Exam 1 STAT 100 Form B Answer Key Spring 2022

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) L2 (Karle Flanagan In Person) **ii)** O1 (Karle Flanagan Online) **iii)** O2 (Jonas Reger Online)

This test is ALL multiple choice. **Circle all answers on this exam and fill in the corresponding bubble on your orange 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 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 UIN number in the Student Number box.
- Print and bubble in your NET ID with **no spaces** in the NETWORK ID box.
- *Write Stat 100 on the COURSE line.*
- *Write your instructor's name (Karle Flanagan or Jonas Reger) on the INSTRUCTOR line.*
- *Write your section (L2, O1, or O2) on the SECTION line.*
- *Sign your name, and right underneath the student signature line PRINT your name.*

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 9 pages including the normal table (57 questions).

There is NO CLASS on Thursday this week!

Scores will be posted on Canvas by Monday at noon. Students may pick up their exam in 0060 Siebel Center for Design during office hours next week.

The next questions pertain to the following situation: Below is the list of score differences for the last 6 Illinois Football games (i.e., Illinois Score - Opponent Score): 8, 0, -3, -4, 6, 2 -4, -3, 0, 2, 6, 8

1. What is the average?
   a) 1 b) 9 c) 12 d) 4.5 e) 1.5

2. What is the median?
   a) -3.5 b) 1 c) 1.5 d) -3 e) -4

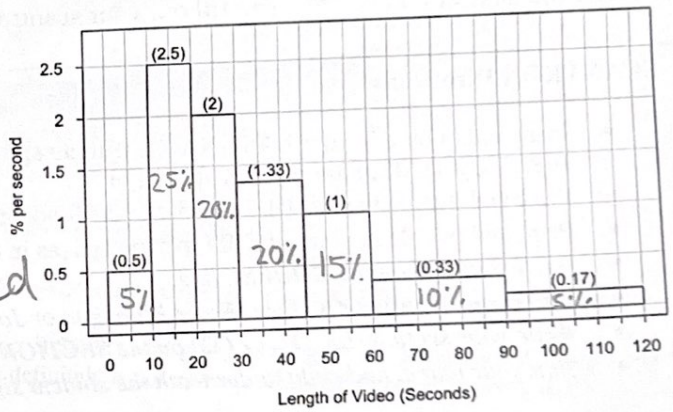
3. What is the standard deviation?
   a) 0 b) 1 c) 1.5 d) 4.39 e) 19.25 -5.5, -4.5, -1.5, 0.5, 4.5, 6.

4. If the score difference of the first game was 5 instead of 8, how would the **median** change?
   a) It would increase b) It would decrease c) It would stay the same

5. If the score difference of the first game was 5 instead of 8, how would the **mean** change?
   a) It would increase b) It would decrease c) It would stay the same

6. If the score difference of the first game was 5 instead of 8, how would the **standard deviation (SD)** change?
   a) It would increase b) It would decrease c) It would stay the same

Questions 7-14 pertain to the following situation:
Karle likes to watch TikTok videos during her breaks. Here is a histogram that shows the length of each TikTok video she has watched. The height of each block is given in parentheses. *Assume an even distribution throughout each interval.*



7. Indicate whether the following statement is true or false: The median is smaller than the average. avg > med
   a) True
   b) False
   c) Not enough information given

8. What percent of videos are between 20 and 30 seconds?
   a) 20 b) 10 c) 2 d) 5 e) 12

9. If Karle loses interest in a video after 45 seconds, what percentage of videos does Karle lose interest in?
   a) 15 b) 30 c) 70 d) 1.5 e) 50

10. The median is closest to:
    a) 40 b) 60 c) 20 d) 30 e) 2

11. The percent of videos that were exactly 80 seconds in length is closest to:
    a) 0.33 b) 30 c) 80 d) 10 e) 1

12. If each video was 10% longer (same as multiplying by 1.1), then…
    a) The average would increase and the SD would decrease
    b) The average would decrease and the SD would decrease
    c) The average would increase and the SD would increase
    d) The average would decrease and the SD would increase
    e) The average would increase and the SD would stay the same

13. If Karle only watched half of every video (i.e., each video is now half as long)…
    a) The average would increase and the SD would decrease.
    b) The average would decrease and the SD would decrease
    c) The average would increase and the SD would increase
    d) The average would decrease and the SD would increase
    e) The average would increase and the SD would stay the same

14. If all of the videos in the 90-120 block doubled in length, but the videos in the 0-90 blocks stayed the same, would the median change?
    a) No b) Yes c) Impossible to tell

15. If you knew the average and SD of the videos displayed in the histogram above, would it be appropriate to use the normal approximation to figure out what percentage of the video lengths fell within various intervals?
    a) Yes, the histogram of the video lengths follows the normal curve
    b) No, the histogram of the video lengths is not close enough to following the normal curve; it has a long right-hand tail
    c) No, the histogram of the video lengths is not close enough to following the normal curve; it has a long left-hand tail
    d) None of the above

Questions 16-23 pertain to the following situation: Karle and her husband Steve recently started going to a gym called "The Gym". Steve was skeptical at first, but ultimately decided to go after Karle convinced him! While looking up statistical studies on going to the gym, Karle came across one that said that going to the gym reduces your risk of cancer. The study looked at thousands of middle-aged adults and saw that those who frequently exercised (at least 30 minute 3 times per week) were significantly less likely to develop cancer in the next 10 years.

16. Which of the following statements best describes this study?
    a) It's an observational study.
    b) It's a randomized controlled experiment without a placebo and without "blind" evaluators.
    c) It's a randomized controlled double-blind experiment.
    d) It's a randomized controlled experiment with a placebo.

17. Based on the results of this study, what can we conclude about the relationship between going to the gym and getting cancer?
    a) There is a clear causal relationship. If Steve keeps going to the gym, he definitely will not get cancer.
    b) There is an association, but there definitely is not a causal relationship between going to the gym and getting cancer.
    c) We see an association, but we cannot say for sure that going to the gym means you won't get cancer.
    d) It's impossible to say. This study shows that people who go to gym are more likely to get cancer, but we don't know who this applies to.

18. In this study, the **treatment** is….
    a) the middle-aged adults b) going to the gym c) whether or not someone gets cancer

19. In this study, the **response** is….
    a) the middle-aged adults b) going to the gym c) whether or not someone gets cancer

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

20. Type of Gym- At certain types of gyms, like Orangetheory Fitness, people are more likely to go because they charge you a fee if you miss a class you signed up for.
    a) Causal Link b) Confounder c) Neither

21. Wealth- Gyms are very expensive! People who are wealthy are more likely to be able to afford a gym membership than people who are not. Also, people who are wealthy are more likely to not get cancer for many reasons.

a) Causal Link **b) Confounder** c) Neither

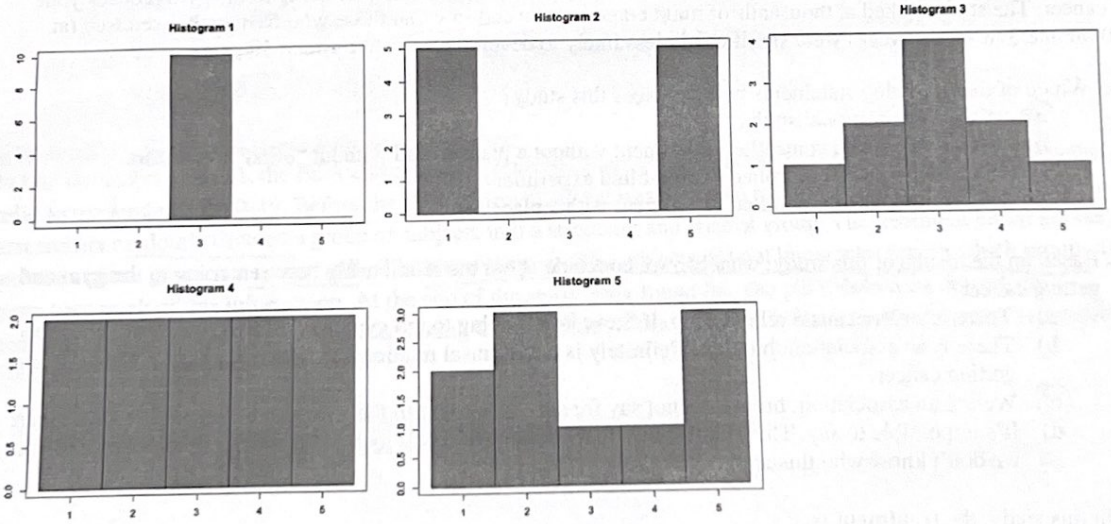
22. Exercise- When people go to the gym, they go to exercise. Exercising has many health benefits that can reduce your risk of cancer.

a) Causal Link b) Confounder c) Neither

23. Eating Healthy- Eating healthy gives people energy! People who eat healthy are more likely to go to the gym and are also more likely to not get cancer than people who eat fast food all the time.

a) Causal Link **b) Confounder** c) Neither

The next 4 questions pertain to the following 5 histograms:



24. Which histogram has SD=0?

a) Histogram 2 b) Histogram 5 c) Histogram 3 **d) Histogram 1** e) Histogram 4

25. Which Histogram has an average of 3?

a) Histogram 4 b) Histogram 1 c) Histogram 2 d) Histogram 3 **e) Histograms 1, 2, 3, and 4**

26. What histogram is has a long left-hand tail?

a) Histogram 3 b) Histogram 2 c) Histogram 5 d) Histogram 4 **e) None of them**

27. Which histogram has the largest SD?

a) Histogram 5 b) Histogram 3 c) Histogram 1 **d) Histogram 2** e) Histogram 4

Questions 28-32 pertain to the following situation:
A study compared the success rate of two treatments designed to help smokers quit smoking. Subjects were classified as either heavy smokers or moderate smokers before treatment began. Heavy smokers have a harder time quitting. The table below gives the results for the two treatments.

	Treatment A				Treatment B			
	Number	# Successes	# Failures	% Success	Number	#Successes	# Failures	% Success
Heavy Smokers	400	200	200	50%	100	30	70	30%
Moderate Smokers	600	600	0	100%	900	870	30	97%
Overall Total	1000	800	200	80%	1000	900	100	90%

28. Based only on the information given above, do you think this was a randomized experiment?
a) No, since it's highly unlikely that randomization would result in 40% heavy smokers in A and only 10% in B.
b) Yes, since it's highly unlikely both groups would end up with exactly 1000 subjects otherwise.
c) There are arguments for both sides, it's hard to tell without more information.

29. Which treatment had a higher success rate for heavy smokers?
a) A **b)** B **c)** Not enough info

30. Which treatment had a higher success rate for moderate smokers?
a) A **b)** B **c)** Not enough info

31. Based only on the information given, which treatment would you say is better?
a) A **b)** B **c)** It depends on whether they are heavy or moderate smokers

32. This is a classic case of....
a) Blocking **b)** Simpson's Paradox **c)** The Placebo Effect **d)** The Normal Curve

Questions 33-37 pertain to the following situation:
In late December of 2021, the FDA authorized *Molnupiravir* as an anti-viral pill for use in individuals 18 or older with mild to moderate COVID-19. Before the FDA approval, a trial was done. To test the effectiveness of this pill, the researchers randomly divided a group of subjects into a treatment and control group. The treatment group got the pill and the control group got a sugar pill. The researchers and the subjects did not know who was in which group- a third party kept track of this information. At the end of the study, they found that the pill (Molnupiravir) reduces hospitalizations by 30%. After 5 days of taking the pill, none of the individuals in the treatment group had viable COVID-19 virus detected.

33. How would you classify the study that was done?
a) A non-randomized controlled experiment
b) A randomized experiment with historical controls
c) A randomized controlled double-blind experiment
d) An observational study

34. Should the researchers be very worried about confounders in this study?
a) Yes- we should always be worried about confounders because they occur in most studies.
b) Yes- the treatment and control group were very different, which leads to confounding.
c) Yes- both a and b are correct.
d) Maybe- it depends on if the study shows that the pill worked.
e) No- randomized controlled experiments do not have confounders.

35. Can we say with confidence that the pill (Molnupiravir) is an effective treatment for reducing hospitalizations from COVID-19?
a) Yes **b)** No

36. In what case would it be useful to use blocking at the start of the study mentioned above?

a) If the sample size is large and you think a variable would not affect the response, you should block based on that variable.
b) If the sample size is small and you think a variable could affect the response, you should block based on that variable.
c) Blocking isn't useful in this situation, however, stratification would be useful for the study.
d) There would never be a case where blocking would be useful.

37. Let's say that during this study, not everyone took their medicine! If we had people who did not adhere to both the treatment and the sugar pill, what comparison should we make to determine whether or not the pill is effective? Choose the comparison that is **best**.

a) We should stratify and compare adherers directly to non-adherers.
b) We should compare just the people who took the pill in the treatment group to everyone in the control group.
c) We should compare adherers in treatment group (people who took the pill) to adherers in the control group (people who took the sugar pill).
d) We should compare everyone in treatment to everyone in control to keep the original randomization.

Normal Curve Questions- please round the percentages on the normal table to the nearest whole number when answering the following questions.

Questions 38-40 pertain to the following scenario: The scores on an Organic Chemistry exam follow the normal curve with an average of 55 and a SD of 15. Let's say there's a nice professor who wants to give half of the class A's and the other half of the class B's.

38. What z-score should be the cutoff between an A and a B? _____

a) 0 **b)** 0.65 **c)** 1 **d)** 0.5 **e)** 2

39. What exam score should be the cutoff between an A and a B? _____

a) 70 **b)** 89 **c)** 62.5 **d)** 55 **e)** cannot be determined

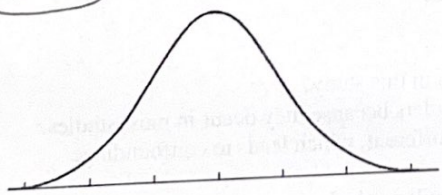
Let's say the nice professor says they want to give every student who is more than 1.5 standard deviations above the average extra credit!

40. What exam score is 1.5 standard deviations above the average? _____

a) 72.5 **b)** 77.5 **c)** 87.5 **d)** 92.5 **e)** 57.5

41. What percent of students will get extra credit? Hint: Use the normal curve!

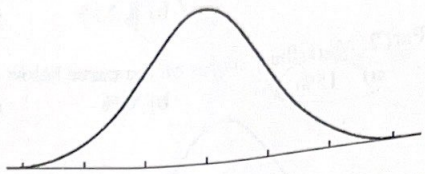
a) 6.5% **b)** 13% **c)** 1.24% **d)** 87% **e)** 0.62%

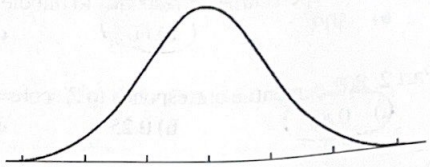


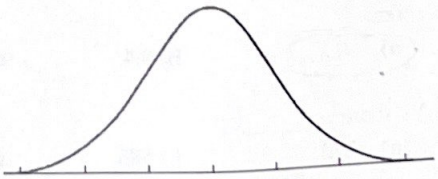
42. If the exam scores follow the normal curve, how does the average compare to the median?

a) Average > Median
b) Average = Median
c) Average < Median
d) We are unable to determine this because there is no histogram given

Questions 43-49 pertain to the following situation: According to survey data, Stat 100 students have an average height of 65" and a SD of 3" and the histogram of their heights is close to the normal curve. Consider the following students: Glo, Flo, and Ro. For each student you're given either their height, z-score, or percentile. Fill in the blanks below to answer the following questions.

Height in Inches	Z Score	Percentile
Glo is 71 inches tall	**43.** Glo's Z-Score = _____ **a)** 1 **b)** 2 **c)** -2 **d)** 3	**44.** Glo's Percentile = _____ *Hint: draw a normal curve **a)** 5 **b)** 95 **c)** 2.5 **d)** 97.5 

Height in Inches	Z Score	Percentile
45. Flo's Height = _____ **a)** 64 **b)** 59 **c)** 62 **d)** 71 **e)** 68	Flo's Z-Score is -1	**46.** Flo's Percentile=_____ *Hint: draw a normal curve **a)** 32 **b)** 16 **c)** 84 **d)** 68 

Height in Inches	Z Score	Percentile
47. Ro's Height = _____ **a)** 58.25 **b)** 65.75 **c)** 66.5 **d)** 67.55	**48.** Ro's Z-Score = _____ **a)** 0.15 **b)** 0.25 **c)** 0.5 **d)** 0.85	Ro is in the 60th Percentile **49.** What middle area corresponds to the 60th percentile? 20 *Hint: draw a normal curve **a)** 10 **b)** 20 **c)** 40 **d)** 60 

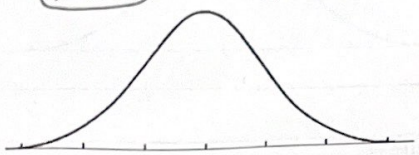
Questions 50-57: The following questions pertain to Math SAT scores that are normally distributed with an average=500 and SD=100.

50. Approximately 68% of Math SAT scores are between _____ and _____
(Fill in the blanks with SAT scores, NOT z scores)
a) 300, 700 **b)** 400, 600 **c)** 432, 568 **d)** 340, 660 **e)** Cannot be determined

What percent of the SAT scores fall in the interval 600 to 735?

51. Part 1: Translate interval into Z scores: _____ to _____
a) -1, 1 **b)** 1, 2.35 **c)** -2.35, 2.35 **d)** -1, -2.35 **e)** -1, 2.35

52. Part 2: Mark the z scores on the curve below, shade the interval and calculate the percent. _____%
a) 15% **b)** 30% **c)** 68% **d)** 83% **e)** 98%



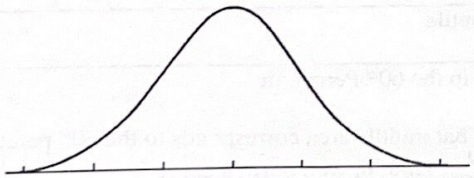
$\frac{98-68}{2} = 15\%$

What score corresponds to the 80th percentile? Mark the 80th percentile on the curve below, find the corresponding z-score, and the corresponding SAT score.

53. Part 1: 80th percentile corresponds to middle area=_____%
a) 30% **b)** 60% **c)** 80% **d)** 20%

54. Part 2: 80th percentile corresponds to Z score= _____
a) 0.85 **b)** 0.25 **c)** 1.3 **d)** 0.4

55. Part 3: 80th percentile corresponds to SAT score (value) =_____
a) 630 **b)** 540 **c)** 525 **d)** 585



What SAT score corresponds to the 20th percentile?

56. Z score= _____
a) -0.85 **b)** 0.4 **c)** -0.25 **d)** 0.85 **e)** -0.4

57. SAT score =_____
a) 370 **b)** 585 **c)** 460 **d)** 325 **e)** 415