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 (Kelly Findley In Person) **ii)** L2 (Karle Flanagan In Person) **iii)** KF (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 7 pages including the normal table (67 questions).

There is NO CLASS on Thursday or Friday!

Scores will be posted on Compass by Friday evening and exams will be returned in class next week. Online students may pick up their exam in 23 Illini Hall during office hours next week.

The following situation pertains to questions 1-4.
To test the effectiveness of an anti-depressant medication, researchers identified 500 adults with diagnosed "Generalized Anxiety Disorder" to participate in a study. The researchers used coin flips to randomly assign 250 participants to receive a new medication pill while the other 250 participants got a placebo. Neither the participants, nor the researchers knew who was in which group. 3 months later, the researchers had the subjects complete an anxiety questionnaire to assess if there were any differences in scores between groups. The researchers found that those receiving the new medication pill reported significantly lower generalized anxiety symptoms than the placebo group.

1. This study is an example of…
- **a)** An observational study with double-blind controls
- **b)** A non-randomized, controlled double-blind experiment
- **c)** A non-randomized controlled experiment with historical controls
- **d)** A randomized controlled double-blind experiment

2. Does this study provide evidence that the new medication is effective at reducing the symptoms of generalized anxiety disorder?
- **a)** Yes, the study provides strong evidence that the new medication is effective at reducing symptoms.
- **b)** No, it only shows that there is an association between the new medication and reduced symptoms because the patients might have experienced a placebo effect.
- **c)** No, since the individuals did not take the anxiety questionnaire at the beginning of the study for comparison, there is a large concern that the medication group may have had lower anxiety scores from the start.

3. Based on the design, should we expect there to be significant confounders in this study? **a)** Yes **b)** No

4. If the researchers did know which subjects were receiving the medication and which were receiving the placebo, should this affect how we categorize this study? **Choose the best answer.**
- **a)** No, it only matters that the patients were never directly informed which group they were in
- **b)** Yes, this creates an evaluator bias where the researchers might have acted differently with subjects from each group.
- **c)** Yes, this creates a subject bias where the control group subjects might have no longer felt they were receiving treatment.
- **d)** Yes, this introduces Simpson's Paradox where the researchers now need to stratify the participants into adherers vs. non-adherers.

The following situation pertains to questions 5-10.
Researchers at the University of Chicago have been studying possible health effects related to vaping. The researchers completed a large study that included 2,551 who reported vaping at least once a day on average for 3 or more years and 3,612 people who reported having never vaped before. The researchers asked the subjects whether or not they had contracted pneumonia in the past 12 months and reported the following results:

	Size	Percent Reporting Pneumonia
Vapers	2,551	4.93%
Non-vapers	3,612	7.86%

5. This study is an example of…
- **a)** An observational study with controls
- **b)** A non-randomized experiment with blocking
- **c)** A non-randomized experiment sorted by the researchers
- **d)** A randomized controlled experiment

6. Which of the following statements best explain what this study shows?
- **a)** Vaping is a likely cause of increased risk for pneumonia
- **b)** There is an association between those who frequently vape and increased risk for pneumonia, but there isn't evidence here that the two have a causal relationship.
- **c)** Vaping is a likely preventative of pneumonia.
- **d)** There is an association between those who frequently vape and a decreased risk for pneumonia, but there isn't evidence here that two have a causal relationship.

For each factor below, identify whether it is a potential confounder for the relationship between vaping and pneumonia (explains how the two could be coincidentally connected), a causal link between vaping and pneumonia (connects the two directly), or neither:

7. Age – People who vape may be younger on average and less susceptible to contracting pneumonia.
a) Confounder **b)** Causal Link **c)** Neither

8. Living Conditions – People who live in cities might be more prone to catching a respiratory illness and getting pneumonia.
a) Confounder **b)** Causal Link **c)** Neither

9. Inhaling – People who vape regularly may clear out their respiratory passageways, making them less susceptible to respiratory illness like pneumonia.
a) Confounder **b)** Causal Link **c)** Neither

10. Of the following statements, which represents the most concerning weakness of this study?
a) The groups sizes are different, so our comparison percentages may not be trustworthy
b) The researchers didn't include a control group, so we can't make a clear comparison
c) The groups may not be balanced because vapers and non-vapers may be different in multiple ways
d) The evaluators were not blind to who was in which group, and their actions may have affected subjects' likelihood of getting pneumonia.

The following situation pertains to questions 11-14.
A study was conducted to test the effectiveness of a new lotion for treating acid burns. The researchers want to split 42 subjects into two groups. To sort the groups, they have medical professionals rate the severity of each patient's burns as mild, moderate, or severe—then the researchers **randomly** sorted patients of each burn category to each group and also ensured an equal number of each category per group. One group receives the new lotion and one group receives a standard, over-the-counter lotion—subjects don't know which type of lotion they are receiving, but the doctors treating them do. The researchers then record how many days it takes before the doctor decides each burn is sufficiently healed.

11. Is this an observational study or designed experiment? **a)** An observational study **b)** Designed experiment

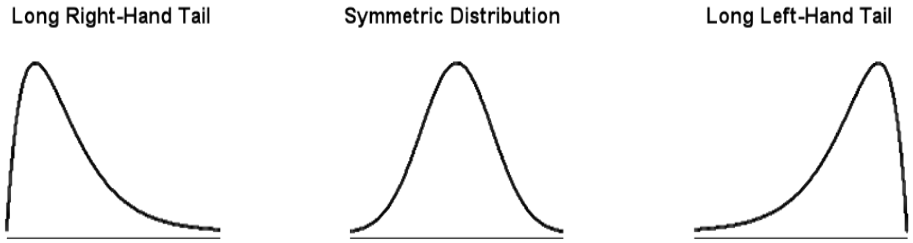
12. How were the groups sorted?
a) Random Assignment without blocking
b) Random Assignment with blocking
c) Non-Random Assignment
d) There was no control group

13. Was there a placebo? **a)** yes **b)** no

14. Which of the following biases are present in this design?
a) Evaluator Bias – The professionals directly interacting with the subjects know who is getting the treatment
b) Subject Bias – Subjects know if they are really getting the treatment
c) Group Selection Bias – The groups were selected with bias and may not be equivalent
d) No Clear Bias – This design represents the gold standard and has no clear bias

15. For which of the following histograms is it reasonable to use the normal approximation?

a) Long Right-Hand Tail Only
b) Long Left-Hand Tail Only
c) Symmetric Only
d) All of them



The following situation pertains to questions 16-19.
For the following data sets described in questions 16-19 below, decide whether you think the histogram would have a long left-hand tail, long right-hand tail, or be fairly symmetrical.

16. ***Dataset:*** Exam scores where the median is 90 but the average is only 80.

a) long right-hand tail **b)** fairly symmetrical **c)** long left-hand tail

17. ***Dataset:*** Heights of all female U of I students.

a) long right-hand tail **b)** fairly symmetrical **c)** long left-hand tail

18. ***Dataset:*** Household income in the US.

a) long right-hand tail **b)** fairly symmetrical **c)** long left-hand tail

19. ***Dataset:*** Exam scores where the median and average are about the same.

a) long right-hand tail **b)** fairly symmetrical **c)** long left-hand tail

The following situation pertains to questions 20-23.
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 of the two treatments.

	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%

20. Which treatment has a higher success rate for deep wrinkles?
a) WTF **b)** LOL **c)** cannot be determined from the information given

21. Which group has a higher success rate for fine lines?
a) WTF **b)** LOL **c)** cannot be determined from the information given

22. Which treatment has the higher overall success rate (combining those who have deep wrinkles and fine lines)?
a) WTF **b)** LOL **c)** cannot be determined from the information given

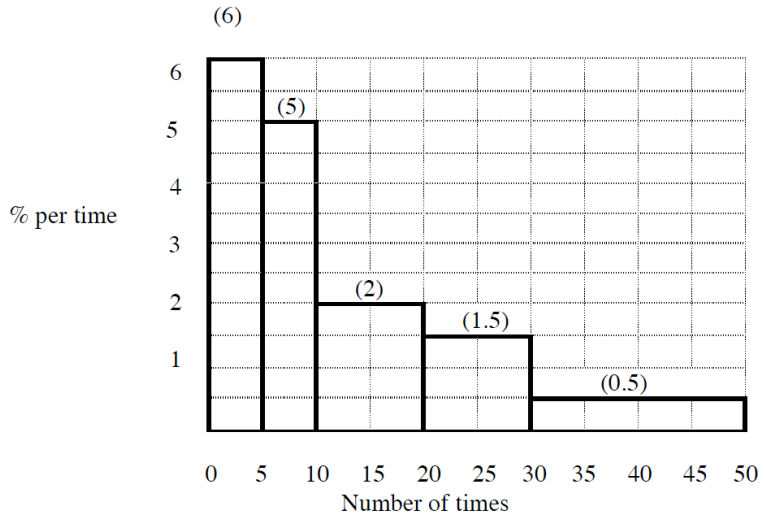
23. What conclusion is best supported by these results?

a) 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.

b) 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.

c) The success rates of the two treatments depend on the doctor who is giving them. We have to consider all of the factors, including many possible confounders that are not shown in the table.

The following situation pertains to questions 24-35.
The figure below is a histogram for the number of times students use their cell phones to text in Stat 100 class last semester. The height of each block is given in parentheses. (Assume an equal distribution throughout each interval.)



25% of the students texted 5-10 times. What percentage of the students texted…

24. 0-5 times _____% **a)** 6% **b)** 25% **c)** 30% **d)** 40% **e)** 50%

25. 10-20 times _____% **a)** 2% **b)** 10% **c)** 20% **d)** 30% **e)** 40%

26. 20-30 times_____% **a)** 1.5% **b)** 10% **c)** 15% **d)** 20% **e)** 50%

27. 30-50 times_____% **a)** 0.5% **b)** 5% **c)** 10% **d)** 15% **e)** 20%

28. The median number of texts is closest to: **a)** 5 **b)** 6 **c)** 9 **d)** 15 **e)** 20

29. The average is _____the median. **a)** less than **b)** greater than **c)** equal to **d)** cannot be determined

30. Did more people text 0-10 times or 20-50 times, or are they the same? **a)** More 0-10 **b)** More 20-50 **c)** Same

31. The 90th percentile is **a)** 5 **b)** 20 **c)** 30 **d)** 40 **e)** 45

32. Assuming an even distribution throughout each interval, what percent of students texted 40-50 times?
a) 0.5% **b)** 5% **c)** 10% **d)** 15% **e)** 20%

33. If everyone had texted 10 more times this semester, would the average, median, and SD all increase by 10?
- **a)** No, only the average would increase by 10.
- **b)** No, only the median would increase by 10.
- **c)** No, only the SD would increase by 10.
- **d)** No, the average and median would increase by 10, but the SD would stay the same.
- **e)** Yes, the average, median, and SD would all increase by 10.

34. If everyone had texted double this semester (their texts were multiplied by 2), would the average, median, and SD all be multiplied by 2?
- **a)** No, only the average would be multiplied by 2.
- **b)** No, only the median would be multiplied by 2.
- **c)** No, only the SD would be multiplied by 2.
- **d)** No, the average and median would be multiplied by 2, but the SD would stay the same.
- **e)** Yes, the average, median, and SD would all be multiplied by 2.

35. Would it be appropriate to use the normal approximation for this data? **a)** Yes **b)** No **c)** It depends

Questions 36-40 pertain to the following list of 5 numbers: -2, 1, 0, 5, 6

36. The average is _____. **a)** 0 **b)** 1 **c)** 2 **d)** 10

37. The median is _____. **a)** 0 **b)** 1 **c)** 2 **d)** 10

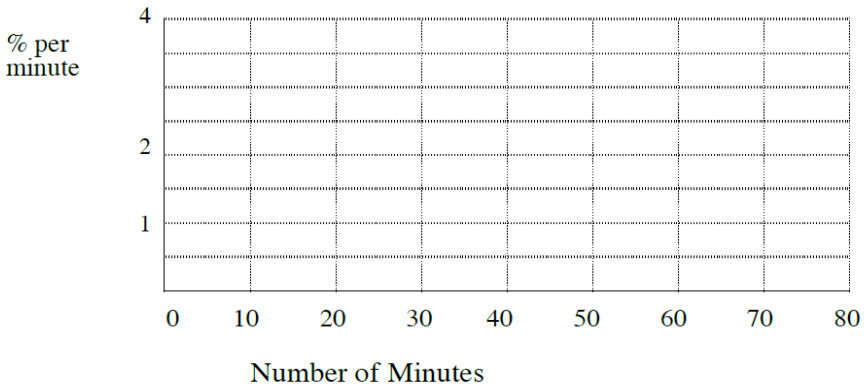
38. The deviations from the average are: **a)** -4, -1, -2, 3, 4 **b)** 4, 1, 2, -3, -4 **c)** -3, 0, -1, 4, 5 **d)** 3, 0, 1, -4, -5

39. The sum of the deviations from the average **should** = _____. **a)** 0 **b)** 1 **c)** 10 **d)** unable to determine

40. Compute the Standard Deviation and round your answer to 2 decimal places. ***What is the SD?***
a) 3.03 **b)** 3.19 **c)** 9.2 **d)** 10.2

The following situation pertains to questions 41-48.
A distribution table for the number of minutes parents spent reading to their pre-school kids each day is shown below. Fill in the column for height and then draw the histogram below.

Minutes	%	Height (% per minute)
0-10	15	
10-20		
20-40	30	
40-80	20	



41. What percent of parents spent 10-20 minutes reading to their kids? **a)** 15 **b)** 35 **c)** 50 **d)** 65

42. What's the height of the 0-10 minute block? **a)** 0.5 **b)** 1.5 **c)** 1.75 **d)** 3.5

43. What's the height of the 10-20 minute block? **a)** 0.35 **b)** 1.5 **c)** 1.75 **d)** 3.5

44. What's the height of the 20-40 minute block? **a)** 0.5 **b)** 0.75 **c)** 1.5 **d)** 2

45. What's the height of the 40-80 minute block? **a)** 0.25 **b)** 0.5 **c)** 1.5 **d)** 2

46. What's the median of the number of minutes parents spent reading to their pre-school kids?
a) 10 **b)** 20 **c)** 35 **d)** 50

47. Assuming an equal distribution throughout the 40-80 interval, the percent of parents who reported spending exactly 42 minutes reading to their child is closest to: **a)** 0.5% **b)** 1.5% **c)** 20% **d)** 40%

48. What value corresponds to the 80th percentile? **a)** 10 **b)** 20 **c)** 40 **d)** 80

The following normal curve situation pertains to questions 49-51. The data from the following question on our Stat 100 survey follows the normal curve: "How old is your mother?" The average is 48 and the SD is 3.

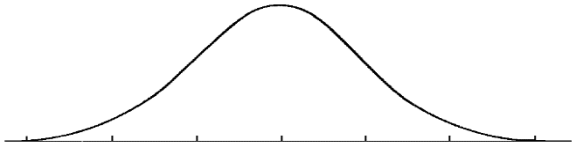
49. Say I wanted to find out what age corresponds to the 26th percentile. I'd have to use our normal table to look up a middle area to get the z-score for the 26th percentile and then convert it to an age. What middle area would I use?
a) 24 **b)** 26 **c)** 48 **d)** 52

50. Would the z-score for the 26th percentile be positive or negative? **a)** positive **b)** negative. **c)** you could use either

51. If a student's mother is 1 standard deviation below average in age, what is their z-score?
a) 1 **b)** -1 **c)** 2 **d)** 50 **e)** impossible to determine

The following situation pertains to questions 52-58. This week, we asked a group of random students who have donated to political campaigns in 2019 how much they donated. Of the people we asked, we found that the amount of money they donated followed the normal curve with an **average of $10 and an SD of $2**.

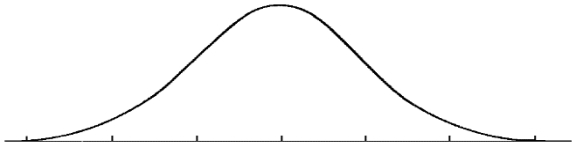
52. What percentage of people donated *more than $14* to political campaigns? First, what is **$14** as a **Z**-score***?***
a) 1.2 **b)** -2 **c)** 2 **d)** – 1.2 **e)** 1.5



53. What percent of people donated over $14?
a) 2.5% **b)** 5% **c)** 50% **d)** 95% **e)** 97.5%

Next, let's find out what percent of students donated *between $7 and $14* to political campaigns?

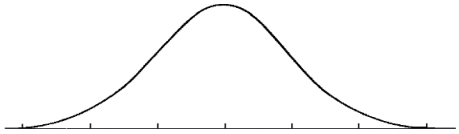
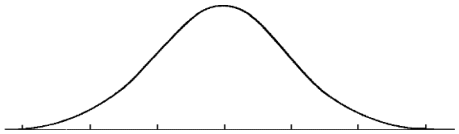
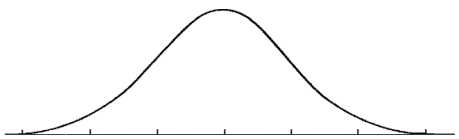
54. Convert **$7** to a **Z**-score. (You've already converted $14 to a Z score above).
a) -0.5 **b)** 0.5 **c)** -1.5 **d)** 1.5 **e)** 2



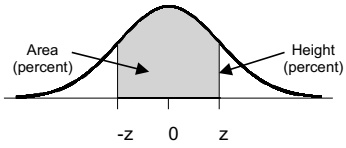
55. Use the normal curve to calculate the percentage of students who donated between $7 and $14?
a) 4% **b)** 8% **c)** 87% **d)** 91% **e)** 95%

56. If you're ***above*** average in donating to political campaigns, your z-score is? **a)** positive. **b)** negative **c)** unknown
57. If you're exactly at the 50th percentile in donating then your Z score is… **a)** -1.5 **b)** 0 **c)** 2 **d)** 50
58. And the 50th percentile corresponds to what contribution amount? **a)** $2 **b)** $7 **c)** $10 **d)** $14

The following situation pertains to questions 59-67. IQ scores among US adults follow the normal curve quite closely with an **average = 100 and SD= 15.** Fill in the missing info.

IQ Score	Z-Score	Percentile
Douglas had an IQ score of **106.**	**59. Z** = _____ a) 0.4 b) -0.4 c) -0.8 d) 0.9	**60.** Douglas is in the _____ percentile a) 18.5 b) 31 c) 34.5 d) 65.5 e) 81.6 
61. Chauncey's IQ score was _____ **a)** 80.5 **b)** 100 **c)** 113.7 **d)** 119.5 **e)** 145	**Z** = -1.3	**62.** Chauncey is in the _____ percentile  **a)** 9.68 **b)** 19.36 **c)** 50 **d)** 80.64 **e)** 90.32
63. Jackie's IQ score was _____ **a)** 79 **b)** 92 **c)** 108 **d)** 155 **e)** 121	**64. Z** = _____ **a)** -1.4 **b)** 1.75 **c)** 0.55 **d)** 1.4	**65.** Jackie is in the **92nd** percentile. What middle area should you look up on the normal table to find the correct Z score?  **a)** 4 **b)** 8 **c)** 42 **d)** 84 **e)** 92
66. Richard's IQ score was **a)** 8 **b)** 79 **c)** 108 **d)** 121 **e)** 155	**67.** Z = _____ **a)** -1.4 **b)** 1.75 **c)** 0.55 **d)** 1.4	Richard is in the **8th** percentile.

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