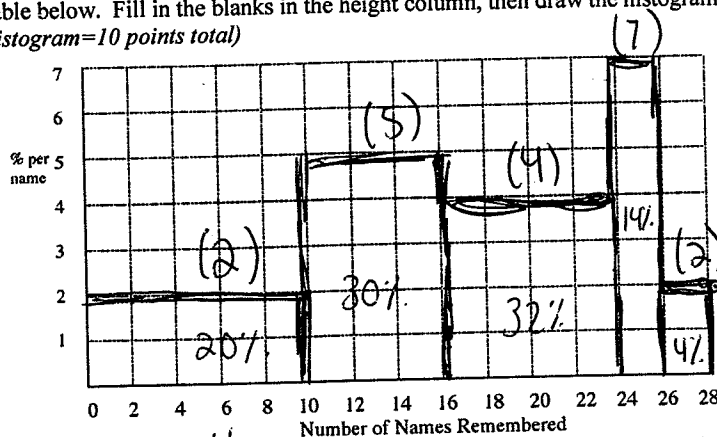
Question 19 (3 points total) Suppose you multiplied all the numbers on the list above by -10...

- a) The average would....
- i) be multiplied by 10.      ☒ ii) be multiplied by -10.      iii) decrease by 10.      iv) increase by 10.      v) stay the same.
- b) The median would....
- i) be multiplied by 10.      ☒ ii) be multiplied by -10.      iii) decrease by 10.      iv) increase by 10.      v) stay the same.
- c) The SD would....
- ☒ i) be multiplied by 10.      ii) be multiplied by -10.      iii) decrease by 10.      iv) increase by 10.      v) stay the same.

**Questions 20-23 (16 points total)**

A group of college seniors were asked the question: "How many names can you remember from your 1<sup>st</sup> grade class?" Their responses are summarized in the table below. Fill in the blanks in the height column, then draw the histogram on the grid below. (5 points+ 5 points for histogram=10 points total)

Names (Width)	% (Area)	Height (%/ name)
0-10	20	2
10-16	30	5
16-24	32	4
24-26	14	7
26-28	4	2



20. (1 point) The median number of names remembered is 16

21. The number of people who remember 0-10 names is \_\_\_\_\_ the number of people who remembered 24-28 names.  
(1 point) Choose one: i) Less Than ii) Greater Than iii) Equal To

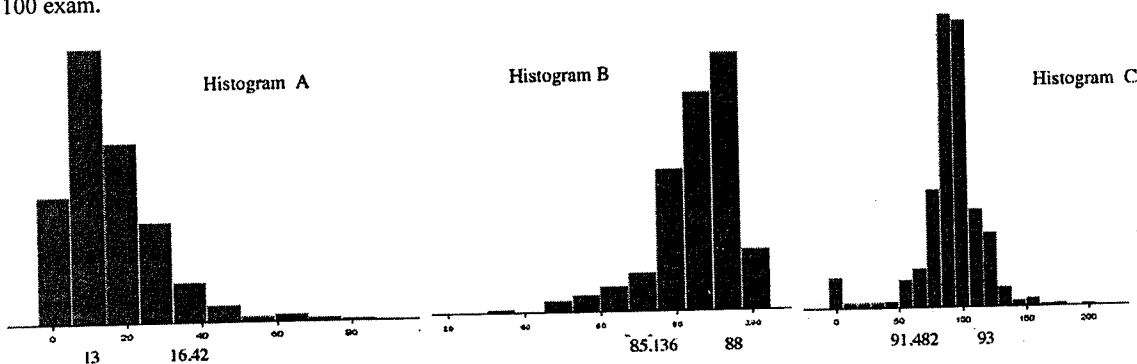
22. (1 point) If Ella is in the 20th percentile, then she remembered 10 names. (Fill in the blank with a #)

23. (3 points) If the people in the 26-28 block all forgot 2 names, how would the average, median, and SD be affected? Check the appropriate boxes below:

	Decrease	Increase	Stay the Same
i) Average would	✓		
ii) Median would			✓
iii) SD would ...	✓		

**Question 24 (6 points total)**

Below are 3 histograms. Two represent our survey responses to the 2 questions: "What is the fastest speed you've ever driven in mph?" and "How many pairs of shoes do you own?" The third represents exam scores from a previous stat 100 exam.



a) Which histogram is which? Fill in the 3 blanks with the correct letter (A, B, or C):

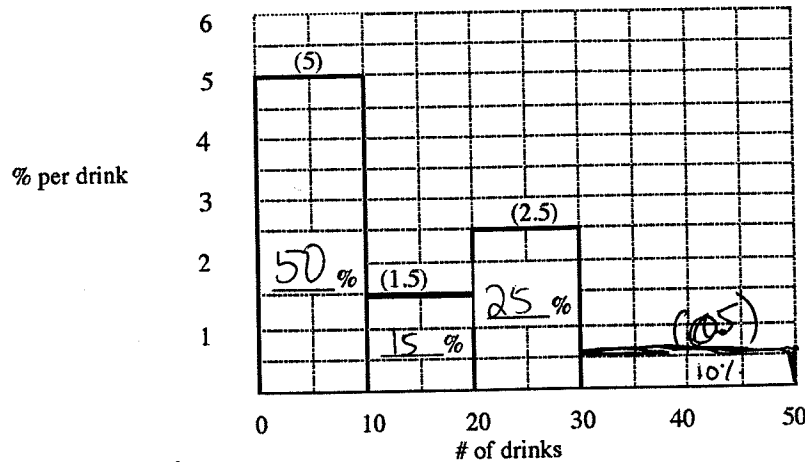
Histogram C represents Speed      Histogram B represents Exam Scores      Histogram A represents Shoes

b) Below each histogram are 2 numbers: one is the average & the other is the median. Fill in the 3 blanks with averages.

The average of Histogram A is 16.42, the average of Histogram B is 85.136, the average of Histogram C is 91.482

Question numbers 25-31 refer to the histogram below: (13 points total)

The figure below is a histogram for the number of alcoholic drinks consumed per week by Stat 100 students (roughly based on a past semester's survey data). The height of each block is given in parentheses and the block over the 30-50 drinks interval is missing.



25. (3 points) What percentage of students fell in 0-10 drinks, 10-20 drinks, and 20-30 drinks? Write the percentages below:

0-10 drinks 50      10-20 drinks 15      20-30 drinks 25

26. (2 points) The block over the 30-50 drinks interval is missing. Draw the missing block on the histogram.

The area of the 30-50 block is 100 - 90 = 10%. Fill in the blank with a percentage.

The height of the 30-50 block is 0.5. Fill in the blank with a number.

27. (2 points) The average is \_\_\_\_\_ the median.  
a) less than      **b) more than**      c) same as      d) none of the given choices

28. (1 point) The median is closest to  
a) 5      **b) 10**      c) 20      d) 30      e) none of the given choices

29. (2 points) What percentage of the subjects reported drinking exactly 22 drinks per week?  
Assume an even distribution throughout the interval.  
a) 5%      b) 1.5%      **c) 2.5%**      d) 25%      e) none of the given choices

30. (1 point) What percentage of subjects drank more than 20 drinks per week?  
**a) 35%**      b) 2.5%      c) 25%      d) 10%      e) none of the given choices

31. (2 points) Would it be appropriate to use the normal approximation for this dataset?  
a) Yes because when we convert to z-scores, our histogram will become normal.  
**b) No** because the histogram shows that the data clearly does not follow the normal approximation.  
c) No, this histogram already looks normal so we don't need to approximate it.  
d) Yes because we know this is normal data since it came from a survey.

**Question 32 (8 points total)**

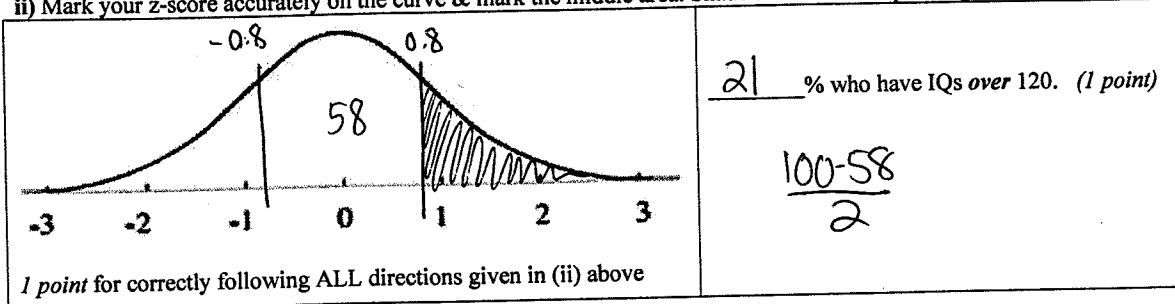
Suppose IQ scores among US adults follow the normal curve with an average of 104 and an SD of 20.

a) What percent of the population have IQ's over 120?

i) First, convert 120 to a Z-score. (Remember: average and SD are given above) Show work for full credit and round your z-score to 2 decimal places. (2 points)

$$Z\text{-score} = \underline{0.8} \quad \frac{120-104}{20}$$

ii) Mark your z-score accurately on the curve & mark the middle area. Shade the area corresponding to scoring over 120.

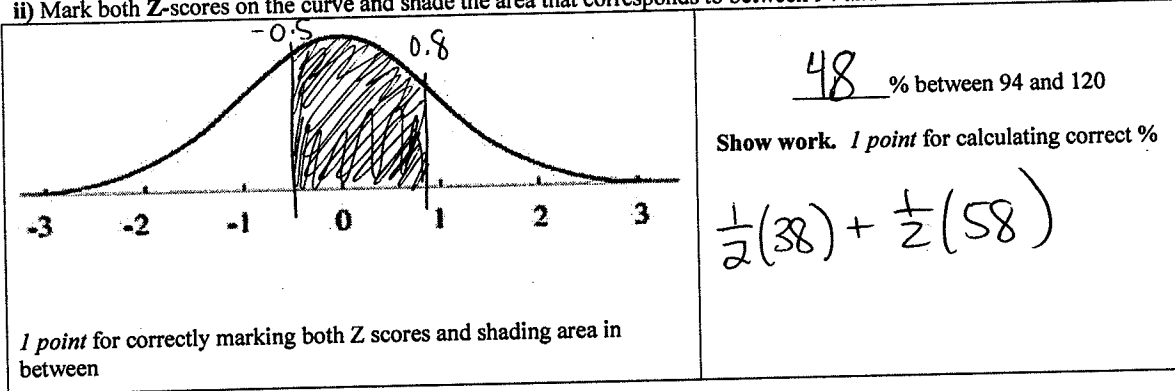


b) What percent of the population have IQ scores between 94 and 120?

i) Convert 94 to a Z-score. (You've already converted 120 to a Z score above) Show work for full credit. (2 points)

$$Z\text{-score for } 94 = \underline{-0.5} \quad \frac{94-104}{20}$$

ii) Mark both Z-scores on the curve and shade the area that corresponds to between 94 and 120.



Questions 33-37: For the first 3 questions, circle your answer. For the last two, fill in the blank with a number.

33. (1 point) If you have a z-score of 0, you are \_\_\_\_\_.  
 a) Exactly Average      b) Above Average      c) Below Average      d) Impossible to tell

34. (1 point) If you have a z-score less than 0, you are \_\_\_\_\_.  
 a) Exactly Average      b) Above Average      c) Below Average      d) Impossible to tell

35. (1 point) If you have a z-score greater than 0, you are \_\_\_\_\_.  
 a) Exactly Average      b) Above Average      c) Below Average      d) Impossible to tell

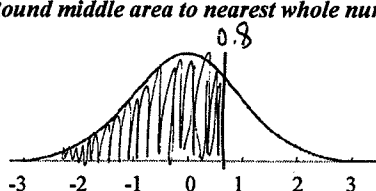
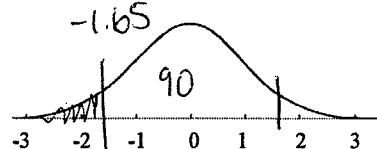
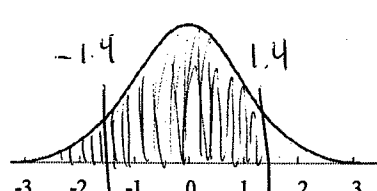
36. (1 point) If you are 2 standard deviations above the average, your z-score is 2.  
 Fill in the blank above with a number.

37. (1 point) If you are 2 standard deviations below average, your z-score is -2.  
 Fill in the blank above with a number.

**Question 38 (12 points total)**

According to previous survey data, male Stat 100 students have an average weight of 170 pounds and an SD of 30 pounds. The histogram of their weights is close to the normal curve. In the table below, you're either given a male's weight, a z-score, or a percentile and you have to fill in the missing blanks.

For all these problems, please round the areas given in the normal table to the nearest whole number.

Weight	Z score	Percentile
<p>Bernie weighs 194 pounds.</p> <p><math display="block">\frac{194 - 170}{30} = .8</math></p>	<p><math>Z = \underline{0.8}</math> (1 point)</p> <p>Show work:</p>	<p>Bernie is in the <u>79<sup>th</sup></u> percentile (Fill in the blank-1 point)</p> <p>Mark z-score on curve &amp; shade the appropriate percentile. (1 point)</p> <p>Round middle area to nearest whole number.</p> <p><math>58 + \frac{100 - 58}{2} = 79</math></p> 
<p>Donald's weight= <u>120.5</u> (Fill in the blank- 1 point)</p> <p>Show work:</p> <p>value = <math>170 + (-1.65)(30)</math> <math>= 120.5</math></p> <p>Do NOT round answer.</p>	<p><math>Z = \underline{-1.65}</math></p>	<p>Donald is in the <u>5<sup>th</sup></u> percentile. (Fill in the blank-1 point)</p> <p>Mark z-score on curve &amp; shade the appropriate percentile. (1 point)</p> <p>Round middle area to nearest whole number.</p> <p><math>\frac{100 - 90}{2} = 5</math></p> 
<p>Jeb's weight= <u>212</u> (Fill in the blank- 1 point)</p> <p>Show work:</p> <p>value = <math>170 + (1.4)(30)</math> <math>= 212</math></p> <p>Do NOT round answer.</p>	<p><math>Z = \underline{1.4}</math> (1 point)</p>	<p>Jeb is in the <u>92<sup>th</sup></u> percentile</p> <p>What middle area should you look up on the normal table to find the correct Z score? <u>84</u> % (Fill in blank- 1 point)</p> <p>If the middle area is between two lines on the table, use the closest line.</p> <p>Mark z-score on curve &amp; shade the appropriate percentile. (1 point)</p> 
<p>Joe's weight= <u>128</u> (Fill in the blank- 1 point)</p> <p>Show work:</p> <p>value = <math>170 + (-1.4)(30) = 128</math></p>	<p><math>Z = \underline{-1.4}</math> (1 point)</p>	<p>Joe is in the <u>8<sup>th</sup></u> percentile.</p> <p>Hint: No work is necessary.</p>