EXAM 2: Statistics 100

Read all of the directions below and make sure you fill out your scantron correctly!

Cover Sheet Questions

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

(Last name) (First name)

2) What's your **net ID** (email)? _____

3) Which **section** are you in? Circle one below.

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 and bubble in your FIRST INITIAL in the right-most column.
- Print and bubble in your University Identification Number (UIN) in the Student Number box.
- Print and bubble in your NET ID with **no spaces** in the NETWORK ID box (ex. kflan).
  - Be sure to include the numbers. Do not bubble in any dashes.
- *Write Stat 100* on the COURSE line.
- *Write your instructor's name* 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

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 8 pages including the normal table (68 questions).

There is NO CLASS on Thursday!

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 1-5 pertain to the following scenario: Karle often walks from *Siebel Center for Design* (SCD) to *Siebel Center for Computer Science* (SCS) (i.e., the Siebel-to-Siebel Trek). Consider the travel times of her recent Siebel-to-Siebel trips (given in minutes):

Walk Times: 16, 20, 36, 18, 24, 28, 32, 23 16, 18, 20, 23 | 24, 28, 32, 36

med

1. What is the median of Karle's walk times? a) 10.5 b) 17 c) 21 (d) 23.5

2. What is the 1st Quartile (Q1) of Karle's walk times? a) 8.5 b) 11 (c) 19 d) 28

3. What is the 3rd Quartile (Q3) of Karle's walk times? a) 13.5 b) 22 c) 28 (d) 30

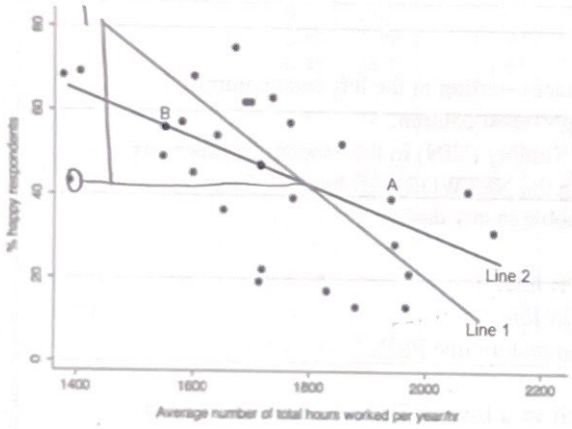
4. What is the IQR of Karle's walk times? Q3-Q1 = 30-19 a) 5 b) 6 c) 10 (d) 11

5. Are there any outliers in Karle's walk times?
   a) Yes, low outlier b) Yes, high outlier c) Yes, high and low outlier (d) No e) Not enough information

low outliers $< 19 - 1.5(11) = 2.5$

high outliers $> 30 + 1.5(11) = 46.5$

The next set of questions pertain to the following situation: A recent Harvard study analyzed the relationship between the number of hours worked per year and reported happiness by nation. The results are displayed below for 34 nations.



6. The SD line and regression line are shown. Which line is the regression line?
   a) Line 1 is the regression line- it is steeper than Line 2.
   (b) Line 2 is the regression line- it is flatter than Line 1.
   c) Cannot be determined.

7. What is the correlation coefficient?
   (a) -0.65 b) 0.95 c) 0 d) 0.65 e) -0.95

8. The average of all the residuals is…
   a) -1 (b) 0 c) 1 d) not enough info is given to know

9. What's the general trend observed from this data?
   a) At a national level, as work hours increase, happiness tends to increase.
   (b) At a national level, as work hours increase, happiness tends to decline.
   c) There is no correlation between the number of hours worked and level of happiness.
   d) In all nations, an individual who works 2000 hours/year will be happier than an individual who works 1800 hours/year.

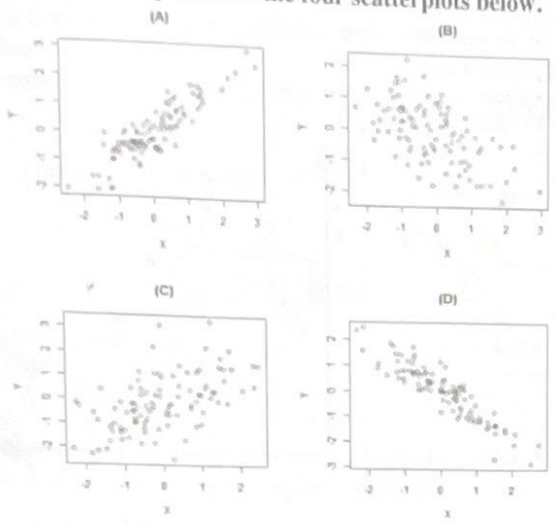
10. What is Nation A's residual? a) 0 (b) 5 c) 15 d) -5 e) -15

11. What is Nation B's residual? (a) 0 b) 5 c) 15 d) -5 e) -15

12. Nation B is _____ predicted. a) Less happy than (b) Exactly as happy as c) More happy than

13. Suppose we made a new scatter plot with every individual plotted instead of their respective nation's average. How would the correlation tend to change?
   a) The two scatter plots would be based off of the same data, so the correlation would be the same.
   b) More data will give us a stronger correlation.
   (c) The extra scatter of the added dots will tend to weaken the correlation. This is an example of ecological correlations.

Questions 14-17 pertain to the four scatterplots below.



Match the 4 plots to their correlation coefficient.

14. r = -0.52
a) Plot A **b) Plot B** c) Plot C d) Plot D

15. r = -0.91
a) Plot A b) Plot B c) Plot C **d) Plot D**

16. r = 0.56
a) Plot A b) Plot B **c) Plot C** d) Plot D

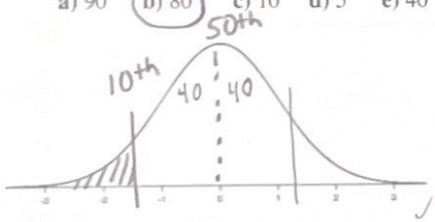
17. r = 0.89
a) Plot A b) Plot B c) Plot C d) Plot D

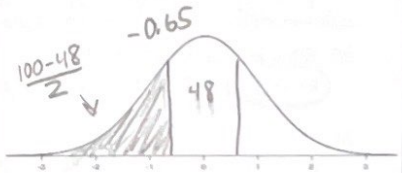
Questions 18-24 are based on this scenario: Suppose listening skills and speaking skills follow the normal curve but have different correlations among different populations. Consider 5 populations where the correlation coefficients between listening and speaking skills are as given in the table below. If someone is in the **90th** percentile in listening skills, estimate his percentile in speaking skills in each population.

Listening Skills Percentile	r	Speaking Skills Percentile
90th	***0***	**18.** a) 10th b) 26th **c) 50th** d) 74th e) 90th
90th	***-1***	**19.** **a) 10th** b) 26th c) 50th d) 74th e) 90th
90th	***0.5***	**20.** a) 10th b) 26th c) 50th **d) 74th** e) 90th

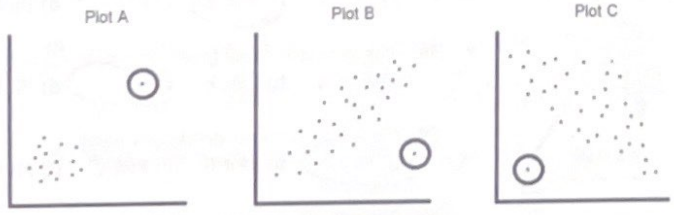
If a student is in the **10th** percentile in listening skills where r = 0.5, estimate her speaking percentile ***by filling in the table below***

Listening Skills Percentile	Listening Z	r	Speaking Z	Speaking Skills Percentile
Student is in the 10th percentile for listening skills. **21.** What middle area on the table should you look up to find the Z score? a) 90 **b) 80** c) 10 d) 5 e) 40	**22.** Z = -1.3 a) 1.3 b) -0.5 c) -0.15 d) 0.5 **e) -1.3**	r=0.5	**23.** Z = -0.65 **a) -0.65** b) -2.6 c) -0.25 d) 0.5 e) -0.075	**24.** Speaking Skills Percentile = 26th a) 40 b) 74 **c) 26** d) 48 e) 52





The next 3 questions pertain to the scatter plots below! Each of these scatter plots has an outlier (circled). Does **including** the outlier make r stronger or weaker for each plot? For *each of the scatter plots*, circle **Stronger** *or* **Weaker** to indicate the outlier's effect on r.



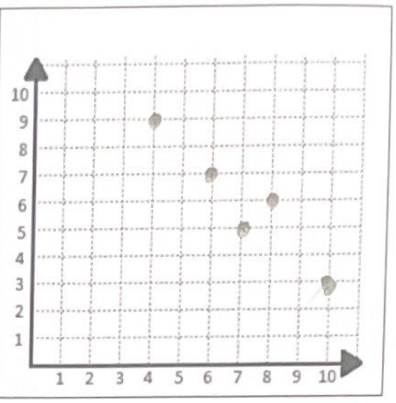
25. Including the outlier in Plot A makes r:
a) stronger b) weaker

26. Including the outlier in Plot B makes r:
a) stronger **b) weaker**

27. Including the outlier in Plot C makes r:
a) stronger **b) weaker**

Questions 28-36 pertain to this set of X and Y points.
Part 1: Fill in the table and plot the 5 points. The **average** of **X=7** and the average of **Y=6**. The **SD** of **X** and **Y** are **both 2**. (NOTE: X and Y have DIFFERENT averages)

X	Y	Z-score for X	Z-score for Y	Products
4	9	Blank 1 $z=\frac{4-7}{2}=-1.5$	$z=\frac{9-6}{2}=1.5$	-2.25
6	7	-0.5	Blank 2 0.5	-0.25
7	5	0	-0.5	0
8	6	0.5	0	Blank 3 0
10	3	1.5	-1.5	-2.25
Totals		**Total *should* =** 0	**Total *should* =** 0	**Total =** -4.75



28. What number goes in Blank 1? **a)** -1.5 **b)** -1 **c)** -0.5 **d)** 0 **e)** 0.5

29. What number goes in Blank 2? **a)** -1.5 **b)** -1 **c)** -0.5 **d)** 0 **e)** 0.5

30. What number goes in Blank 3? **a)** -1.5 **b)** -1 **c)** -0.5 **d)** 0 **e)** 0.5

31. The totals of the z-score for X and the z-score for Y columns should equal: **a)** 0 **b)** 1 **c)** 100 **d)** a positive number

32. To find the correlation coefficient, you should:
a) sum the products **b)** square the products and take the average **c)** sum the z-scores **d)** take the average of the products

33. The scatterplot you drew above should:
a) slope down and to the right **b)** slope up and to the right **c)** form no linear pattern

Part 2: The correlation of the Original List in Part 1 is -0.95. What would be the new r if the lists were changed as shown below. **HINT:** Compare the X Y data sets below to the original X Y data set in Part A (also listed below) and think about changes affect r?

Original List

X	Y
4	9
6	7
7	5
8	6
10	3

New List 1 (×2)

X	Y
8	9
12	7
14	5
16	6
20	3

New List 2 (switch)

X	Y
9	4
7	6
5	7
6	8
3	10

New List 3 (×-10, switch signs)

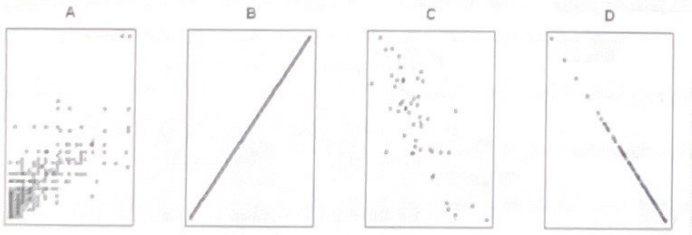
X	Y
4	-90
6	-70
7	-50
8	-60
10	-30

34. What is r for New List 1?
a) -0.95 **b)** 0.95 **c)** -1 **d)** 1

35. What is r for New List 2?
a) -0.95 **b)** 0.95 **c)** -1 **d)** 1

36. What is r for New List 3?
a) -0.95 **b)** 0.95 **c)** -1 **d)** 1

Questions 37-40: Match the 4 plots to their descriptions.



37. Temperature in Fahrenheit versus Temperature in Celsius
a) Plot A **b)** Plot B **c)** Plot C **d)** Plot D

38. Missed classes and grade in a math class
a) Plot A **b)** Plot B **c)** Plot C **d)** Plot D

39. Party hours versus drinks per week
a) Plot A **b)** Plot B **c)** Plot C **d)** Plot D

40. Number of questions correct versus number of questions incorrect on Exam 1
a) Plot A **b)** Plot B **c)** Plot C **d)** Plot D

Questions 41-48 pertain to the number of hours worked per week and percent of tuition paid by parents for 563 female students who responded to Survey 2 this semester. Here are the 5 rounded summary statistics:
Hours Worked: Avg = 6, SD = 5 **% Tuition Parents Paid**: Avg = 64, SD = 20 **Correlation: r = -0.25 (note the negative sign)**
Compute the regression estimates by filling in the blanks in the table below.

Hours Worked	Hours Z Score	r	Tuition % Z Score	% Tuition Parents Paid
16 Hours	Blank 1 Z = 2 $Z = \frac{16-6}{5} = 2$	r = -0.25	Blank 2 Z = -0.5	Blank 3 54 % 64 + (-.5)(20)
Blank 6 4.75 hours	Blank 5 Z = -0.25 6 + (-.25)(5)	r = -0.25	Blank 4 Z = 1	84 $\frac{84-64}{20} = 1$

41. What goes in Blank 1? **a)** 1 **b)** 2 **c)** 5 **d)** -0.25 **e)** -2.4

42. What goes in Blank 2? **a)** 0.5 **b)** 1 **c)** -0.5 **d)** -0.25 **e)** 0.6

43. What goes in Blank 3? **a)** 3.5 **b)** 54 **c)** 74 **d)** 59 **e)** 8.5

44. What goes in Blank 4? **a)** 1 **b)** 1.4 **c)** 2 **d)** -1 **e)** 15.6

45. What goes in Blank 5? **a)** -4 **b)** 0.25 **c)** 1 **d)** 0.75 **e)** -0.25

46. What goes in Blank 6? **a)** 20 **b)** 4.75 **c)** 7.25 **d)** 2.5 **e)** 11

$r_X \frac{SD_Y}{SD_X}$

47. What is the slope of the regression equation when predicting the percent of tuition paid by parents from the number of hours worked?

a) 1 **b)** $\sqrt{1-(-0.25)^2} \times 5$ **c)** $\sqrt{1-(-0.25)^2} \times 20$ **d)** -0.25*(5/20) **e)** -0.25*(20/5)

$\sqrt{1-r^2} \cdot SD_Y$

48. What is the SD of the prediction errors when predicting the percent of tuition paid by parents from the number of hours worked?

a) 1 **b)** $\sqrt{1-(-0.25)^2} \times 5$ **c)** $\sqrt{1-(-0.25)^2} \times 20$ **d)** -0.25*(5/20) **e)** -0.25*(20/5)

Questions 49-54 pertain to the shoe size and the fastest speed ever driven in mph for the 1,046 students who responded to Survey 1. Here are the 5 rounded summary statistics:

	Average	SD
Shoe Size	9	2
Fastest Speed (in mph)	90	20

Correlation: r = 0.25

The regression equation for predicting Fastest Speed from shoe size is: **Fastest Speed** = 2.5 x **Shoe Size** + _____

49. What is the slope of the regression line? **a)** 0.25 **b)** 2.5 **c)** 10 **d)** 19.36 **e)** 0.025

$m = r \times \frac{SD_Y}{SD_X} = 0.25 \times \frac{20}{2}$

50. What is the y-intercept? **a)** 19.36 **b)** 0.25 **c)** 67.5 **d)** 2.5 **e)** -216

$y = 2.5x + b$

$90 = 2.5(9) + b$

$b = 67.5$

51. Now use the regression equation to predict the speed of a student with a size 11 shoe.
a) 90 **b)** 0 **c)** 95 **d)** 60 **e)** none of the above

$\hat{Y} = 2.5(11) + 67.5 = 95$

52. The SD of the prediction errors (the RMSE) when predicting speed from shoe size is closest to…
a) 2 **b)** 19.36 **c)** 1.9 **d)** 17.3 **e)** 2.5

$RMSE = \sqrt{1-.25^2} \times 20 = 19.36$

53. About 68% of the time, our predictions for fastest speed from shoe size will be right within….
a) 1 SD of fastest speed **b)** 1 RMSE **c)** 2 RMSEs **d)** 1 SD of shoe size **e)** none of the above

54. What is the best explanation for why our survey responses show a positive correlation between shoe size and speed?
a) Bigger shoes are longer and can exert more leverage on the gas pedal making the car go faster.
b) Males tend to have bigger feet and males tend to drive faster (Males are the confounder.)
c) Bigger shoes are heavier and tend to exert more pressure on the gas pedal increasing the speed of the car.
d) The correlation coefficient is inflated due to ecological correlations.
e) Clowns often wear very big shoes and clowns tend to drive at high speeds. (Clowns are the confounder)

Questions 55-60 pertain to the following situation:
Suppose a machine contains 8 fair dice-- 4 red, 3 blue and 1 yellow. The machine shakes up the dice and then randomly rolls one out at a time, without replacement (so each is equally likely to land 1, 2, 3, 4, 5, or 6.)

55. What's the probability that the first 2 rolls sum to 5? 2,3 3,2 4,1 1,4
a) 2/36 **b)** 3/36 **c)** 4/36 **d)** 5/36 **e)** 6/36

56. What's the probability that the machine first rolls out a blue?
a) 3/8 **b)** 1/8*1/6 **c)** 1/8 **d)** 3/8*1/6 **e)** 4/8

57. What's the probability that **none** of the first 3 rolls are 6's?
a) $1-(5/6)^3$ **b)** $1-(1/6)^3$ **c)** $(1/6)^3$ **d)** $(5/6)^3$ **e)** 5/6+5/6+5/6

58. What's the probability that the machine first rolls out a blue and that it lands 6?
a) 4/8 **b)** 3/8*1/6 **c)** 3/8 **d)** 1/8*1/6 **e)** 1/6

59. What's the probability that **not all** of the first 3 rolls are 6's? 1 - P(all)
a) $(1/6)^3$ **b)** $1-(1/6)^3$ **c)** $(5/6)^3$ **d)** $1-(5/6)^3$ **e)** 1/6+1/6+1/6

60. What's the probability that the first is a blue **and** the second is a red (remember it's without replacement)?
a) 3/8 + 4/7 **b)** 3/8*4/7 **c)** 7/8 **d)** 3/8+1/6 - (3/8*1/6) **e)** 3/8*4/8

Questions 61-66: The table below shows our class survey responses to the questions "Are you a member of a fraternity or a sorority (Greek)?" and "How many classes do you plan to skip on Unofficial?" Suppose you randomly draw from 776 students who answered this survey:

	How many classes students plan to skip				
	None	Exactly One	Some (more than one but less than all)	All	Total
Greek	112	53	20	69	254
Not Greek	389	65	23	45	522
Total	501	118	43	114	776

61. About 60% of those who plan to skip all classes are Greek, about what percent of Greeks plan to skip all classes? 69/254
a) 45% **b)** 27% **c)** 60% **d)** 40% **e)** 33%

62. What is the chance that you'll get a student who plans to skip all classes?
a) 275/776 **b)** 662/776 **c)** 114/776 **d)** 501/776

$\frac{522}{776} + \frac{501}{776} - \frac{389}{776}$ 522+501-389

63. What's the chance you'll get someone who either answered "Not Greek" or answered "None"?
a) 1 – 1023/776 b) 634/776 c) 522/776*501/776 d) 1023/776 e) 776/766

64. What is the chance that you'll get a student who plans to skip at least one class? = 1 − P(none)
a) 662/776 b) 275/776 c) 501/776 d) 114/776 $= 1 - \frac{501}{776}$

65. What's the chance of drawing 3 students **without** replacement and getting all Greeks?
a) 522/776*521/775*520/774 b) $(522/776)^3$ c) 254/776*253/775*252/774 d) $(254/776)^3$

66. What's the chance of drawing 3 students **with** replacement and getting all Greeks?
a) $(254/776)^3$ b) 522/776*521/775*520/774 c) $(522/776)^3$ d) 254/776*253/775*252/774

Questions 67-68 pertain to the following situation: Suppose you decide to have 3 children. (Assume the same chance of having a boy or a girl each time).

67. What is the probability that you have all girls?
a) $(1/2)^3 + (1/2)^3$ b) $1-(1/2)^3$ c) $(1/2)^3$ d) 1/2+1/2+1/2 e) $(1/2)^3 + (1/2)^3 + (1/2)^3$

68. What is the probability that you have either all boys or all girls?
a) $(1/2)^3 + (1/2)^3$ b) $1-(1/2)^3$ c) $(1/2)^3$ d) 1/2+1/2+1/2 e) $(1/2)^3 + (1/2)^3 + (1/2)^3$

EXAM 2 FORMULAS

IQR=Q3-Q1
Low outliers < Q1 – 1.5*IQR
High outliers > Q3 + 1.5*IQR

Slope of Regression Line = r*SDy/SDx
RMSE = sqrt(1-r^2) * SDy

P(at least one) = 1 - P(none)
P(not all) = 1 - P(all)