EXAM 2: Statistics 100

Fall Key

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 (Tues/Thurs at 12:30 pm)   **ii)** KF (Flanagan Online)   **iii)** ONL (Yu 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 (70 questions).

There is NO CLASS on Thursday!

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.

Fall Page 1

Questions 1 to 5 pertain to the figures below.
Determine whether the correlation coefficient is appropriate for analyzing the plots. If so, choose the r which best represents the plot.

Plot A	Plot B	Plot C	Plot D	Plot E
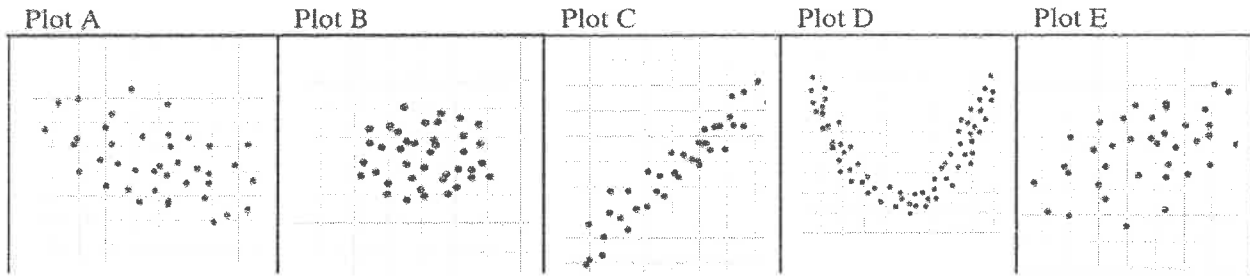
1. Plot A

    a) 0.9   b) 0.5   c) 0   **(d) -0.5**   e) Not appropriate to use r

2. Plot B

    a) 0.9   b) 0.5   **c) 0**   d) -0.5   e) Not appropriate to use r

3. Plot C

    a) 0.9   b) 0.5   c) 0   d) -0.5   e) Not appropriate to use r

4. Plot D

    a) 0.9   b) 0.5   c) 0   d) -0.5   **e) Not appropriate to use r**

5. Plot E

    a) 0.9   **b) 0.5**   c) 0   d) -0.5   e) Not appropriate to use r

Use the following information to answer questions 6-10. For each of the following pairs of variables, select the response that best describes its correlation coefficient, **r**.

6. Y is always exactly 1 less than X.

    a) Exactly -1   b) Between 0 and -1   c) About 0   d) Between 0 and +1   **e) Exactly +1**

7. As X increases, Y tends to increase.

    a) Exactly -1   b) Between 0 and -1   c) About 0   **d) Between 0 and +1**   e) Exactly +1

8. X increases, Y tends to decrease.

    a) Exactly -1   **b) Between 0 and -1**   c) About 0   d) Between 0 and +1   e) Exactly +1

9. X and Y have no relationship, they are just randomly paired.

    a) Exactly -1   b) Between 0 and -1   **c) About 0**   d) Between 0 and +1   e) Exactly +1

10. X and Y always add up to 10

    a) Exactly -1   b) Between 0 and -1   c) About 0   d) Between 0 and +1   e) Exactly +1

Use the following table to answer question 11. The average of X and Y are both 3 and the SD of X and Y are both 2.

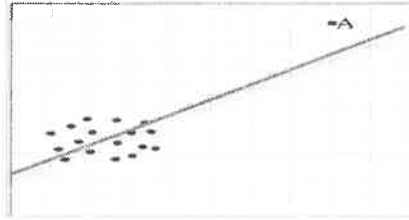| X	Y	Zx	Zy	Products
0	3	-1.5	0	0
2	6	-0.5	1.5	-0.75
3	4	0	0.5	0
4	2	0.5	-0.5	-0.25
6	0	1.5	-1.5	-2.25

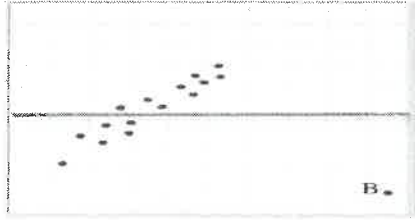**11.** What is the correlation coefficient, r? (Hint: fill out the rest of the table!)

$$r = -0.65$$

    a) -3.25   b) -0.5   c) 0.5   d) -0.75   **e) -0.65**

$$r = \frac{-0.75 + (-0.25) + (-2.25)}{5} = -0.65$$

1

Fall Page 2

Questions 12-13 relate to the 2 scatterplots below.

12. If we removed point A, the correlation coefficient would…

    a) increase   **b)** decrease   **c)** stay the same   **d)** not enough info

13. If we removed point B, the correlation coefficient would…

    a) increase   **b)** decrease   **c)** stay the same   **d)** not enough info

Use this scenario for questions 14-18.

Below is a distribution table for US income (in thousands of dollars) in 2018. The right-hand column shows the % of the population in each interval.

Income (in thousands of $)	%
0-10	5
10-30	20
30-50	25
50-90	15
90-180	10
180 and up	25

14. Median = __50__   **a)** 10 **b)** 30 **c) 50** **d)** 90 **e)** 180

15. Q1= __30__   **a)** 10 **b) 30** **c)** 50 **d)** 90 **e)** 180

16. Q3= __180__   **a)** 10 **b)** 30 **c)** 50 **d)** 90 **e) 180**

17. What percent of the population are low outliers?

    a) 0% **b)** 5% **c)** 20% **d)** 25% **e)** not enough info

low outliers $< Q1 - 1.5(IQR) = 30 - 1.5(150) = -195$

18. What percent of the population are high outliers?

    a) 0% **b)** 5% **c)** 20% **d)** 25% **e)** not enough info

accept a + e

high outliers $> Q3 + 1.5(IQR) = 180 + 1.5(150) = 40$

19. Say we also looked at the income in Canada (also in thousands of dollars) in 2018. I don't have the whole distribution table like we do above, but I know that Q3 is equal to $200,000. What percent of people in Canada made less than that in 2018?

   **a)** Impossible to tell   **b)** 25%   **c)** 50%   **d) 75%**

Question 20-23 pertain to the following information:

- The standard deviations of X and Y are both equal to 3.
- The means of X and Y are both equal to 10.
- The correlation coefficient between X and Y is **r = 0.5**

What would be the new **value** of **r** if….

20. Bert subtracts 5 from each Y value.

    a) r = 0.5   **b)** r = -4.5   **c)** r = -0.5   **d)** r = 0   **e)** not enough information

21. Flannie doubles each X value.

    a) r = 0.5   **b)** r = 1   **c)** r = - 0.5   **d)** r = 0   **e)** not enough information

22. Albie multiplies each X value by -2

    a) r = 0.5   **b) r = - 0.5**   **c)** r = -1   **d)** r = 0   **e)** not enough information

23. Jake (from State Farm) swaps the X and Y columns

    a) r = 0.5   **b)** r = - 0.5   **c)** r = 0   **d)** r = 1   **e)** not enough information

The following situation is used for questions 24-30. The graph below plots the number of violent crimes per 100,000 people vs. the percent of people living in poverty. The 2 straight lines (Line 1 and Line 2) are the regression and SD lines.

Poverty versus Crime

Utah is on Line 1 and New York is on Line 2
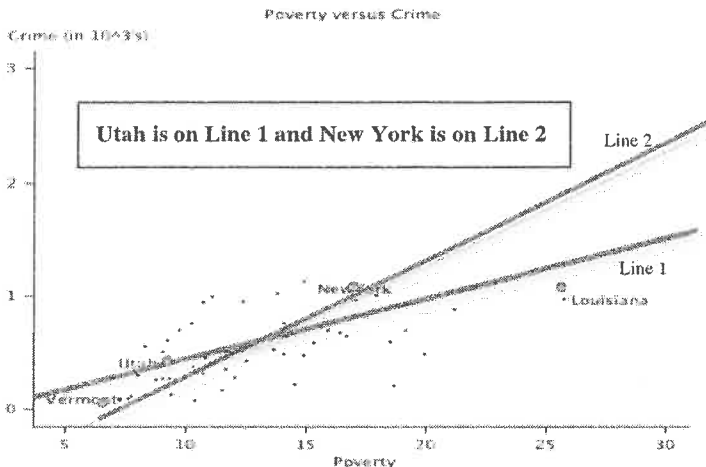
24. Which line is the regression line?
a) Line 1   b) Line 2   c) Unable to determine

25. The correlation coefficient r is closest to…
a) -0.9   b) -0.5   c) 0.5   d) 0.9   e) 0.2

26. What is the average poverty rate, (in %)?
a) 5   b) 9   c) 13   d) 18   e) 0.7

27. Which state has a prediction error = 0?
a) Vermont   b) Utah   c) New York   d) Louisiana

28. Which state has the same z-scores for Poverty and Crime? a) Vermont   b) Utah   c) New York   d) Louisiana

29. The largest residual is 1,000,000 and the smallest residual is -500,000. What's the average of all the residuals?
a) 0   b) 250,000   c) 2.5   d) 1   e) 100,000

30. The above graph has 50 points, one for each individual state's poverty and crime level. If we divided the 50 states into 9 geographical regions and calculated the average poverty and crime level within each region, we'd condense the 50 points into 9 points. Would the correlation coefficient for the 9 points be the same as for the 50 points?
a) Yes, it would be exactly the same since it's the same information.
b) No, it would probably be lower since we have fewer points.
c) No, it would probably be higher since there would be less points and less scatter.

Questions 31-36 pertain to this scenario: Suppose the correlation between IQ scores and ACT scores have the following summary statistics among the population of students who take both exams and that the scatter plot is football-shaped:

	Average	SD
IQ	100	15
ACT	20	5

Correlation: r = 0.8

In the table below you are given the IQ score of Roger. Compute the regression estimate for Roger's ACT score. You are also given Bernard's ACT score. Compute the regression estimate for Bernard's IQ score. Fill out the chart below! Make sure your numbering is correct on your scantron.

IQ Score	IQ Z Score	r	ACT Z Score	ACT Score
Roger's IQ = 130 $Z = \frac{130-100}{15} = 2$	**31.** Z = __2__ ✗ a) -0.6 b) 22 c) 0.6 d) 2 e) 1	r = 0.8	**32.** Z = __1.6__ a) 1.6 b) 2.8 c) 0.75 d) 0.8 e) 2.25	**33.** ACT = __28__ Val = 20 + 1.6(5) a) 34 b) 28 c) 30 d) 36 e) 24
34. IQ = __88__ Val = 100 + (-0.8)(15) a) 81 b) 96 c) 88 d) 112 e) 24	**35.** Z = __-0.8__ a) -0.42 b) -0.27 c) 0.8 d) -1.25 e) -0.8	r = 0.8	**36.** Z = __-1__ a) -1 b) -5.67 c) 1 d) -0.33 e) 0.33	Bernard's ACT = 15 $Z = \frac{15-20}{5} = -1$

Using the same summary statistics from the previous page, answer questions 37-43.

	Average	SD
IQ	100	15
ACT	20	5

Correlation: r = 0.8

37. What is the slope of the regression equation when predicting IQ scores from ACT scores?

a) 3   **b) 2.4**   c) 0.8   d) 0.333   e) 0.267

$slope = r \times \frac{SD_Y}{SD_X} = 0.8 \times \frac{15}{5}$

38. What is the SD of the prediction errors (the RMSE) when predicting IQ scores from ACT scores?

a) 5   b) 15   **c) $\sqrt{1 - 0.8^2} * 15 = 9$**   d) $\sqrt{1 - 0.8^2} * 5$ e) 4

39. The regression equation predicts an IQ score of 112 for those who score 25 on the ACT. Of course not everyone with an ACT of 25 will score 112 on the IQ test. Instead there's a range of IQ scores, with about 68% of them scoring between…

a) 85 and 115   b) 100 and 124   **c) 103 and 121**   d) 97 and 127

prediction $\pm$ 1RMS

$112 \pm 1(9)$

40. The regression equation for predicting ACT scores from IQ scores is: ACT = 0.267 x (IQ) + _____ .
What is the y-intercept? (the y-intercept is the blank in the above equation.)

a) 94.66   b) -94.66   **c) -6.7**   d) 6.7   e) 0
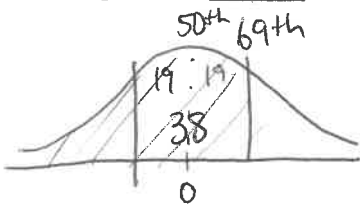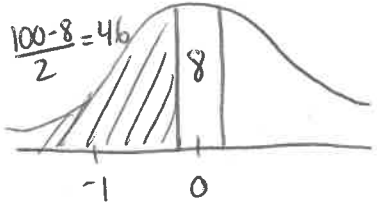
$20 = 0.267(100) + b$
$b = -6.7$

41. Let's say we have a student named Rodney and the regression line predicted that he'd get a 29 on the ACT. This was a pretty good prediction because Rodney's residual (prediction error) was only -1. What was Rodney's actual score?

a) 29   **b) 28**   c) 30   d) 15   e) -29

residual = actual - predict
$-1 = actual - 29$

42. The slope for predicting ACT scores from IQ scores is 0.267 (see question 40). How do you interpret the slope?

a) For every one unit increase in ACT score, IQ score increases 0.267 points on average.
b) For every one unit increase in IQ score, the ACT score increases 0.267 points on average.
c) On average, if you get a 0 on the ACT, your IQ score will be 0.267.
d) On average, if you get a 0 on the IQ test, your ACT score will be 0.267.
e) The correlation between ACT score and IQ score is 0.267.

43. In general, the larger the RMSE, the _____ our predictions are.

a) better   **b) worse**   c) unable to determine   d) RMSE tells us nothing about predictions

Questions 44-47. Suppose science skills and social studies skills of grade school children follow the normal curve but have different correlations among different populations. Consider a population where r = -0.2. Given that Douglas is in the 69th percentile for science skills, make a prediction for his percentile for social skills.

Science Skills Percentile	Science Z	r	Social Z	Social Skills Percentile
DOUGLAS is in the 69th percentile for science skills. 44. What middle area on the normal curve does that correspond to? __38__ % a) 9 b) 18 c) 31 **d) 38** e) 69	45. Z = __0.5__ $\times$ r = - 0.2 a) 1 b) 0.4 **c) 0.5** d) -0.70 e) 0.25		46. Z = __-0.1__ **a) -0.1** b) 0.2 c) -0.5 d) -0.25 e) -1	47. DOUGLAS Social Skills Percentile = __46th__ $\frac{100-8}{2} = 46$ a) 4th. b) 8th. c) 42nd **d) 46th** e) 54th

Questions 48-51: Consider 4 populations where the correlation coefficients between science and social studies skills are as given in the table below. If a child is in the **80th percentile in science skills**, estimate her percentile in social studies skills in each population. **For Example:**

Science Skills Percentile	r	Social studies skills percentile
80th	(provided)	?

48. $r = 1$. Social studies percentile = _____
    a) 20th   b) 50th   c) 66th   **(d) 80th**   e) 100th

49. $r = 0.5$. Social studies percentile = _____
    a) 20th   b) 33rd   c) 50th   **(d) 66th**   e) 80th

50. $r = 0$. Social studies percentile = _____
    a) 0th   b) 20th   **c) 50th**   d) 80th   e) 100th

51. $r = -1$. Social studies percentile = _____
    a) -1st   b) 0th   **c) 20th**   d) 50th   e) the limit does not exist!

Use the following information to answer questions 52-56. The question, "Are you a good person?" and which Harry Potter House would you say you'd be in was asked of 100 random students. Results are shown in the table.

	Yes	No	Total
Hufflepuff	5	40	45
Slytherin	55	0	55
Total	60	40	100

52. What is the probability that a randomly chosen student claims to be a good person?
    a) 35/100   b) 55/100   c) 25/100   **d) 60/100**   e) 100/100

53. What is the probability that a randomly chosen student claims to be a good person and is Slytherin?
    a) 5/100   b) 45/100   **c) 55/100**   d) 60/100   e) 100/100

54. What is the probability that a randomly chosen student is a Hufflepuff or not a good person? $\dfrac{45}{100} + \dfrac{40}{100} - \dfrac{40}{100}$
    a) 5/100   b) 60/100   **c) 45/100**   d) 5/40   e) 5/55

55. What is the probability of a student claiming to be a good person given that they are a Hufflepuff?
    a) 5/60   b) 5/100   **c) 5/45**   d) 5/55   e) 60/100

56. Given that a student claim to be a good person, what is the probability that they are in Slytherin?
    a) 55/55   b) 5/100   c) 5/45   d) 5/55   **e) 55/60**

Use the following information to answer questions 57-59. It seems like everyone on campus is sick (notice the coughing?)! Let's say that 75% of people go to McKinley and get tested for strep actually have it. Suppose 92% of students who have strep will correctly get a positive result. 2% of students who do not have strep will also get a positive result. Fill in the following table and then answer the questions below.

	Tests Positive	Tests Negative	Total
Has Strep	Blank 1 690	60	750
Does Not Have Strep	5	Blank 2 245	250
Total	695	305	1,000

57. What goes in Blank 1?   a) 750   b) 536   **c) 690**   d) 15   e) 200

58. What goes in Blank 2?   a) 750   b) 536   c) 690   d) 15   **e) 245**

59. In general, is the probability of A given B always the same as the probability of B given A?   a) Yes   **b) No**   c) It depends

Question 60-63 pertain to a well-shuffled deck of 52 cards. A deck of cards has 4 suits: clubs, diamonds, hearts and spades. There are 13 cards in each suit: 2 through 10, jack, queen, king, ace. (So there are 4 Aces, 4 Jacks, and 13 Hearts.)

60. Draw 2 cards with replacement. What is the chance that the first card is an Ace and the second is a Heart? $\frac{4}{52} \times \frac{13}{52}$

    a) 4/52 + 13/52   b) 4/52 + 13/52 -1/52   c) 4/52*3/51   d) 4/52*13/51   **e) 4/52*13/52**

61. Draw 2 cards without replacement. What is the chance that the both cards are Aces?

    a) 4/52 + 4/52   b) 4/52 + 4/52 -1/52   **c) 4/52*3/51**   d) 4/52*4/51   e) 4/52*4/52

62. Draw 2 cards without replacement. What is the chance that the first card is an Ace and the second is a Jack?

    a) 4/52 + 4/52   b) 4/52 + 4/52 -1/52   c) 4/52*3/51   **d) 4/52*4/51**   e) 4/52*4/52

63. Draw one card. What is the chance that it's either a Jack or a Heart?

    a) 4/52 + 13/52   **b) 4/52 + 13/52 -1/52**   c) 4/52*3/51   d) 4/52*13/51   e) 4/52*13/52

Question 64 – 70 pertains to rolling fair dice.

64. Two dice are rolled. What is the chance that the sum of the dots is 4?
$\quad \begin{matrix} 2,2 \\ 3,1 \\ 1,3 \end{matrix}$

    a) 2/36   **b) 3/36**   c) 4/36   d) 5/36   e) 1/6*1/6

65. One die is rolled 4 times. What is the chance of getting all 4's?

    a) $(5/6)^4$ **b) $(1/6)^4$**   c) $1- (5/6)^4$ d) $1- (1/6)^4$ e) 4/6

66. One die is rolled 4 times. What is the chance of not getting all 4's?

    a) $(5/6)^4$ b) $(1/6)^4$ c) $1- (5/6)^4$ **d) $1- (1/6)^4$**   e) 4/6

67. One die is rolled 3 times. What is the chance of getting no 2's?

    a) $(5/6)^3$ b) $(1/6)^3$ c) $1- (5/6)^3$ d) $1- (1/6)^3$ e) 3/6

68. One die is rolled 3 times. What is the chance of getting at least one 2?

    a) $(5/6)^3$ b) $(1/6)^3$ **c) $1- (5/6)^3$**   d) $1- (1/6)^3$ e) 3/6

69. One die is rolled twice. What is the chance that the first roll is a 2 and the second roll is a 3?

    a) 1/36   b) 2/36   c) 6/36   d) 11/36   e) 12/36

70. One die is rolled twice. What is the chance that the first roll is a 2 or the second roll is a 3?

    a) 1/36   b) 2/36   c) 6/36   **d) 11/36**   e) 12/36

EXTRA CREDIT

71. This extra credit question pertains to zodiac signs. There are 12 zodiac signs. Assume that each of the signs is equally likely. In a group of 5 random people, what is the chance that there's at least 1 zodiac match? In other words, what's the chance that at least 2 people share a sign? Round your answer to two decimal places

    a) 35.27%   b) 99.99%   c) 76.39%   d) 41.67%   **e) 61.81%**

STANDARD NORMAL TABLE

Standard Units
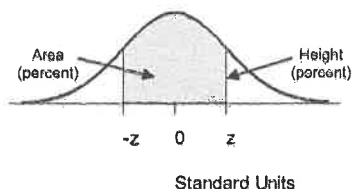
z	Area		z	Area		z	Area
0.00	0.00		1.50	86.64		3.00	99.730
0.05	3.99		1.55	87.89		3.05	99.771
0.10	7.97		1.60	89.04		3.10	99.806
0.15	11.92		1.65	90.11		3.15	99.837
0.20	15.85		1.70	91.09		3.20	99.863
0.25	19.74		1.75	91.99		3.25	99.885
0.30	23.58		1.80	92.81		3.30	99.903
0.35	27.37		1.85	93.57		3.35	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