

# Statistics Page 1

Statistics 100 Exam 2

Oct. 14<sup>th</sup>, 2015

**Question 1: Follow the steps below and draw a boxplot for the data given. (10 points total)**

Here is the data from 9 random students this semester who answered the survey question: "How many hours a day do you spend on social media?": 0, 3, 3, 4, 5, 6, 6, 7, 12

**Step 1:** Find the median, Q1, Q3, and the IQR.  
Circle your answers.

(1 point) Median = 5

(1 point) Q1 = 3

(1 point) Q3 = 6.5

(1 point) IQR = 3.5

**Step 2:** Check for outliers. Fill in the blanks with yes or no.

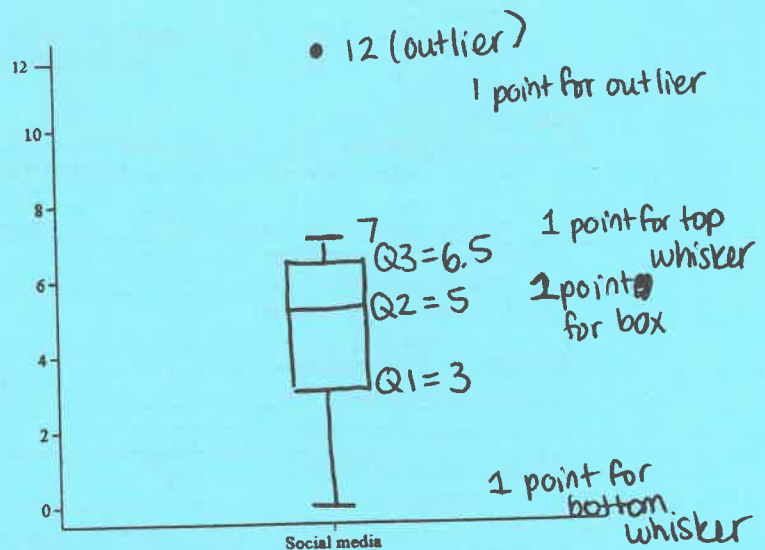
(1 point) Are there low outliers? No  
Yes/No

$$Q1 - 1.5(IQR) = -2.25$$

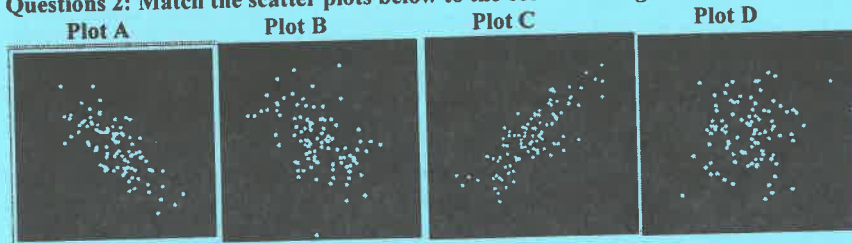
(1 point) Are there high outliers? Yes  
Yes/No

$$Q3 + 1.5(IQR) = 11.75$$

**Step 3:** Draw the boxplot on the axis below. (4 points)



**Questions 2: Match the scatter plots below to the correlations given in the table. (4 points)**



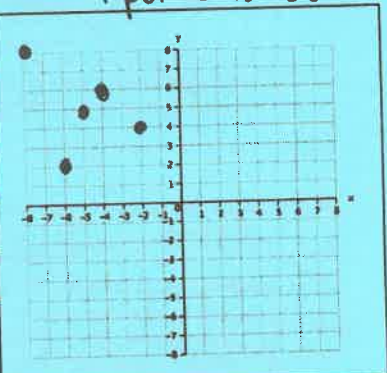
r	Write the letter of the plot that matches r
0.76	<u>C</u>
-0.51	<u>B</u>
0.22	<u>D</u>
-0.87	<u>A</u>

**Question 3: Part a)** Compute the correlation coefficient (r) by filling in the table below including the totals.

The average of X = -5 and the SD of X = 2. The average of Y = 5 and the SD of Y = 2.

Plot the points on the graph below to create a scatterplot. (13 points total)

X	Y	Z-score for X	Z-score for Y	Products
-2	4	1.5	-0.5	-0.75
-4	6	0.5	0.5	0.25
-5	5	0	0	0
-6	2	-0.5	-1.5	0.75
-8	8	-1.5	1.5	-2.25
Totals		Total should = <u>0</u> 1 point	Total should = <u>0</u> 1 point	Total = <u>-2</u> 1 point



1 point for each row

4 points for scatterplot

(-2 if in wrong quadrant)

**Part b)** (1 point) The correlation coefficient  $r = \underline{-0.4}$  (Fill in the blank and round to 2 decimal places)

1 of 7 pages (16 problems)

# Statistics Page 2

Statistics 100 Exam 2

Oct. 14<sup>th</sup>, 2015

**Question 4:** X and Y are 2 sets of numbers with a correlation coefficient of  $r=0.75$ . Circle your answers below.

- a) (1 point) If we multiply all the X values by -3, r will be...  
 i) 0.75      ii) -0.75      iii) impossible to tell
- b) (1 point) If we add 0.2 to all of the Y values, r will be...  
 i) 0.75      ii) -0.75      iii) impossible to tell
- c) (1 point) If we switch all the X values and Y values, r will be...  
 i) 0.75      ii) -0.75      iii) impossible to tell

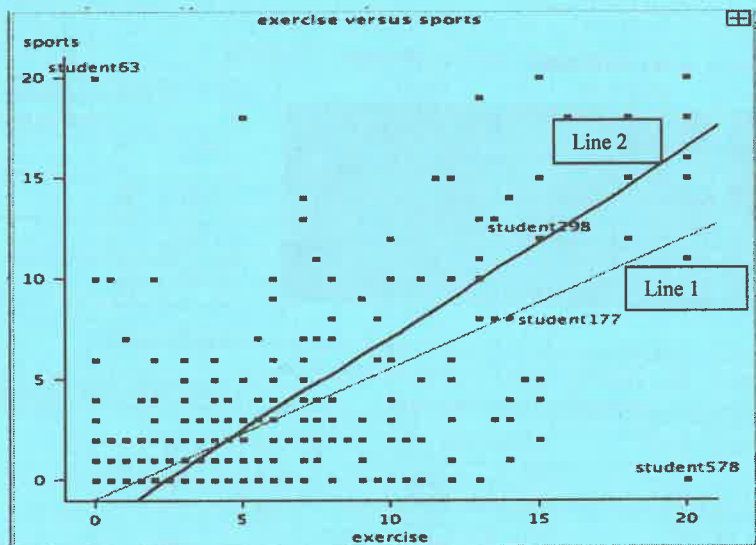
**Question 5** (5 points total)

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
As X gets larger, corresponding Y values get smaller	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
X and Y are paired by a random lottery	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Number of days you attend Stat 100 and your score on the final	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Y is always equal to -1 times X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Height in inches and height in centimeters	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Question 6:** This question pertains to the scatter plot below that depicts the survey responses of 764 Stat 100 students to the 2 questions: "How many hours per week do you typically exercise?" and "How many hours per week do you typically spend playing sports?" (7 points total)



- a) (1 point) Look at Line 1 and Line 2 on the scatter plot above. One is the SD line and one is the regression line. Which is the regression line?  
**Choose one:**  
 i) Line 1      ii) Line 2
- b) (1 point) The correlation (r) between exercise and sports is closest to ...  
**Choose one:**  
 i) 0.7      ii) -0.7  
 iii) -0.3      iv) 0.3

- c) (1 point) The average number of hours people said they played sports (y value) is closest to:  
 i) 0 hours      ii) 2 hours      iii) 5 hours      iv) 10 hours

d) (4 points) Look at students 63, 177, 298 and 578 on the graph. Each question below has one correct answer.

- i) Which student has exactly the same Z scores for exercise and sports?      a) 63      b) 177      c) 298      d) 578  
 ii) Which student has a prediction error of 0?      a) 63      b) 177      c) 298      d) 578  
 iii) Which student has a prediction error of -12?      a) 63      b) 177      c) 298      d) 578  
 iv) Which student has a prediction error of 21?      a) 63      b) 177      c) 298      d) 578

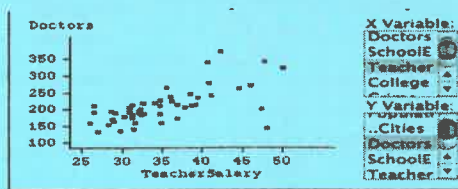


# Statistics Page 3

Statistics 100 Exam 2

Oct. 14<sup>th</sup>, 2015

**Question 7:** The following scatter plots show the relation between teacher salaries and number of doctors by state and by geographical region in 1995. The graph on the left has 50 points, one for each individual state's teacher salary and doctor level. The graph on the right has the same information condensed into 9 points, one for each of 9 geographical regions. (In other words, the 50 states were divided in 9 regions and the average teacher salary and average number of doctors was computed for each region.)



(2 points) Which statement is best?

- i) The two graphs must have the same correlation coefficient since they're based on the same exact data.
- ii) The graph of the 50 states contains more information, giving it a higher correlation coefficient.
- iii) The graph of the 9 regions eliminates within region scatter, giving it a higher correlation coefficient.

**Question 8:** The table below gives the 5 summary statistics of the 300 random US adults who responded to the questions: "How happy are you on a scale of 1-10, 1 being very sad and 10 being very happy?" and "How many hours per week do you spend working at your job?" (10 points total)

	Average	SD
Happiness	5	2
Work Hours	40	10

Correlation Coefficient ( $r$ ) = 0.6

a) Make regression estimates for Philip and Elizabeth by filling in the blanks in the table below. For Philip, you're given work hours per week and asked to predict happiness. For Elizabeth you're given happiness rating and are asked to predict work hours per week. Round your answers to two decimal places.

Happiness Rating	Happiness z-score	$r$	Work Hours z-score	Work Hours per Week
5.84 Happiness Rating (1 point) Show work for full credit.	$Z = 0.42$ (1 point)	$r = 0.6$	$Z = 0.7$ (1 point)	Philip: 47 Hours
Elizabeth: 2 Happiness Rating	$Z = -1.5$ (1 point)	$r = 0.6$	$Z = -0.9$ (1 point)	31 Hours (1 point) Show work for full credit

b) (1 point) If we drew a scatterplot for the 2 variables above, the average of the prediction errors (residuals) would equal? 0 (Fill in the blank above with a number)

c) (1 point) Say we made a prediction of work hours per week for another person, Paige. We predict that Page will work 45 hours per week. Knowing that her prediction error is 3, how many hours per week did she actually work? Show work below and fill in the blank with a number.

Paige worked 48 hours per week.

$$3 = \text{actual} - 45$$

d) (1 point) Suppose another person, Henry, has a negative prediction error. This means that his actual value is \_\_\_\_\_ than his predicted value.

- i) bigger than
- ii) less than
- c) equal to

e) (1 point) Your predicted value always ends up being closer to the average than the value that you start with. Because the correlation is between 0 and 1, there is a drift towards the average due to the football shaped cloud of points. This phenomenon is called:

- i) Regression Fallacy
- ii) Regression Effect
- iii) Confounders
- iv) Bayes Rule

# Statistics Page 4

Statistics 100 Exam 2

Oct. 14<sup>th</sup>, 2015

## Question 9 (5 points total)

There are a total of 50 Faculty members in the English and Math Departments at Parkland and 60% of them are male. Only 5% of the females teach math and only 40% of the males teach English. Fill in the chart below.

	Math	English	Total
Female	1 (1 point)	19	20 (1 point)
Male	18 (1 point)	12	30
Total	19	31	50

- a) (1 point) Given that a person teaches math, what's the chance that the person is male?  $\frac{18}{19}$  continued error
- b) (1 point) Suppose a person is female, what's the chance that the person teaches English?  $\frac{19}{20}$

## Question 10 (16 points total)

The two variables below are negatively correlated: Amount of time spent exercising per day (in min) and weight. The scatter plot was roughly football shaped. Below are the 5 summary statistics.

	Average	SD
Time Spent Exercising (minutes)	30	8
Weight	154	5

Correlation Coefficient ( $r$ ) = -0.4

- a) Find the regression equation for predicting Weight from Minutes Spent Exercising.  $Y = -0.25x + 161.5$

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

$$m = r \times \frac{SD_y}{SD_x} = -0.4 \times \frac{5}{8} = -0.25 \quad \text{-1 for switching x + y or forgetting negative}$$

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

$$y = -0.25x + b$$

$$154 = -0.25(30) + b \quad b = 161.5 \quad \text{continued error}$$

- b) (2 points) Using the regression equation from part a, predict the weight of someone who spends 20 min exercising per day. Round your answer to the nearest whole number.

$$y = -0.25(20) + 161.5 = 156.5 \quad \text{continued error}$$

- c) (2 points) Suppose the person above actually weighs 150 pounds. What is his residual? Show work below:

$$\text{Residual} = \text{Actual} - \text{Predicted}$$

$$= 150 - 156.5 = -6.5 \quad \text{accept +156 or 157} \quad \text{continued error}$$

- d) (2 points) What is the SD of the prediction errors (the RMSE) when predicting weight from minutes spent exercising?

Show work below and round your answer to 2 decimal places.

$$RMSE = \sqrt{1 - (-0.4)^2} \times 5 = 4.58$$

One point for formula

4 of 7 pages (16 problems)



# Statistics Page 5

Statistics 100 Exam 2

Oct. 14<sup>th</sup>, 2015

- e) (1 point) Of course, the prediction we made in part b isn't going to be exactly correct. Instead there's a **range** of weights, with about **85%** of them falling within ...

Choose one:

i) 1 SDerrors

ii) 2 SDerrors

iii) 1.45 SDerrors

iv) 2.3 SDerrors

- f) (1 point) The regression equation for exercise time from weight is:

Exercise =  $-0.16 * (\text{weight}) + 54.64$ . How would you interpret the slope for this equation?

Choose one:

i) For each extra pound of weight, people exercise about 0.16 minutes less on the average.

ii) If you exercise 0 minutes per week, on average you will weigh about 54.64 pounds.

iii) For each extra pound of weight, people exercise about 54.64 minutes more on average.

iv) If you weigh zero pounds, you will exercise -0.16 minutes on average.

- g) (1 point) Say another person, Chelsea, has a z-score of -2 for weight and a z-score of -2 for minutes spent exercising. Which line does she fall on?

Choose one:

i) SD line

ii) Regression Line

iii) Both

iv) Neither

**Question 11:** (1 point) In a different dataset that contained heights and weights, one student was very tall and skinny. His prediction error was huge, the largest one implying he was an outlier. When I saw his point on the scatter plot, I decided to remove it. Before I removed the point  $r = 0.4$ . After I removed the point,  $r$  would

i) Be the same

ii) increase

iii) decrease

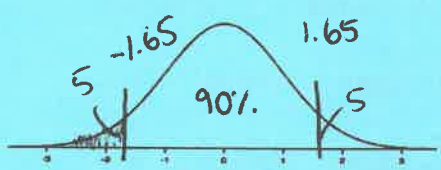
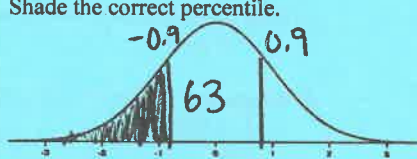
**Question 12** (10 points total)

Suppose scores on Spatial and Social Skills tests follow the normal curve, but have different correlations among different populations.

- a) Consider 5 populations where the correlation coefficients between the spatial and social skills are as given in the table below. If someone is in the **30th** percentile in spatial skills, estimate his social skills percentile in each population.

Percentile in Spatial Skills	r	Percentile in Social Skills							
30 <sup>th</sup>	-1	Choose One:	25 <sup>th</sup>	30 <sup>th</sup>	38 <sup>th</sup>	50 <sup>th</sup>	62 <sup>nd</sup>	70 <sup>th</sup>	75 <sup>th</sup>
30 <sup>th</sup>	1	Choose One:	25 <sup>th</sup>	30 <sup>th</sup>	38 <sup>th</sup>	50 <sup>th</sup>	62 <sup>nd</sup>	70 <sup>th</sup>	75 <sup>th</sup>
30 <sup>th</sup>	0.6	Choose One:	25 <sup>th</sup>	30 <sup>th</sup>	38 <sup>th</sup>	50 <sup>th</sup>	62 <sup>nd</sup>	70 <sup>th</sup>	75 <sup>th</sup>
30 <sup>th</sup>	0	Choose One:	25 <sup>th</sup>	30 <sup>th</sup>	38 <sup>th</sup>	50 <sup>th</sup>	62 <sup>nd</sup>	70 <sup>th</sup>	75 <sup>th</sup>
30 <sup>th</sup>	-0.6	Choose One:	25 <sup>th</sup>	30 <sup>th</sup>	38 <sup>th</sup>	50 <sup>th</sup>	62 <sup>nd</sup>	70 <sup>th</sup>	75 <sup>th</sup>

- b) If someone is in the **5th** percentile for spatial skills where  $r = 0.55$ , 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.)

Percentile in Spatial Skills	Spatial Z	R=0.55	Social Z	Percentile in Social Skills
<p>5<sup>th</sup> Percentile <i>if they put 5, no points</i></p> <p>What middle area on the table should you look up to find the Z score? <u>90</u> % (1 point)</p>  <p>Correctly mark the z-score and shade the area corresponding to the 5th percentile. (1/2 point for shading correctly)</p>	<p><math>Z = -1.65</math> <math>r = 0.55</math></p> <p>(1 point)</p>		<p><math>Z = -0.9</math></p> <p>(1 point)</p> <p>Round your z-score to the nearest tenth.</p>	<p>Social Skills Percentile = <u>18.5<sup>th</sup></u></p> <p>(1 point)</p> <p><math>\frac{100-63}{2} = 18.5</math></p> <p>Mark the Z score on the graph below. Shade the correct percentile.</p>  <p>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)</p>

# Statistics Page 6

Statistics 100 Exam 2

Oct. 14<sup>th</sup>, 2015

**Question 13** pertains to the table below which shows our survey responses for gender and handedness. (7 points total)

	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:

- What is the chance of getting a male?  
i)  $647/963$  ii)  $90/963$  iii)  $737/963$  iv)  $647/963$  v)  $16/963$  vi)  $826/963$  vii)  $1142/963$
- What is the chance of getting someone who is left-handed?  
i)  $30/90$  ii)  $90/963$  iii)  $62/500$  iv)  $63/500$  v)  $826/963$  vi)  $266/826$  vii)  $37/65$
- What is the chance of drawing once and getting either a male or someone who is left-handed?  
i)  $60/963$  ii)  $737/963$  iii)  $677/963$  iv)  $30/963$  v)  $406/826$  vi)  $376/963$  vii)  $60/90$
- What is the chance you'll get a male if you draw only from the left-handers?  
i)  $90/316$  ii)  $30/90$  iii)  $30/316$  iv)  $60/90$  v)  $60/647$  vi)  $90/947$
- What is the chance you'll get a left hander if you draw only from the males?  
i)  $90/316$  ii)  $30/90$  iii)  $30/316$  iv)  $60/90$  v)  $60/647$  vi)  $90/947$
- Draw 3 students *without replacement*. What is the chance that all 3 students are left-handed?  
i)  $(90/963)^3$  ii)  $1 - (873/963)^3$  iii)  $1 - 90/963 * 89/962 * 88/961$  iv)  $90/963 * 89/962 * 88/961$
- Draw 3 students *with replacement*. What's the chance that not all 3 students are left-handed?  
i)  $1 - (873/963)^3$  ii)  $1 - (90/963)^3$  iii)  $1 - 90/963 * 89/962 * 88/961$  iv)  $873/963 * 872/962 * 871/961$

**Question 14** pertains 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. (3 points total)

- Draw 2 cards *without replacement*. What is the chance that both cards are fours?  
i)  $8/52$  ii)  $13/52 * 12/51$  iii)  $16/52$  iv)  $13/52 * 13/52$  v)  $4/52 * 4/52$  vi)  $4/52 * 3/51$  vii)  $17/52$
- Draw *one* card at random. What is the chance that it's *either* a Heart or a Diamond?  
i)  $8/52$  ii)  $26/52$  iii)  $16/52$  iv)  $13/52 * 13/52$  v)  $4/52 * 4/52$  vi)  $4/52 * 3/51$  vii)  $17/52$
- Draw *one* card at random. What's the chance that it's *either* a Queen or a Diamond?  
i)  $8/52$  ii)  $13/52 * 12/51$  iii)  $16/52$  iv)  $13/52 * 13/52$  v)  $4/52 * 4/52$  vi)  $4/52 * 3/51$  vii)  $17/52$

**Question 15** pertains to rolling fair dice. (5 points total)

- Two dice are rolled. What is the chance that the sum of the spots is 10?  
i)  $2/36$  ii)  $3/36$  iii)  $4/36$  iv)  $5/36$  v)  $1/6 * 1/6$  vi)  $7/36$
- One die is rolled. What is the chance of getting a 3 or a 4?  
i)  $2/6$  ii)  $1/6 + 1/6 - 1/36$  iii)  $4/36$  iv)  $6/36$  v)  $1/6 * 1/6$  vi)  $7/36$
- What is the chance of rolling a die 7 times and getting at least one "3"?  
i)  $(5/6)^7$  ii)  $1/6 * 1/6$  iii)  $1 - (5/6)^7$  iv)  $1/6 + 1/6 - 1/36$  v)  $1/6 + 1/6$
- What is the chance of rolling a die 7 times and getting no "3"s?  
i)  $(5/6)^7$  ii)  $(1/6)^7$  iii)  $1 - (5/6)^7$  iv)  $1 - (1/6)^7$  v)  $7 * (5/6)$
- What is the chance of rolling a die 7 times and getting all "3"s?  
i)  $(5/6)^7$  ii)  $1/6^7$  iii)  $1 - (5/6)^7$  iv)  $1 - (1/6)^7$  v)  $7 * (1/6)$

**Question 16** pertains to tossing fair coins. (2 points total)

- What is the chance of tossing a fair coin 3 times and getting all tails?  
i)  $1/2 * 1/2 * 1/2$  ii)  $1/2 + 1/2 + 1/2$  iii)  $1 - (1/2 * 1/2 * 1/2)$  iv)  $3/6$
- What is the chance of tossing a fair coin 3 times and getting this particular sequence: HHT?  
i)  $1/2 * 1/2 * 1/2$  ii)  $1/2 + 1/2 + 1/2$  iii)  $1 - (1/2 * 1/2 * 1/2)$  iv)  $3/6$