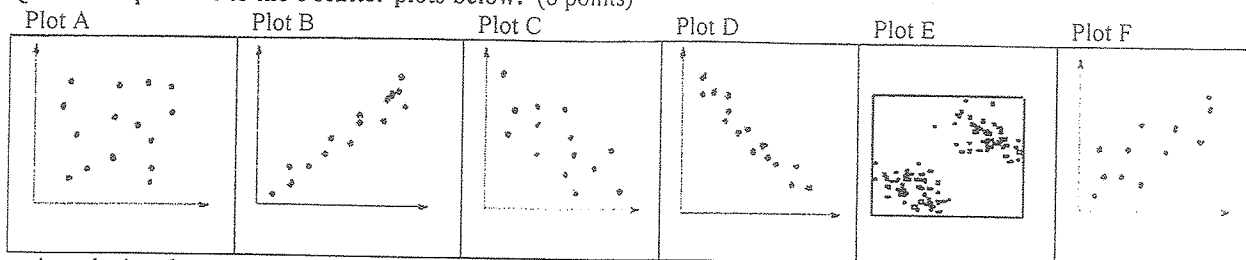


Question 1 pertains to the 6 scatter plots below: (6 points)



Determine whether the correlation coefficient is appropriate for analyzing the plots. If so, choose the r which best represents the plot. Check only one box per row! READ THIS- Each column will be used exactly once!

	$r = 0$	$r = +0.8$	$r = +0.3$	$r = -0.3$	$r = -0.8$	Not appropriate to use r
Plot A	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Plot B	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Plot C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Plot D	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Plot E	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Plot F	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 2 (5 points)

For each of the following pairs of variables, check the box that best describes its correlation coefficient r .

Check only one box per row! READ THIS- Each column will be used exactly once!

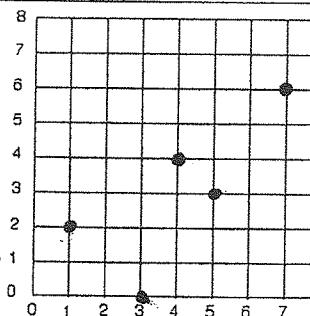
	Exactly +1	Between 0 and +1	About 0	Between 0 and -1	Exactly -1
Amount of exercise and percent body fat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Ice cream sales and temperature	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Weight in pounds and weight in kilograms	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Number of siblings you have and GPA	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Two numbers that always add up to 1000.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Question 3 (18 points)

Part A: Compute the correlation coefficient (r) by filling in the table below. Then fill in the totals row.The averages of $X=4$ and SD of $X=2$. The average of $Y=3$ and SD of $Y=2$. (13 points) Plot the points on the graph below.

Chart (10.5 points)

X	Y	Z-score for X	Z-score for Y	Products
1	2	$\frac{1-4}{2} = -1.5$	$\frac{2-3}{2} = -0.5$	$(-1.5)(-0.5) = 0.75$
3	0	$\frac{3-4}{2} = -0.5$	$\frac{0-3}{2} = -1.5$	$(-0.5)(-1.5) = 0.75$
4	4	$\frac{4-4}{2} = 0$	$\frac{4-3}{2} = 0.5$	$(0)(0.5) = 0$
5	3	$\frac{5-4}{2} = 0.5$	$\frac{3-3}{2} = 0$	$(0.5)(0) = 0$
7	6	$\frac{7-4}{2} = 1.5$	$\frac{6-3}{2} = 1.5$	$(1.5)(1.5) = 2.25$
Totals		Total should = 0	Total should = 0	Total = 3.75



Part B: From the previous page, find the correlation coefficient r . Round to 2 decimal places.

$$r = 0.75 \quad (1 \text{ point})$$

$$\frac{3.75}{5} = 0.75$$

Part C: What happens to the correlation coefficient r if ... (Note: X and Y below refer to the original X and Y values given in part A) (Use the r you calculated in Part B to answer Part C. If you didn't answer Part B, then use $r = 0.3$ to answer Part C) (4 points)

i) We add 5 to each Y value. $r = 0.75$

ii) We multiply each X value by 2. $r = 0.75$

iii) We divide each X value by -4. $r = -0.75$

iv) We switch all X and Y values. $r = 0.75$

Question 4: In the very first survey of the year, we asked students "How old is your mother?" and "How old is your father?" The scatter plot of mothers' and fathers' ages is shown below along with the summary statistics. The point where the SD line and the regression line intersect is marked in black. (12 points)

- a) The SD Line and the Regression Line are shown. Which is the SD Line? (2 points)

Choose One: ☒ i) Line A ☐ ii) Line B

- b) The correlation is closest to ... (2 points)

Choose One: i) 0.2 ii) 0.4 ☒ iii) 0.8 iv) -0.2 v) -0.6

- c) About what is the average Mother's Age? (1 point)

Choose One: i) 40 ☒ ii) 50 iii) 55 iv) 60 v) 65

- d) About what is the average Father's Age? (1 point)

Choose One: i) 40 ☒ ii) 50 iii) 58 iv) 60 v) 65

- e) We think that the point circled on the scatter plot is an outlier. What does removing this outlier do to the correlation coefficient? (2 points)

i) Outliers have no effect on the correlation coefficient.
☒ ii) Removing this outlier raises the correlation coefficient.
 iii) Removing this outlier lowers the correlation coefficient.

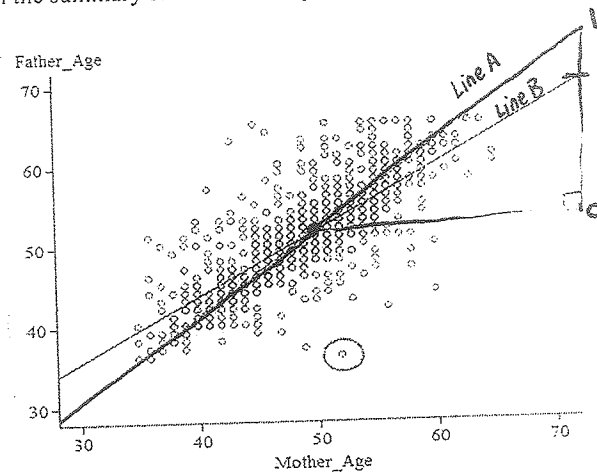
- f) Points above the regression line have prediction errors (residuals) that are: (1 point)

Choose One: i) negative ☒ ii) positive iii) equal to zero

- g) The average of all the prediction errors is always 0. Fill in the blank with a number. (1 point)

- h) Which must be true if you lie on the SD line? (2 points) Choose One:

☒ i) You have the same z-scores for mother age and father age
 ii) Your prediction error is equal to 0.
 iii) You are exactly average.



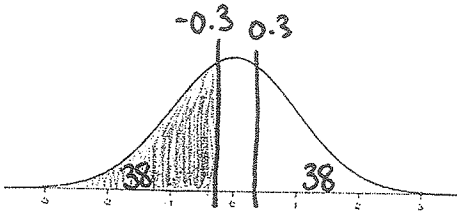
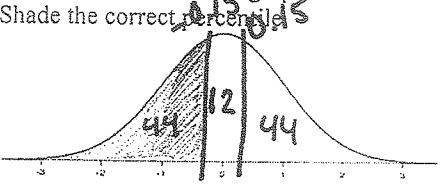
Question 5 (10 points)

Suppose people's scores on a math skills test and a social skills test follow the normal curve, but have different correlations in different countries.

- a) Consider 5 countries where the correlation coefficients between people's math and social skills are given in the table below. If a child is in the 25th percentile in math, estimate her percentile in social skills in each country. (5 points)

Math Skills Percentile	r	Social Skills Percentile						
25 th	-1	Choose One:	10 th	25 th	37 th	50 th	63 rd	<input checked="" type="radio"/> 75 th
25 th	1	Choose One:	10 th	<input checked="" type="radio"/> 25 th	37 th	50 th	63 rd	75 th
25 th	0	Choose One:	10 th	25 th	37 th	<input checked="" type="radio"/> 50 th	63 rd	75 th
25 th	-0.5	Choose One:	10 th	25 th	37 th	50 th	<input checked="" type="radio"/> 63 rd	75 th
25 th	0.5	Choose One:	10 th	25 th	<input checked="" type="radio"/> 37 th	50 th	63 rd	75 th

b) If someone is in the 38th percentile in math where $r = 0.5$, what percentile would you estimate for his social skills? Solve by filling in the table below. (You may round areas and z-scores to fit the nearest line on the table). (5 points)

Math Skills Percentile	Math Z	$r = 0.5$	Social Z	Social Skills Percentile
38 th Percentile What middle area on the table should you look up to find the Z score? <u>24</u> % (1 point)  Correctly mark the z-score and shade the area corresponding to the 38 th percentile. (1/2 point for shading correctly)	$Z = -0.3$ (1 point) negative	$r = 0.5$	$Z = -1.5$ (1 point)	Social Skills Percentile = <u>44</u> (1 point) Mark the Z score on the graph below. Shade the correct percentile.  Round the middle area given in the Normal Table to the nearest WHOLE number and then calculate the percentile. (1/2 point for shading correctly)

Question 6 (10 points)

This question pertains to the survey questions: "What is the fastest you've ever driven in mph?" and "What is the most you have ever drank in a 24 hour period?" To the right are the 5 summary statistics.

	Average	SD
Speed	96mph	20mph
Drinks	13	10

Correlation: $r = 0.4$

a) Make regression estimates for Students A and B by filling in the table below. For student A, you're given speed and asked to predict drinks using the 3 step method. For student B, you're given drinks and asked to predict speed using the 3 step method.

Speed (in mph)	Speed Z-score	r	Drinks Z-score	Number of Drinks
Student A: 106 mph $Z = \frac{106-96}{20} = 0.5$ (2 points) Show all work for full credit.	$Z = 0.5$ (2 points)	$r = 0.4$	$Z = 0.2$ (1 point)	Drinks = <u>15</u> (2 points) value = $13 + (0.2)(10) = 15$
<u>112</u> mph (2 points) value = $96 + 0.8(20) = 112$ Show all work for full credit.	$Z = 0.8$ (1 point)	$r = 0.4$	$Z = 2$ (2 points)	Student B: 33 Drinks $Z = \frac{33-13}{10} = 2$

Question 7 (2 points)

In 2010, millions of students nationwide took the math and verbal SAT tests. Within each state, the average verbal and the average math SAT scores were calculated. The correlation between these 50 pairs of averages was 0.97. Does that mean that the correlation between the millions of individual students' verbal and math SAT scores would also be 0.97?

Choose one:

- i) Yes, since the state averages are computed from the individual scores, the correlation for individual verbal and math scores must also be 0.97.
- ☒ ii) No, the correlation for individuals is likely to be lower than 0.97 since the individuals within each state would add more scatter so the points wouldn't follow a straight line so closely.
- iii) No, the correlation for individuals is likely to be higher than 0.97 since it's based on millions of data points instead of just 50.

Question 8 (2 points)

When are the regression line and the SD line the same?

Choose one:

i) When $r=0$

ii) Never

☒ iii) When $r=1$

iv) Need more information to answer

Question 9 (12 points)

A large study gave the following results on the heights of fathers and the heights of their daughters. Assume the scatter plot football-shaped:

	Average	SD
Father's Height	69"	2.5"
Daughter's Height	64"	1.5"

is

Correlation: $r = 0.7$

- a) Find the regression equation for predicting fathers' heights from daughters' heights. $Y = 1.17X + -5.88$

i) First calculate the slope. Show work below. Round to 2 decimal places! (2 points)

$$m = r \times \frac{SD_y}{SD_x} = 0.7 \times \frac{2.5}{1.5} = 1.17 \quad \text{accept } 1.17 \text{ or } 1.16$$

ii) Next, calculate the y-intercept. Show work below. Round to 2 decimal places! (2 points)

$$y = 1.17x + b$$

$$69 = 1.17(64) + b \quad b = -5.88 \quad \text{accept answers between } -5.24 \text{ and } -6$$

- b) Using the regression equation from part a, predict the height of a father whose daughter is 62 inches tall. Round your answer to the nearest whole number. (2 points)

$$y = 1.17(62) - 5.88 = 67 \quad \text{make sure it makes sense with part a}$$

- c) Suppose the father whose daughter is 62 inches tall is actually 70 inches tall. What is his *residual*? (2 points) Show work below:

$$\text{residual} = \text{actual} - \text{predicted} = 70 - 67 = 3 \quad \text{make sure it makes sense w/ part b}$$

- d) What is the SD of the prediction errors (the RMSE) when predicting fathers' heights from daughters' heights? Choose one: (2 points)

i) 2.5 ii) 1.5 iii) $\sqrt{1-0.7^2} \times 1.5$ iv) $\sqrt{1-0.7^2} \times 2.5$ v) 0

- e) Of course, the prediction we made in part b isn't going to be exactly correct. Instead there's a range of father's ages, with about 95% of them falling within ... Choose one: (2 points)

i) 1 SD errors ii) 2 SD errors iii) 3 SD errors iv) Not enough info

Question 10 pertains to a roadside drunk driving test. Suppose only 10% of those who get stopped and tested for drunk driving are actually drunk. 70% of the drunk drivers will correctly fail the test. But 30% of the sober drivers will also fail the test (due to nervousness or other reasons). Fill in the table for 100 drivers total who get stopped and tested. (6 points)

	Fail Test	Pass Test	Total
Drunk Drivers	7 (1 point)	3	10 (1 point)
Sober Drivers	27 (1 point)	63	90
Total	34	66	100

Check that rows and columns sum correctly. (1 point)

- a) If a driver fails the test, what's the chance he or she is actually sober? $\frac{27}{34}$ (leave answer as a fraction) (1 point)
- b) If a driver passes the test, what is the chance that he or she is actually drunk? $\frac{3}{66}$ (leave answer as a fraction) (1 point)

Question 11 pertains to the table below which shows the survey responses of 389 students who identified themselves as "white" & the 201 students who identified themselves as "non-white" to the question "Who do you plan to vote for in the 2012 election?"

	Obama	Romney	Undecided/Other	Won't Vote	Totals
White	207	82	77	23	389
Non-White	124	9	25	43	201
Totals	331	91	102	66	590

Suppose you randomly draw a student from those who answered this survey: (6 points)

- a) What is the chance that you'll get a student voting for Romney?
 i) $422/590$ ii) $9/91$ iii) $82/389$ iv) $9/201$ v) $82/91$ vi) $91/590$ vii) $91/201$
- b) What is the chance you'll get a student voting for Romney if you draw only from the non-white students?
 i) $91/389$ ii) $91/590$ iii) $82/389$ iv) $9/201$ v) $82/91$ vi) $9/91$ vii) $91/201$
- c) What is the chance that you'll get a student voting for Romney if you draw only from white students?
 i) $91/389$ ii) $91/590$ iii) $82/389$ iv) $9/201$ v) $82/91$ vi) $9/91$ vii) $91/201$
- d) What is the chance of getting a white student if you draw only from Romney voters?
 i) $91/389$ ii) $91/590$ iii) $82/389$ iv) $9/201$ v) $82/91$ vi) $9/91$ vii) $91/201$
- e) What is the chance that a randomly selected student is either voting for Obama or white?
 i) $331/590 + 389/590 - 207/590$ ii) $331/590 * 389/590$ iii) $331/590 + 207/590$ iv) $331/590 + 389/590$
- f) Draw 3 students *with replacement*. What is the chance that at least one of them is a Romney supporter?
 i) $331/590$ ii) $331/590 + 33/590$ iii) $(331/590)^3$ iv) $1 - (499/590)^3$ v) $1 - (91/590)^3$

Question 12 pertains to a well-shuffled deck of 52 cards. (4 points)

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.

- a) Draw 2 cards without replacement. What is the chance that they are both diamonds?
 i) $4/52 + 4/52$ ii) $13/52 * 12/51$ iii) $16/52$ iv) $13/52 * 13/52$ v) $4/52 * 4/51$ vi) $4/52 * 3/51$
- b) Draw 2 cards with replacement. What is the chance that the both cards are spades?
 i) $4/52 + 4/52$ ii) $13/52 * 12/51$ iii) $16/52$ iv) $13/52 * 13/52$ v) $4/52 * 4/51$ vi) $4/52 * 3/51$
- c) Draw *one* card at random. What is the chance that it's *either* a Jack or an Ace?
 i) $4/52 + 4/52$ ii) $13/52 * 12/51$ iii) $17/52$ iv) $13/52 * 13/52$ v) $4/52 + 4/52 - 1/52$ vi) $5/52 * 3/51$
- d) Draw *one* card at random. What is the chance that it's *either* a Queen or a Club?
 i) $4/52 + 13/52 - 1/52$ ii) $4/52 + 13/52$ iii) $4/52 * 3/51$ iv) $4/52 * 13/51$ v) $4/52 * 13/52$ vi) $4/52 + 4/52 - 1/52$

Question 13 pertains to rolling fair dice. (7 points)

- a) Two dice are rolled. What is the chance that the sum of the spots is 3?
 i) $1/36$ ii) $2/36$ iii) $3/36$ iv) $4/36$ v) $1/6 * 1/6$ vi) $1/6 + 1/6$
- b) One die is rolled 4 times. What is the chance of getting *no* 2's?
 i) $(5/6)^4$ ii) $(1/6)^4$ iii) $1 - (5/6)^4$ iv) $1 - (1/6)^4$ v) $4/6$
- c) One die is rolled 4 times. What is the chance of *at least one* 2?
 i) $(5/6)^4$ ii) $(1/6)^4$ iii) $1 - (5/6)^4$ iv) $1 - (1/6)^4$ v) $4/6$
- d) One die is rolled 3 times. What is the chance of getting *all* 2's?
 i) $(5/6)^3$ ii) $3/6$ iii) $1 - (5/6)^3$ iv) $1 - (1/6)^3$ v) $(1/6)^3$
- e) One die is rolled 3 times. What is the chance of getting *not all* 2's?
 i) $(5/6)^3$ ii) $(1/6)^3$ iii) $1 - (5/6)^3$ iv) $1 - (1/6)^3$ v) $3/6$
- f) One die is rolled twice. What is the chance that the first roll is a 4 *or* the second roll is a 5?
 i) $1/6 * 1/6$ ii) $1/6 + 1/6$ iii) $4/6 + 5/6$ v) $1/6 + 1/6 - 1/36$
- g) One die is rolled twice. What is the chance that the first roll is a 4 *and* the second roll is a 5?
 i) $1/6 * 1/6$ ii) $1/6 + 1/6$ iii) $4/6 + 5/6$ v) $1/6 + 1/6 - 1/36$
- h) A die is rolled once. What is the chance of getting either a 4 or a 5?
 i) $1/6 * 1/6$ ii) $1/6 + 1/6$ iii) $4/6 + 5/6$ v) $1/6 + 1/6 - 1/36$