READ THE DIRECTIONS BELOW TWICE!

Cover Sheet Questions

1) What's your **name**? FORM   A   ANSWER   KEY

                           (Last name)   (First name)

2) What's your **net ID** (email)? _____@illinois.edu

3) Which **section** are you in? *Circle one:*

      i) L2 (In Person Section)   ii) O1 (Online Section)

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) on the INSTRUCTOR line.*
- *Write your section (L2 or O1) 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.

There is NO CLASS on Friday this week! Have a wonderful Spring Break ☺

Scores will be posted on Canvas by Monday at noon. Students may pick up their exam in 171 Computing Applications Building during office hours next week.

Questions 1-6 pertain to the 2 box plots below: The 2 box plots below depict the survey responses of 365 males and 765 females to the question: "On a scale of 0 to 10, rate how strongly you believe in ghosts. (0 is not at all and 10 is extremely). All answers are whole numbers.

Ghosts: Splitting on Gender
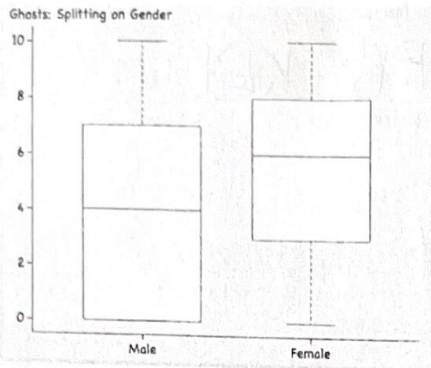
	Males	Females
Average	4.135	5.417
SD	3.515	3.381
Min	0	0
Q1	Blank 1	3
Median	4	Blank 2
Q3	7	8
Max	10	Blank 3
IQR	Blank 4	5
n	365	765

1. What goes in Blank 1?
   a) 0 ⟵   b) 1   c) 3   d) 6   e) 10

2. What goes in Blank 2?
   a) 0   b) 1   c) 3   (d) 6   e) 10

3. What goes in Blank 3?
   a) 0   b) 1   c) 3   d) 7   (e) 10
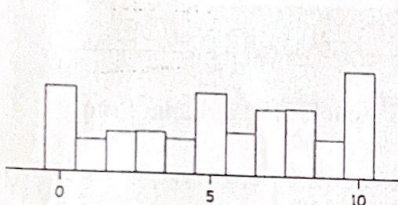
4. What goes in Blank 4?
   a) 0   b) 1   c) 3   (d) 7   e) 10

5. Are there any outliers for males or females?   a) Yes, only for males   b) Yes, only for females   c) Yes, for both   (d) No
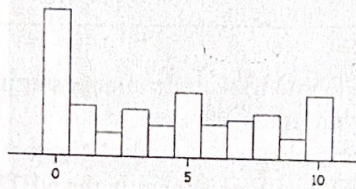
6. The histograms below depict the ghost ratings of the 365 males and 765 females. Which histogram depicts the male responses?
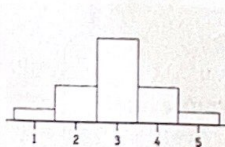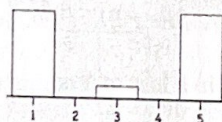   a) A   (b) B

Histogram A   Histogram B

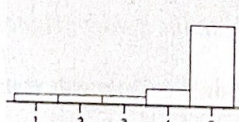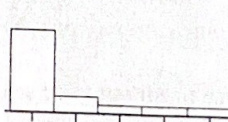**Questions 7-10 pertain to the 5 histograms and box plots below.** Which histograms correspond to which box plots?

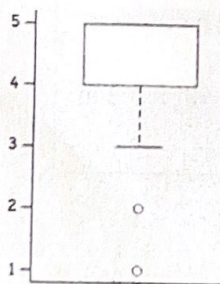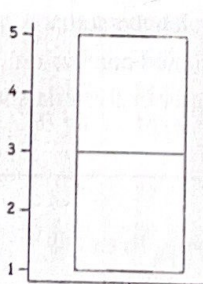Histogram A   Histogram B   Histogram C   Histogram D

Box Plot 1   Box Plot 2   Box Plot 3   Box Plot 4
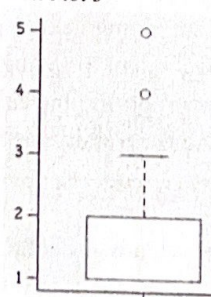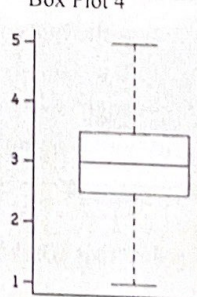
7. Which histogram corresponds to Box Plot 1?   a) Histogram A   b) Histogram B   (c) Histogram C   d) Histogram D

8. Which histogram corresponds to Box Plot 2?   a) Histogram A   (b) Histogram B   c) Histogram C   d) Histogram D

9. Which histogram corresponds to Box Plot 3?   a) Histogram A   b) Histogram B   c) Histogram C   (d) Histogram D

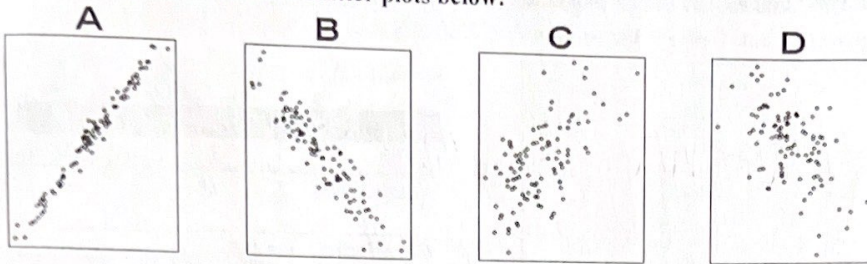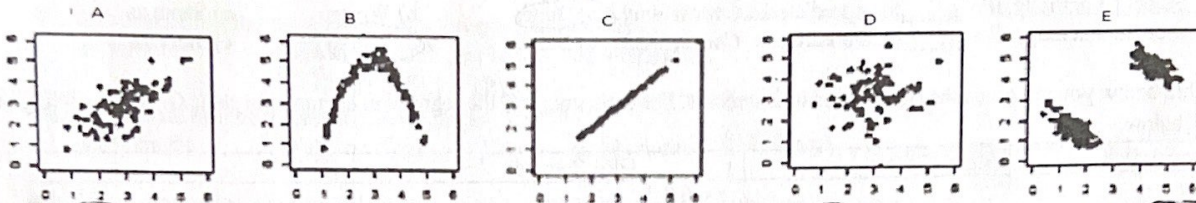10. Which histogram corresponds to Box Plot 4?   (a) Histogram A   b) Histogram B   c) Histogram C   d) Histogram D

Questions 11-14 pertains to the 4 scatter plots below: Spring 2023

A **B** **C** **D**

11. The correlation of the points in scatterplot A is? a) 0.48 b) -0.85 c) -0.4 **d) 0.92** (circled)
12. The correlation of the points in scatterplot B is? a) 0.48 **b) -0.85** (circled) c) -0.4 d) 0.92
13. The correlation of the points in scatterplot C is? **a) 0.48** (circled) b) -0.85 c) -0.4 d) 0.92
14. The correlation of the points in scatterplot D is? a) 0.48 b) -0.85 **c) -0.4** (circled) d) 0.92

Questions 15-19: Is r an appropriate statistic to use in the plots below? Under each plot circle YES if r is appropriate or NO if it's not.

A B C D E

15. **a) YES** (circled) b) NO   16. a) YES **b) NO** (circled)   17. **a) YES** (circled) b) NO   18. **a) YES** (circled) b) NO   19. a) YES **b) NO** (circled)

Questions 20-24 pertain to the table below: For each of the following pairs of variables, check the box that best describes its correlation coefficient **r**.

	Exactly +1	Between 0 and +1	Exactly 0	Between 0 and -1	Exactly -1
20. The slope of the regression line for predicting Y from X is 0. (Assume SD's $\neq 0$)	a) ☐	b) ☐	c) ☑	d) ☐	e) ☐
21. The regression line has a negative slope and a RMSE = 0	a) ☐	b) ☐	c) ☐	d) ☐	e) ☑
22. Y is *always exactly* 0.5 times X	a) ☑	b) ☐	c) ☐	d) ☐	e) ☐
23. Years of education and income among US adult men.	a) ☐	b) ☑	c) ☐	d) ☐	e) ☐
24. Temperature and heating bill	a) ☐	b) ☐	c) ☐	d) ☑	e) ☐

25. If x and y are 2 sets of numbers with the same average and same SD, what do we know about the correlation coefficient, r?
    a) It's positive   b) It's negative   c) It's 0   **d) This information does not tell us anything about r** (circled)

Questions 26-28 pertain to 2 sets of numbers. X and Y are 2 sets of numbers with a correlation coefficient of **r = 0. 3**.

26. If all the original X-values are multiplied by *negative* -2, the new correlation coefficient would be?
    a) 0.3   **b) -0.3** (circled)   c) impossible to tell

27. If all the original X values are increased by 10%, the new correlation coefficient would be?
    a) 0.3 (circled)   b) -0.3   c) impossible to tell

28. If all the X and Y values are switched, the new correlation coefficient would be?
    a) 0.3 (circled)   b) -0.3   c) impossible to tell

Questions 29-37 pertain to the scatterplot below: The scatter plot below shows the GPA and hours studied per week for 16 students. The line shown is the regression line.
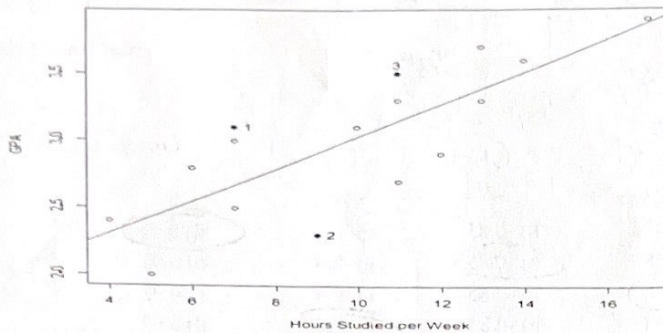
	Average	SD
GPA	3.0	0.5
Study Hours	10	4

Correlation: $r = 0.8$

Look at students **1 and 2** on the graph. How did their actual GPA's compare to their predicted scores? For each student circle whether their actual GPA was better than, worse than, or the same as the regression line predicted from how many hours per week they studied.

29. Student 1 actually did _____ than predicted. *Choose one:* **(a) Better**   b) Worse   c) Same as
30. Student 2 actually did _____ than predicted. *Choose one:* a) Better   **(b) Worse**   c) Same as

In the table below you are given the study hours of 2 students. For each, compute the regression estimate for their GPA's, by filling out the chart below.

# Hours Studied	Study Z score	r	GPA Z score	GPA
14 hours	**31.** Z= __1__ a) -1 b) 0 **c) 1** d) 0.8 e) 1.8	$r = 0.8$	**32.** Z= __0.8__ a) -1 b) 0 c) 1 **d) 0.8** e) 1.8	**33.** GPA = __3.4__ **a) 3.4** b) 2.6 c) 3.8 d) 4.8 e) 0.8

$Z = \dfrac{val - avg}{SD}$

$Z = \dfrac{14 - 10}{4} = 1$

$val = avg + (SD)(Z) \rightarrow val = 3 + (.8)(.5)$

The regression equation for predicting GPA from hours studied is: GPA = __0.1__ x Study Hours + __2__

34. Fill in the first blank in the equation above with the slope.   a) 0.8   b) 6.4   c) 8   d) 0.125   **e) 0.1**

$$m = r \times \frac{SD_y}{SD_x} \rightarrow 0.8 \times \frac{0.5}{4} = 0.1$$

35. Fill in the second blank in the equation above with the y intercept.
   a) -9.2   b) 6.4   **c) 2**   d) 9.7   e) 3

$$y = 0.1x + b \rightarrow 3 = 0.1(10) + b \rightarrow 3 = 1 + b \rightarrow b = 2$$

plug in avgs.

In questions 34 and 35 above you were predicting GPA from study hours. Would the slope, and y-intercept change if you were predicting study hours from GPA instead?

36. The slope would…   **(a) change**   b) stay the same   c) cannot be determined
37. The y-intercept would…   **(a) change**   b) stay the same   c) cannot be determined

Questions 38-41 pertain to the table below: A set of exam scores follows a **normal distribution**. In the table below, you are given either the z score or the percentile for 4 students scores. Fill in the missing blanks.

Student	Exam Percentile	Exam z score
A	50th	Blank 1
B	16th	Blank 2
C	Blank 3	-1.75
D	Blank 4	2.5

38. What goes in Blank 1?   a) 50   **b) 0**   c) 1   d) -1   e) 0.2

50th %tile   z-score   always = 0

39. What goes in Blank 2?   a) 50   b) 0   c) 1   **d) -1**   e) 0.2
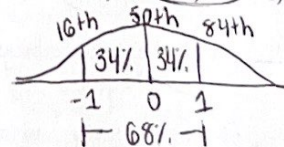
40. What goes in Blank 3?   a) 92   b) 8   **c) 4**   d) 96   e) 68

41. What goes in Blank 4?   a) 50   b) 80   c) 98.76   d) 1   **e) 99.5**

$\dfrac{100 - 92}{2} = 4\%$

$\dfrac{100 - 98.76}{2} = .62\%$

$.62 + 98.76 = \boxed{99.38\%}$

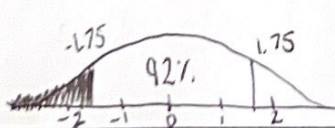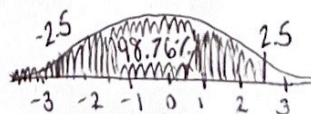**Questions 42-50 pertain to this situation:** The scatter plot below shows the percent of people living in poverty versus the number of people in prison per 100,000 for the 50 states in the US. **Vermont is on Line 2 and Mississippi is on Line 1.**

42. The average poverty level (in %) is closest to...
    a) 10   b) 13   c) 16   d) 18   e) 20

43. The average number of prisoners (per 100,000) is closet to....
    a) 10   b) 40   c) 60   d) 75   e) 100

44. Which is the regression line?   a) Line 1   b) Line 2

45. The correlation between poverty and prisoners is closest to
    a) 0   b) -0.4   c) 0.4   d) 0.8   e) -0.8

46. Which of the labeled states has a residual = 0?   *on reg. line*
    a) Vermont   b) Mississippi   c) Texas   d) Minnesota

47. Which state has the same z scores for poverty and prisoners?
    a) Vermont   b) Mississippi   c) Texas   d) Minnesota
                      on SD line

48. Which state has the largest prediction error?
    a) Vermont   b) Mississippi   c) Texas   d) Minnesota   *furthest from reg. line*

49. The average of all the residuals is?   a) 1   b) 100   c) always a different number   d) 0
                                    always 0

50. The above graph has 50 points, one for each **individual** state's poverty and prisoner level. If we divided the 50 states into 9 geographical regions and calculated the **average** poverty and prisoner 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? *Choose one:*
    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 the within region scatter would disappear and we'd just see the between region scatter.

Question 51: Fill in the table and plot the points. The **average** of X = 4 and the average of Y = 3. The **SD** of X and Y are both **2**.

X	Y	Z-score for X	Z-score for Y	Products
1	6	-1.5	$z = \frac{6-3}{2} = 1.5$	-2.25
3	4	$z = \frac{3-4}{2} = -0.5$	0.5	-0.25
4	2	0	$z = \frac{2-3}{2} = -0.5$	0
5	3	$z = \frac{5-4}{2} = 0.5$	0	0
7	0	1.5	-1.5	-2.25
Totals		Total *should* = 0	Total *should* = 0	Total = -4.75
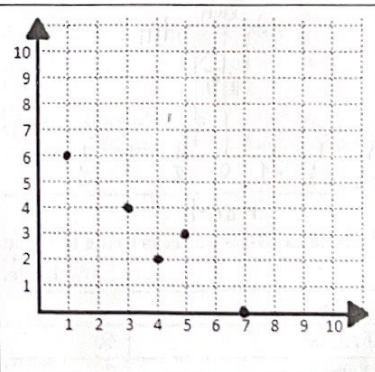
The correlation coefficient r is?   a) 0.5   b) 0.95   c) -4.75   d) -0.95   e) 0 $\frac{-4.75}{5} = -0.95$

Questions 52-54 pertain to the following situation: Suppose Verbal SAT and Math SAT scores among students who take both exams have the following summary statistics and the scatter plot is football shaped: **Verbal SAT: avg = 500, SD=100 Math SAT: avg = 500, SD=100 and r =0.8**

52. The regression equation when predicting Math scores from Verbal scores is: **Predicted Math = 0.8 * Verbal + 100.** Use the regression equation to predict the Math score of a student who got a 640 on the Verbal.
    a) 640   b) 350   c) 500   d) 580   e) 612

$y = 0.8(640) + 100 = 612$

4

53. What is the SD of the prediction errors (the RMSE) when predicting Math SAT scores from Verbal SAT scores?

    a) 10   **b) 60**   c) 68   d) 80   e) 100

$$RMSE: \sqrt{1-r^2} \cdot SD_y \rightarrow \sqrt{1-.8^2} \cdot 100 = 60$$

54. The regression equation predicts Math SAT scores of 500 for those who score 500 on the Verbal SAT. Of course not all will get exactly 500. Instead there's a range of scores, with about **95%** of them scoring between …

    a) 300 and 700   b) 400 and 600   **c) 380 and 620**   d) 340 and 660

$$\text{prediction} \pm 2(RMSE) \rightarrow 500 \pm 2(60) = (380, 620)$$

Questions 55-60 pertain to the following situation: Suppose blood pressure and temperature follow the normal curve but have different correlations among different populations.

55. Imagine a population where there is a perfect positive correlation (r = 1) between the 2 measurements. If someone's blood pressure is in the 70th percentile then the regression estimate for his temperature would be the _____ percentile.

    a) 15th b) 30th c) 50th **d) 70th**   e) impossible to tell $r=1 \rightarrow$ exactly equal

56. Imagine a population where there is a perfect negative correlation (r = -1) between the 2 measurements. If someone's blood pressure is in the 70th percentile then the regression estimate for his temperature would be the _____ percentile.

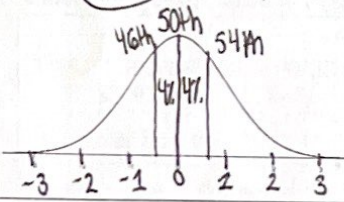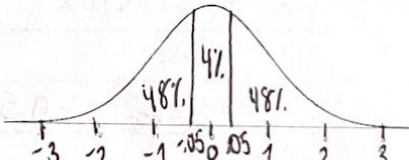   a) 15th **b) 30th**   c) 50th d) 70th e) impossible to tell $r=-1 \rightarrow$ add to 100

57. Imagine a population where there is a zero correlation (r = 0) between the 2 measurements. If someone's blood pressure is in the 70th percentile then the regression estimate for his temperature would be the _____ percentile.

    a) 15th b) 30th **c) 50th**   d) 70th e) impossible to tell $r=0 \rightarrow$ always 50th

If someone's blood pressure is in the **54th** percentile where r = 0.5, estimate his temperature percentile *by filling in the table below.*

Blood Pressure Percentile	Blood Pressure Z	r	Temperature Z	Temperature Percentile
Steve is in the 54th percentile for blood pressure. **58.** What middle area does that correspond to? a) 4% **b) 8%** c) 23% d) 46% e) 54% 	**59.** What is the z-score? a) 0.05 b) 0.3 **c) 0.10** d) 0.75 e) 1	$r = 0.5 \rightarrow Z = 0.05$		**60.** Temperature Percentile = **52nd** a) 69th b) 4th **c) 52nd** d) 48th e) 72nd $48\% + 4\% = 52\%$

Questions 61-68 pertains to the table below that shows our survey responses for gender and handedness.

	Left-Handed	Ambidextrous	Right-Handed	Totals
Male	30	20	266	316
Female	60	27	560	647
Totals	90	47	826	963

Suppose you randomly draw from these students.

61. What is the chance of getting a female?

    a) 647/963   b) 90/963   c) 316/963   d) 737/963   e) 826/963

62. What is the chance of getting someone who is left-handed?

    a) 647/963   **b) 90/963**   c) 60/963   d) 30/963   e) 826/963

63. What is the chance of drawing once and getting a female **or** someone who is left-handed?

    a) 60/963   b) 737/963   **c) 677/963**   d) 30/963   e) 406/826

$$\frac{647}{963} + \frac{90}{963} - \frac{60}{963} = \frac{677}{963}$$

64. What is the chance you'll get a female if you draw only from the left-handers?
    a) 90/316   b) 30/90   c) 30/316   d) 60/90   e) 60/647

65. What is the chance you'll get a left-hander if you draw only from the females?
    a) 90/316   b) 30/90   c) 30/316   d) 60/90   e) 60/647

66. Draw 3 students *without* replacement. What is the chance that *all* 3 students are left-handed?
    a) $(873/963)^3$ b) $1-(873/963)^3$ c) $90/963*89/962*88/961$ d) $1-(90/963)^3$ e) $1- 90/963*89/962*88/961$

67. Draw 3 students *without* replacement. What's the chance that *not all* 3 students are left-handed?
    a) $(873/963)^3$ b) $1-(873/963)^3$ c) $90/963*89/962*88/961$ d) $1-(90/963)^3$ e) $1- 90/963*89/962*88/961$

$$P(not\ all) = 1 - P(all)$$

68. Draw 3 students *with* replacement. What is the chance that *at least one* student is right-handed?
    a) $(826/963)^3$ b) $1-(826/963)^3$ c) $1-(137/963)^3$ d) $1-(90/963)^3$ e) $90/963*89/962*88/961$

$$P(at\ least\ one) = 1 - P(none)$$

Questions 69-74 pertains to rolling fair dice.

69. Two dice are rolled. What is the chance that the sum of the spots is 9?
    a) 2/36   b) 3/36   c) 4/36   d) 5/36   e) 1/6*1/6

    $(5,4)(4,5)$
    $(6,3)(3,6)$

70. Two dice are rolled. What is the chance that the sum of the spots is 9 **or** 10?
    a) 2/36   b) 3/36   c) 4/36   d) 5/36   e) 7/36

71. Two dice are rolled what is the chance the sum of the spots is either 6 **or** doubles (doubles is (1,1), (2,2), (3,3) etc.)?
    a) 6/36   b) 7/36   c) 8/36   d) 9/36   e) 10/36

72. What is the chance of rolling a die 7 times and getting no "3"s?
    a) $(1/6)^7$ b) $1-(1/6)^7$ c) $1-(5/6)^7$ d) $7*(1/6)$ e) $(5/6)^7$

73. What is the chance of rolling a die 7 times and getting all "3"s?
    a) $(5/6)^7$ b) $(1/6)^7$ c) $1-(5/6)^7$ d) $1-(1/6)^7$ e) $7*(1/6)$

74. What is the chance of rolling a die 7 times and getting at least one "3"?
    a) $(5/6)^7$ b) $1-(1/6)^7$ c) $1-(5/6)^7$ d) $7*(1/6)$ e) $(1/6)^7$

$$P(at\ least\ one) = 1 - P(none)$$

Questions 75 and 76 pertain to tossing fair coins.

75. What is the chance of tossing a fair coin 3 times and getting all tails?
    a) ½ + ½ + ½   b) ½ * ½ * ½   c) 1- (½ * ½ * ½)   d) 3/6

76. What is the chance of tossing a fair coin 3 times and getting this particular sequence: HHT?
    a) ½ + ½ + ½   b) ½ * ½ * ½   c) 1- (½ * ½ * ½)   d) 3/6

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)
z=(value-average)/SD