Exam 1 Statistics 100     Version B Key     **Fall 2021**

**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)? _____

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. <u>**Circle all answers on this exam and fill in the**</u> <u>**corresponding bubble on your orange scantron**</u>. 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. <u>If you don't do both, you will get a 0.</u>

---

**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 the date in the Date box.
- Print and bubble in your NET ID with **no spaces** in the NETWORK ID box. Print and bubble in the Section Box. See section codes on the projector.
- *Write Stat 100 on the COURSE line.*
- *Write your instructor's name* (Karle Flanagan) on the INSTRUCTOR line.
- *Write your section on the SECTION line.*
- Sign your name, and right underneath the student signature line <u>PRINT</u> 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 8 pages including the normal table (67 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 1-4 pertain to the following situation:**
As you know, the U of I developed a saliva test that is convenient, repeatable, and large-scale to use as a key weapon to help control the COVID-19 pandemic. They released a paper with some data out of the saliva test! There were 100 students who were tested for this sample. Of the 100 samples, 10 students actually had COVID. 90% of people with COVID-19 will correctly test positive. However, 10% of people without COVID-19 will also test positive. This is called a false positive. Fill in the following table for this sample of 100 students, rounding to the nearest whole number.

| | Tested Positive | Tested Negative | Total |
|---|---|---|---|
| **Actually Had COVID-19** | 9 | **1. What number should go in this blank?** <br> a) 1  b) 8  c) 9  d) 2 | **2. What number should go in this blank?** <br> a) 9  b) 10  c) 90  d) 1 <br> 90 |
| **Did NOT have COVID-19** | **3. What number should go in this blank?** <br> a) 1  b) 9  c) 10  d) 90 | 81 | 90 |
| **Total** | 18 | 82 | 100 |

4. Given that a student receives a positive test result, what's the chance that they have COVID-19?
   a) 1/9    b) 8/9    c) 9/18    d) 9/100    e) 9/90

5. Suppose a student does not have COVID-19. What is the chance that they will get a negative test result?
   a. 1/91    b) 81/90    c) 1/100    d) 10/90    e) 90/100

**The next 3 questions 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. In other words, there are 13 hearts, 4 kings, and one king of hearts.

6. Draw 2 cards with replacement. What is the chance that both cards are kings?
   a. 13/52 * 13/52    b) 13/52 * 12/51    c) 8/52    d) 4/52 * 4/52    e) 4/52 * 3/51

7. Draw 2 cards without replacement. What is the chance that both cards are hearts?
   a. 13/52 * 13/52    b) 13/52 * 12/51    c) 8/52    d) 4/52 * 4/52    e) 4/52 * 3/51

8. Draw one card at random. What's the chance that it's either a king or a heart?
   a. 8/52    b) 16/52    c) 17/52    d) 26/52    e) 13/52 * 4/52

**The following 7 questions pertain to the table below.**
This table shows the survey responses of the 290 males and the 474 females to the question on survey 3 last semester: "Will you go to any of your classes drunk on Unofficial?". Suppose you draw randomly from the students who answered this survey.

| | No | Possibly | Probably | Certainly | Total |
|---|---|---|---|---|---|
| **Males** | 204 | 35 | 30 | 21 | 290 |
| **Females** | 349 | 76 | 24 | 25 | 474 |
| **Total** | 553 | 111 | 54 | 46 | 764 |

9. What is the chance of getting a student certain to attend a class drunk?
   a. 46/764    b) 21/290    c) 25/474    d) 25/46    e) 100/764

**10.** What is the chance of getting a student certain to attend a class drunk if you draw only from the females?

    a. 46/764      b) 21/290      c) 25/474      d) 25/46      e) 100/764

**11.** What is the chance that a randomly selected student will either probably or certainly come to a class drunk?

    a. 46/764      b) 21/290      c) 25/474      d) 25/46      e) 100/764

**12.** What is the chance that a randomly selected student is either female or certain to come to a class drunk?

    a. 46/764      b) 520/764      c) 25/474      d) 495/764      e) 100/764

**13.** Draw 3 students without replacement. What is the chance that not all of them are certain to attend a class drunk?

    a. $46/764 * 45/763 * 44/762$

    b. $46/764 + 45/763 + 44/762$

    c. $1 - 46/764 * 45/763 * 44/762$

    d. $1 - (46/764)^3$

**The next 4 questions pertain to rolling fair dice.**

*36 possibilities*

    *1,4   3,2*

**14.** Two dice are rolled. What is the chance that the sum of the spots is 5?   *4,1   2,3*

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

**15.** One die is rolled 3 times. What is the chance of getting all 6's?

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

**16.** One die is rolled 3 times. What is the chance of not getting all 6's?

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

**17.** One die is rolled 3 times. What is the chance of getting at least one 6?

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

**Questions 46-49 pertains to the following situation:**

If Steve (Karle's husband) is in the 80th percentile for height, where r = 0.6, what percentile would you estimate for his wife's (Karle's) height? Solve by filling in the table below.

| Husband's Percentile for Height | Husband's z-score | r = 0.6 | Wife's z-score | Wife's Percentile for Height |
|---|---|---|---|---|
| 80th percentile<br><br>**18.** What middle area on the table should you look to find the z-score?<br><br>a) 30   b) 50   c) 60   d) 80 | **19.** Z = _____<br><br>a) 0.7   b) 0.85<br>c) 1.3   d) 0.4 | r = 0.6 | **20.** Z = _____<br><br>a) 0.25   b) 0.4<br>c) 0.5   d) 0.8 | **21.** Wife's height percentile is _____<br><br>a) 38   b) 69   c) 60<br>d) 19   e) 31 |

**Questions 22-24 pertain to the following scenario:**
While Karle is water skiing, Steve is busy finding new data sets to analyze! Suppose Steve finds a data set that shows that husbands' and wives' heights follow the normal curve, but have different correlations in different countries.

22. If a husband is in the 15th percentile for height, estimate his wife's percentile in height if r = -1.
    a) 15th        b) 34th        c) 50th        d) 66th        e) 85th

23. If a husband is in the 15th percentile for height, estimate his wife's percentile in height if r = 1.
    a) 15th        b) 34th        c) 50th        d) 66th        e) 85th

24. If a husband is in the 15th percentile for height, estimate his wife's percentile in height if r = 0.
    a) 15th        b) 34th        c) 50th        d) 66th        e) 85th

**The next 7 questions pertain to the following situation:**
Steve and Karle love going to the lake in the summer! Karle LOVES water skiing so much that Steve thinks that there is a linear relationship between the amount of time Karle spends water skiing and how happy she feels. He records how long she water skis and how happy she feels on a scale of 0-10 (0 being not happy at all and 10 being very happy) for each week during the summer of 2020. The plot and summary statistics are shown below.
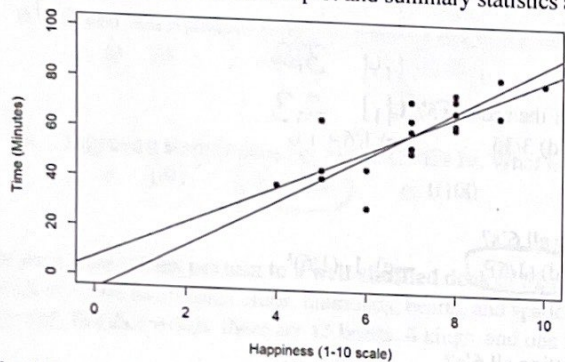


|  | Average | SD |
|---|---|---|
| Y Time | 60.0 | 15 |
| X Happiness | 7.0 | 1.6 |
| Correlation | 0.8 | |

Find the regression equation for predicting the amount of time Karle spends water skiing from her happiness level.

25. Start by finding the slope of the regression equation for predicting time spent water skiing from happiness level
    a) 0.085        b) 0.107        c) 7.5        d) 9.375

26. Now find the y-intercept for the regression line for predicting time spent water skiing from happiness level
    a) -10.3        b) -443        c) 1.88        d) 6.36        e) 7.5

27. What is the RMSE for predicting time spent water skiing from happiness level?
    a) 0.96        b) 2.32        c) 4.2        d) 9        e) 36

28. If both variables follow the normal curve, 87% of the time our predictions will be correct within _____ RMSEs. What goes in the blank?
    a) 0        b) 1        c) 1.5        d) 1.65        e) 2

29. If Karle's happiness level is at 5, how many minutes do we predict she will spend water skiing? *You can use your regression equation of the 3-step process.*
    a. 45        b) 78.75        c) 41.25        d) 60        e) 75

**30.** The range for our predictions to be correct 87% of the time is...
    a)  45 - 87       b) 30.15-59.85       c) 31.5 – 58.5     d) 4.36 - 9.64     e) 43.5 – 46.5

**31.** Suppose Steve observes that during one week, Karle actually spent 72 minutes water skiing when her happiness level is at 7 rather than the predicted time of 60 minutes. What is Karle's prediction error?
    a)  -12       b) 12       c) 0       d) Cannot be determined from the information given

**32.** If we now want to predict happiness from time spent water skiing, which of the following statistics will change?
    a)  Slope of the regression line     b) Y-intercept     c) RMSE     d) All of the above

For each of the following pairs of variables in the next 5 questions, choose the option that best describes its correlation coefficient, r. *Each option will be used exactly once!*

**33.** As X gets larger, corresponding Y values get smaller
    a)  Exactly +1     b) Between 0 and +1     c) About 0     d) Between 0 and -1     e) Exactly -1

**34.** X and Y are paired by a random lottery
    a)  Exactly +1     b) Between 0 and +1     c) About 0     d) Between 0 and -1     e) Exactly -1

**35.** Number of STAT 100 videos you watch and your score on the final
    a)  Exactly +1     b) Between 0 and +1     c) About 0     d) Between 0 and -1     e) Exactly -1

**36.** Y is always equal to -1 times X
    a)  Exactly +1     b) Between 0 and +1     c) About 0     d) Between 0 and -1     e) Exactly -1

**37.** Height in inches and height in centimeters
    a)  Exactly +1     b) Between 0 and +1     c) About 0     d) Between 0 and -1     e) Exactly -1

**Questions 38 and 39 pertain to the following figures:**
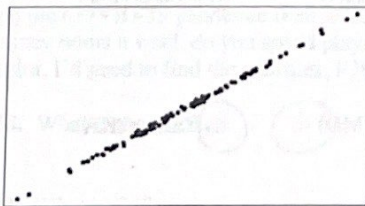


**38.** Is r an appropriate statistic to use in the plot below? Mark 'yes' if r is appropriate or 'no' if it is not.
    a)  No
    b)  Yes
    c)  Cannot be determined from the information given
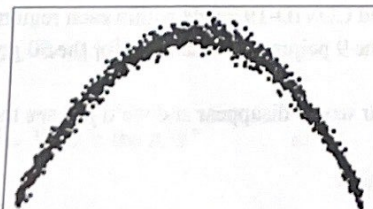


**39.** Is r an appropriate statistic to use in the plot below? Mark 'yes' if r is appropriate or 'no' if it is not.
    a)  No
    b)  Yes
    c)  Cannot be determined from the information given

**The next 3 questions pertain to the following situation:**
X and Y are 2 sets of numbers with a correlation coefficient of r = 0.75

40. If we divide all the X values by -3, r will be...
    a) -0.75          b) 0.75          c) Impossible to tell

41. If we remove one X value and its corresponding Y value, r will be....
    a) - 0.75          b) 0.75          c) Impossible to tell

42. If we switch all the X values and Y values, r will be...
    a) -0.75          b) 0.75          c) Impossible to tell

**The next 9 questions refer to the scatter plot below:**
The scatter plot below shows the percent of people who were confirmed with COVID-19 versus the population density for the 50 states in the U.S. and Puerto Rico (the District of Columbia was omitted since it is an outlier). Oklahoma (OK) is on Line 1 and Maryland (MD) is on Line 2
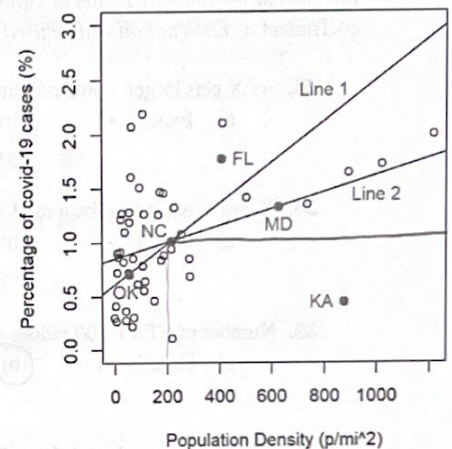
43. The average population density (in p/mi²) is closest to...
    a) 217          b) 1.0          c) 600          d) 82

44. The average number of confirmed COVID-19 cases (in %) is closes to....
    a) 217          b) 0.75          c) 1.0          d) 1.5

45. Which is the SD line?
    a) Line 1          b) Line 2

46. The correlation between the percentage of COVID-19 cases and the population density is closest to...
    a) 0.4          b) -0.4          c) 0          d) 0.8          e) -0.8

47. Which of the labeled states has a residual that equals 0?
    a) Oklahoma (OK)          b) Florida (FL)          c) Maryland (MD)          d) Kansas (KA)

48. Which state has the same z-score for COVID-19 cases and population density?
    a) Oklahoma (OK)          b) Florida (FL)          c) Maryland (MD)          d) Kansas (KA)

49. Which state has the largest prediction error?
    a) Oklahoma (OK)          b) Florida (FL)          c) Maryland (MD)          d) Kansas (KA)

50. The above graph has 50 points, one for each individual state's COVID-19 and density levels. If we divided the states into 9 geographical regions and calculated the average density and COVID-19 levels within each region, we'd condense 50 into 9 points. Would the correlation coefficient for the 9 points be the same as for the 50 points?
    a) No, it would probably be lower since we have fewer points
    b) No, it would probably be higher since the within-region scatter would disappear and we'd just see the between-region scatter
    c) Yes, it would be exactly the same since it's the same information

**51.** Looking at the scatter plot, about how many COVID-19 cases would we predict for a population density of 200?

     a) 1.5%      (b) 1.0%)      c) 2.25%      d) 0%

**52.** The average of the residuals is _____.

     a) unknown      b) sometimes 0      (c) always 0      d) always a different number

**The next 5 questions below refer to this data. Compute the correlation coefficient (r) by filling in the table below.**
The average of X = -5 and the SD of X = 2. The average of Y = 5 and the SD of Y = 2.   $z = \dfrac{val - avg}{SD}$

| X | Y | Z-score for X | Z-score for Y | Products $z_x \cdot z_y$ |
|---|---|---|---|---|
| -2 | 4 | **53.** What number goes in this blank? <br> a) 0   (b) 1.5)   c) -0.5 <br> d) 0.5   e) -1.5 | -0.5 | -0.75 |
| -4 | 6 | 0.5 | **54.** What number goes in this blank? <br> a) 0   b) 1.5   c) -0.5 <br> (d) 0.5)   e) -1.5 | 0.25 |
| -5 | 5 | 0 | 0 | 0 |
| -6 | 2 | -0.5 | -1.5 | 0.75 |
| -8 | 8 | -1.5 | 1.5 | **55.** What number goes in this blank? <br> a) 1.5   b) -1.5   (c) -2.25) <br> d) 0   e) None of the above |
| Totals | | Total should = 0 | Total should = 0 <br> **56.** What number goes in this blank? <br> a) 1   b) -0.5   c) 2 <br> (d) 0   e) None of the above | Total = -2 |

**57.** What is the correlation coefficient, r?

     a) 0.5 <br>
     (b) -0.4) <br>
     c) -0.8 <br>
     d) 0.25 <br>
     e) None of the above

**The following situation pertains to questions 58-63:**
During the COVID-19 pandemic (Fall 2020), I asked 9 random students on the quad the following question: "On average, how many hours a week do you spend playing the game "Among Us"?" Here is their data: 3, 5, 12, 7, 6, 0, 3, 4, 6. To draw a boxplot, I'd need to find the quartiles, IQR, and check for outliers.

0, 3, 3, 4, 5, 6, 6, 7, 12

**58.** What is the median?      (a) 5)      b) 6      c) 5.11      d)

**59.** What is Q1?      a)      b)      (c) 3)      d)

**60.** What is Q3?      a)      (b) 6.5)      c)      d)

*accept any answer*

**61.** What is the IQR?      a)      b)      c)      (d) 3.5)

62. Are there any low outliers?     a) Yes     (b) No)     c) Cannot determine from the information given

low outliers $< Q1 - 1.5(IQR) = 3 - 1.5(3.5) = -2.25$

63. Are there any high outliers?     (a) Yes)     b) No     c) Cannot determine from the information given
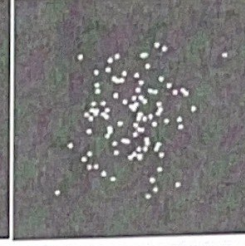
high outliers $> Q3 + 1.5(IQR) = 6.5 + 1.5(3.5) = 11.75$
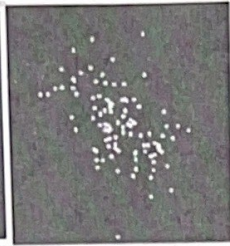
**The following 4 questions pertain to these 4 scatterplots:**

| Plot A | Plot B | Plot C | Plot D |
|---|---|---|---|



Each plot will be used once.

64. Which plot matches with r=0.51?
   a. Plot A
   b. Plot B
   c. Plot C
   (d. Plot D)

65. Which plot matches with r= - 0.76
   a) Plot A)
   b) Plot B
   c) Plot C
   d) Plot D

66. Which plot matches with r=0.87?
   a. Plot A
   b. Plot B
   (c. Plot C)
   d. Plot D

67. Which plot matches with r= - 0.22
   a) Plot A
   (b) Plot B)
   c) Plot C
   d) Plot D

## EXAM 2 FORMULAS

$IQR = Q3 - Q1$

Low outliers $< Q1 - 1.5*IQR$

High outliers $> Q3 + 1.5*IQR$

Slope of Regression Line $= r*SD_y/SD_x$

$RMSE = sqrt(1-r^2) * SD_y$

Residual $=$ actual $-$ predicted

$r = [sum (Z_x * Z_y)] / n$