EXAM 1: Statistics 100

READ THE DIRECTIONS BELOW TWICE!

Cover Sheet Questions

1) What's your **name**? _____ _____

(Last name) (First name)

2) What's your **net ID** (email)? _____@illinois.edu

3) Which **section** are you in?

Circle one: **i)** L1 (Karle Flanagan In Person) **ii)** ONL (Karle Flanagan Online)

This test is ALL multiple choice. Circle all answers on this exam and fill in the corresponding bubble on your scantron. All questions have exactly one answer. If you circle/bubble in more than one answer, you will automatically be marked wrong. Make sure to circle the answers on this test and fill out your scantron. If you don't do both, you will get a 0.

SCANTRON Form Directions

- Print and bubble in your LAST NAME with **no spaces** starting in the left most column. Print your FIRST INITIAL in the right-most column.
- Print and bubble in your Student ID number (UIN) in the Student Number box.
- Print and bubble in your NET ID with **no spaces** in the NETWORK ID box.
- *No need to bubble in anything for Section*.

READ THIS: Failure to fill out your scantron correctly will result in a loss of 2 points on your exam!

WARNING- The exams look alike but you are sitting next to people who actually have a different version than you. Copying from anyone is equivalent to giving a signed confession.

All cheating including being caught with a non-permissible calculator or formula sheet will result in a 0 and an academic integrity violation on your University record.

Make sure you have all 11 pages including the normal table (60 questions).

There is NO CLASS on Friday!

Scores will be posted on Canvas by Monday at noon and exams will be returned in class next week. Online students may pick up their exam in 0060 Siebel Center for Design during office hours next week.

Questions 1-3 pertain to the following situation:
When I was in high school, I tore my ACL from playing basketball! Recently, I saw a study that 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. 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. 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 number of males and females.

3. 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 4-6 pertain to the following situation:
Over the last few months, several studies explored the potential effectiveness of hydroxychloroquine in the treatment of COVID-19. Here are two that were done:

Study 1: One hospital treating patients with COVID-19 gave their patients who came in and tested positive hydroxychloroquine and compared them to people at another hospital who tested positive and did not get hydroxychloroquine. They found that the patients receiving hydroxychloroquine had higher survival rates and shorter infection periods than those who didn't receive it.

Study 2: One hospital treating patients with COVID-19 took 50 volunteers who tested positive and randomly gave half of them hydroxychloroquine and the other half a sugar pill. The patients and the doctors didn't know who had the hydroxychloroquine and who had the sugar pill. They found that both groups experienced very similar survival rates and infection periods.

4. How would you classify the two studies?
   a. Study 1 is a randomized controlled experiment, Study 2 is an observational study.
   b. Study 2 is a randomized controlled experiment, Study 1 is an observational study.
   c. Studies 1 and 2 are both randomized controlled experiments.
   d. Studies 1 and 2 are both observational studies.

5. Which study is more likely to have confounders?
   a. Study 1 is more likely to have confounders.
   b. Study 2 is more likely to have confounders.
   c. Both studies are very likely to have confounders.
   d. Both studies shouldn't have any confounders.
   e. Impossible to tell.

6. Can we say with confidence that hydroxychloroquine is an effective treatment for COVID-19?
   a. Yes
   b. No

Questions 7-12 pertain to the following situation:
A study reported in Time Magazine claimed that people who abstain from drinking alcohol die sooner than those who drink, even sooner than those who drink 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.

7. Based only on the information above, this study is an example of…
   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.

8. Based only on the information above, which statement is best?
   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. ***Indicate which is which***.

9. 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

10. Cardiovascular Benefits- Alcohol consumption increases the good kind of cholesterol (HDL), which lowers one's risk for heart disease and causes people to live longer.

a. Confounder b. Causal Link c. Neither

11. Age- The non-drinking group may have included older people who are likely to die sooner.

a. Confounder b. Causal Link c. Neither

12. Religion- The non-drinking group may have included more people whose faith prevents them from drinking.

a. Confounder b. Causal Link c. Neither

Questions 13-16 pertain to the following list of numbers: 0, 7, 1, -8, -3, -9

13. What is the average of this list of numbers?
   a. -2
   b. -1.5
   c. -12
   d. -3.5

14. What is the median of this list of numbers? a) -1.5 b) -3.5 c) -2 d) -4.5

15. The sum of the deviations from the average should always be equal to ____.
a. 0
b. 1
c. 100
d. The average
e. Unable to determine from the information given

16. Compute the standard deviation.
a. 5.48
b. 30
c. 13.42
d. 6.78

Questions 17-24 pertain to the following situation:
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**
Bernie S weighs 194 pounds.	**17. Z= _____** a. 1 b. 0.95 c. -0.8 d. 0.7 e. 0.8	**18. *Bernie is in the*** _____ percentile Hint: Use the normal curve to answer this! a. 58th b. 21st c. 42nd d. 79th
19. Donald T's weight= _____ a. 231 b. 290.1 c. 120.5 d. 144.5	**Z= -1.65**	**20. *Donald is in the*** _____ percentile. *Hint: Use the normal curve to answer this!* a. 90th b. 10th c. 5th d. 95th
21. Joe B's weight= _____ a. 212 b. 236 c. 222.5 d. 268 e. 200	**22. Z= _____** a. 1 b. 1.75 c. 1.4 d. 0.10 e. 0.55	**23. *Joe is in the* 92nd percentile** What middle area should you look up on the normal table to find the correct Z score? _____% *Hint: Use the normal curve to answer this!* a. 92 b. 8 c. 42 d. 84 e. 68

24. Now, let's say another male Stat 100 student, Barack O. is in the 8th percentile for weight. What is his z-score?
   a. -1
   b. -1.4
   c. -1.75
   d. 1.4
   e. 1.75

Questions 25-28 pertain to the following situation:
Luke Skywalker conducted a survey on an island inhabited by Porgs, sea-dwelling birds, and found that the weights of female Porgs follow the normal curve with an ***average = 10 lbs*** and ***SD = 2 lbs***

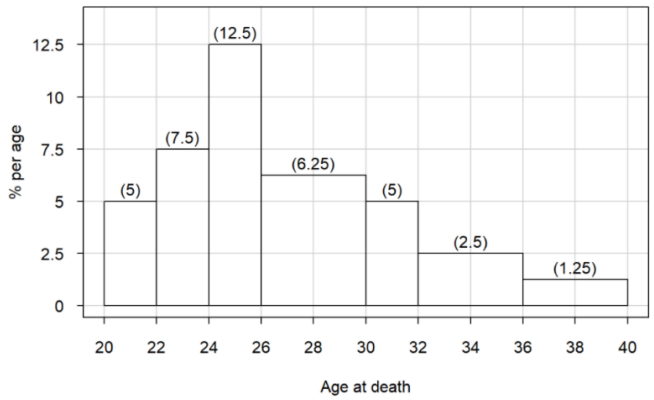
25. If a female Porg is above average in weight, is its z-score positive or negative?
   a. Positive
   b. Negative
   c. It will be zero
   d. Unable to determine from the information given

26. If a female Porg is exactly at the 50th percentile in weight, then its z-score is…
   a. 0.5
   b. 50
   c. 0
   d. 10
   e. 0.67

27. If a female Porg is exactly at the 50th percentile in weight, then its weight is…
   a. 10
   b. 2
   c. 15
   d. 11
   e. 5

28. If two female Porgs have the same z-score in absolute value, but opposite signs, then their percentiles must sum to ___%
   a. 0
   b. 100
   c. 50
   d. Unable to determine from the information given

Questions 29-36 pertain to the following situation:
In Star Trek, low-level cadet officers wear red Starfleet uniforms. The histogram below shows the age at death of a large population of cadet officers. The height of each block is given in parentheses. *Assume an even distribution throughout each interval.*



29. The median is _____ the average. a) Less than b) Equal to c) Greater than

30. If everyone lived 2 more years, the average would…
- a. Increase by 2 years
- b. Stay the same
- c. Decrease by 2 years
- d. Unable to determine from the information given

31. What percent of cadet officers died in the 24-26 interval?
- a. 2
- b. 12.5
- c. 25
- d. 50

32. The median is closest to _____.
- a. 26
- b. 30
- c. 12.5
- d. 24

33. The 25th percentile is _____.
- a. 24
- b. 30
- c. 22
- d. 26
- e. 7.5

34. The percent of the population who died at age 25 is closest to _____.
- a. 1
- b. 12.5
- c. 37.5
- d. 25

35. If everyone lived 2 more years, the SD would…
- a. Increase by 2 years
- b. Decrease by 2 years
- c. Stay the same
- d. Unable to determine from the information given

36. If you knew the average and the SD of the ages displayed in the histogram above, would it be appropriate to use the normal approximation to figure out what percentage of the ages fell within various intervals?
- a. No, the histogram of the ages is not close enough to following the normal curve; it has a long right-hand tail
- b. No, the histogram of the ages is not close enough to following the normal curve; it has a long left-hand tail
- c. Yes, the histogram of the ages follows a normal curve

The following situation pertains to questions 37-38

A study published in the Daily Illini a few years ago looked at whether or not using a tablet 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 normally would, 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 notetaking can have negative effects on educational assessments. The students using laptops for notes did worse when answering both types of questions.

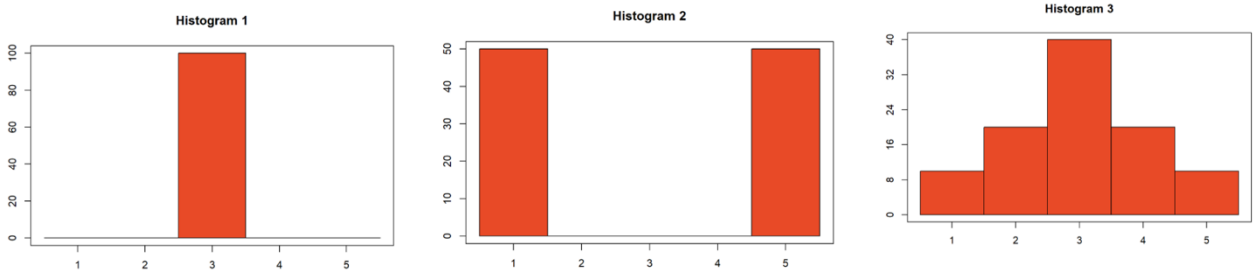
37. Which of the following could be a potential **causal link**?

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.

38. 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 39-41 pertain to the 3 histograms below.
Each histogram represents a data set of 10 whole numbers



39. Which histogram has an average of 3?

a. Histogram 1
b. Histogram 2
c. Histogram 3
d. None of the histograms
e. All of the histograms

40. Which histogram has an SD = 0?

a. Histogram 1
b. Histogram 2
c. Histogram 3
d. None of the histograms
e. All of the histograms

41. Which histogram has the largest SD?

a. Only histogram 1
b. Only histogram 2
c. Only histogram 3
d. Both histograms 1 and 2
e. Both histograms 2 and 3

Questions 42-43 pertain to the following situation:
Statistician Raccoon lives in the attic above the STAT 100 office. He records how many STAT 100 students come to office hours each day during exam week and the week after, as shown in the two sets of numbers:

During exam week: 18, 27, 0, 0, 0
The week after: 0, 2, 1, 0, 0

42. Statistician Raccoon doesn't have a calculator to do any calculations. What would he conclude about the averages of exam week compared to the week after?
   a. They have roughly the same average.
   b. The week after has a larger average.
   c. The exam week has a larger average.
   d. It is impossible to say without calculating the averages.

43. Statistician Raccoon doesn't have a calculator to do any calculations. What would he conclude about the standard deviations of exam week compared to the week after?
   a. They have roughly the same standard deviation.
   b. The exam week has a larger standard deviation.
   c. The week after has a larger standard deviation
   d. It is impossible to say without calculating standard deviations.

Questions 44-46 pertain to the following situation:
Bert thinks that romantic relationships may be better able to thrive and endure a longer period of time if each partner has a disposition for humor and laughter. To test this, he designed and performed an experiment on 85 romantic couples at UIUC of similar characteristics such as age, political beliefs, and cultural background. He asked them a series of questions related to how they feel about being laughed at and ridiculed, their relationship quality, and how they communicate and cope with disagreements. The couples were then randomly assigned to two groups. Group A couples were told to avoid laughing, using humor, and teasing when around each other for 30 days, while Group B couples were told to laugh, use humor, and tease when around each other for 30 days. At the end of the study, Bert asked each couple the same questions as the ones he asked in the beginning and measured the changes in their responses.

44. What are the subjects of Bert's experiment?
   a. The romantic couples.
   b. The change in the responses to Bert's questions.
   c. The ability for couple's relationships to thrive due to laughter and humor.
   d. The task to laugh and tease more or less when around each other.

45. What are the treatments?
   a. The romantic couples.
   b. The change in the responses to Bert's questions.
   c. The ability for couple's relationships to thrive due to laughter and humor.
   d. The task to laugh and tease more or less when around each other.

46. What is the response?
   a. The change in the responses to Bert's questions.
   b. The assignment of laughing more or less during the experiment
   c. One couple's responses to Bert's questions.
   d. Whether relationships can thrive with increased disposition for humor and laughter in each partner.

The following situation pertains to questions 47-57.
According to some local high school survey data, the histogram for the weights of the 500 women in the freshmen class is close to the normal curve with an **average of 136 lbs**. and a **SD of 24 lbs**. Answer the questions below using this information! *Round middle areas to the nearest whole number. Hint: Draw the normal curve to help you!*

47. About 68% of the women are between _____ pounds and _____ pounds. (Fill in the blanks with weights)
- a. 1, -1
- b. 134, 137
- c. 112, 160
- d. 88, 184

48. If a student is 0.6 SD's below average, what is their z-score?
- a. -0.6
- b. 0.6
- c. 1
- d. -1
- e. 0.5

49. Approximately what percent of the females in the freshmen class are between 100 and 160 pounds? Answer this in 3 parts. First, translate the interval into z-scores: _____ to _____.
- a. -1.5, 1.5
- b. -1.5, 1
- c. -1, 1
- d. -1, 1.5

50. Next, calculate the percentage between those z-scores.
- a. 87
- b. 68
- c. 9.5
- d. 77.5

51. One female listed her weight as 148 pounds. What is her z-score?
- a. 1
- b. 2
- c. -0.4
- d. 1.5
- e. 0.5

52. What percentile is she in?
- a. 38^{th}
- b. 31^{st}
- c. 69^{th}
- d. 34.5^{th}
- e. 95^{th}

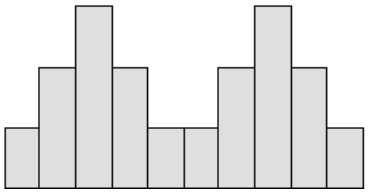
53. One female is in the 92^{nd} percentile. The 92^{nd} percentile corresponds to a middle area of _____%
- a. 84
- b. 92
- c. 8
- d. 16
- e. 50

54. The z-score corresponding to the 92^{nd} percentile is….
- a. 1
- b. 1.4
- c. -1.4
- d. -1

55. Lastly, if a female is in the 92nd percentile, they weigh how much? Pick the closest answer.
- a. 170 lbs
- b. 136 lbs
- c. 178 lbs
- d. 138 lbs

56. If a student is in the 8th percentile, how much does she weigh?
- a. 170 lbs
- b. 102 lbs
- c. 138 lbs
- d. 134 lbs

57. If the histogram for the data looked like this, should we still use the normal approximation?



- a. Yes – we can convert any data to normal data.
- b. No – this histogram is not normal, so it wouldn't make sense to use the normal approximation.
- c. Maybe – it depends if the data is already in z-scores.
- d. Yes – this looks like two normal curves

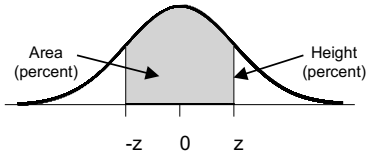
Questions 58-60 pertain to the following situation:
A recent Israeli study looked at currently active serious COVID-19 cases and vaccination status of Israeli residents 12 years and older. In this case, the overall results were misleading, showing that vaccinated patients were more likely to be hospitalized from COVID-19 than unvaccinated patients. This was a classic example of Simpson's Paradox where the overall results described were misleading. The study mentioned that age was a confounder. Answer the questions below about Simpson's Paradox.

58. Simpson's Paradox occurs in…
- a. Randomized controlled experiments
- b. Observational studies
- c. All of the above
- d. None of the above

59. To deal with age being a confounder, the researchers should…
- a. Stratify the results based on age
- b. Block based on age at the start of the study
- c. Remove all old people from the study
- d. There is nothing they can do to control for this confounder

60. When age was controlled for, the researchers found the opposite results: those who were vaccinated were way less likely to be hospitalized than those who were unvaccinated. In observational studies, it is best to…
- a. Always look at overall results.
- b. Always stratify based on confounders if possible and look at those results.
- c. Throw away all of your data.
- d. Do a randomized controlled experiment even if it is unethical.

STANDARD NORMAL TABLE



z	*Area*		*z*	*Area*		*z*	*Area*
0.00	0.00		1.50	86.64		3.00	99.730
0.05	3.99		1.55	87.89		3.05	99.771
0.10	7.97		1.60	89.04		3.10	99.806
0.15	11.92		1.65	90.11		3.15	99.837
0.20	15.85		1.70	91.09		3.20	99.863
0.25	19.74		1.75	91.99		3.25	99.885
0.30	23.58		1.80	92.81		3.30	99.903
0.35	27.37		1.85	93.57		3.35	99.919
0.40	31.08		1.90	94.26		3.40	99.933
0.45	34.73		1.95	94.88		3.45	99.944
0.50	38.29		2.00	95.45		3.50	99.953
0.55	41.77		2.05	95.96		3.55	99.961
0.60	45.15		2.10	96.43		3.60	99.968
0.65	48.43		2.15	96.84		3.65	99.974
0.70	51.61		2.20	97.22		3.70	99.978
0.75	54.67		2.25	97.56		3.75	99.982
0.80	57.63		2.30	97.86		3.80	99.986
0.85	60.47		2.35	98.12		3.85	99.988
0.90	63.19		2.40	98.36		3.90	99.990
0.95	65.79		2.45	98.57		3.95	99.992
1.00	68.27		2.50	98.76		4.00	99.9937
1.05	70.63		2.55	98.92		4.05	99.9949
1.10	72.87		2.60	99.07		4.10	99.9959
1.15	74.99		2.65	99.20		4.15	99.9967
1.20	76.99		2.70	99.31		4.20	99.9973
1.25	78.87		2.75	99.40		4.25	99.9979
1.30	80.64		2.80	99.49		4.30	99.9983
1.35	82.30		2.85	99.56		4.35	99.9986
1.40	83.85		2.90	99.63		4.40	99.9989
1.45	85.29		2.95	99.68		4.45	99.9991