

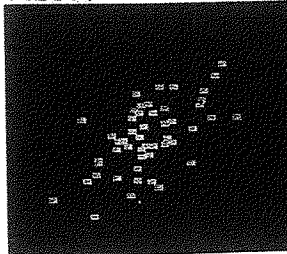
Page 1 Statistics

Statistics 100 Exam 2

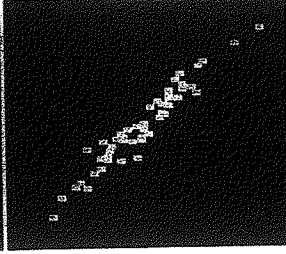
March 11th, 2015

Questions 1-4 pertain to the 4 scatter plots below. Match the scatter plots to the correlations given in the table.

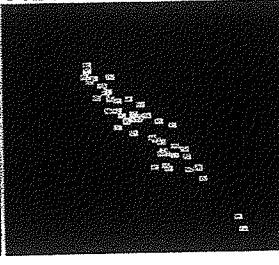
Plot A



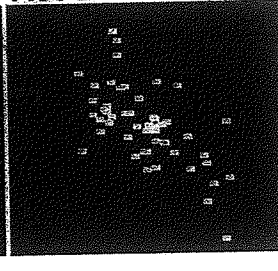
Plot B



Plot C



Plot D



Question 1 (1 point): Plot A

a) $r = -0.62$ b) $r = -0.95$ c) $r = 0.97$ d) $r = 0.60$

Question 2 (1 point): Plot B

a) $r = -0.62$ b) $r = -0.95$ c) $r = 0.97$ d) $r = 0.60$

Question 3 (1 point): Plot C

a) $r = -0.62$ b) $r = -0.95$ c) $r = 0.97$ d) $r = 0.60$

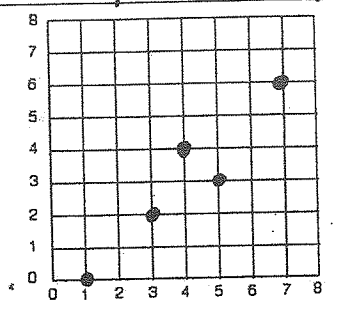
Question 4 (1 point): Plot D

a) $r = -0.62$ b) $r = -0.95$ c) $r = 0.97$ d) $r = 0.60$

Question 5: Part a) Compute the correlation coefficient (r) by filling in the table below including the totals. The average of $X=4$ and the SD of $X=2$. The average of $Y=3$ and the SD of $Y=2$. Plot the points on the graph below to create a scatterplot. (13 points total)

$\frac{1}{2}$ pt for each point

X	Y	Z-score for X	Z-score for Y	Products
1	0	-1.5	-1.5	2.25
3	2	-0.5	-0.5	0.25
4	4	0	0.5	0
5	3	0.5	0	0
7	6	1.5	1.5	2.25
Totals		Total should = 0	Total should = 0	Total = 4.75



Part b) (1 point) The correlation coefficient $r = 0.95$ (Fill in the blank and round to 2 decimal places)

Question 6: X and Y are 2 sets of numbers with a correlation coefficient of $r=0.4$. Fill in the 4 blanks below with numbers (NOT "increase" or "stay the same"). (4 points total)

- i) We add 0.5 to all X values. $r = 0.4$ ii) We subtract each Y value by 0.2. $r = 0.4$
 iii) We interchange all X and Y values. $r = 0.4$ iv) We multiply each X by -3. $r = -0.4$

Question 7 (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!

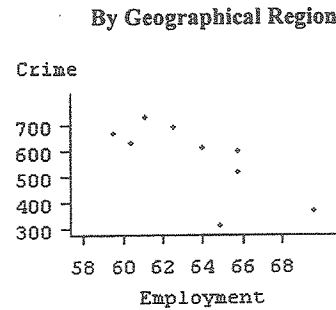
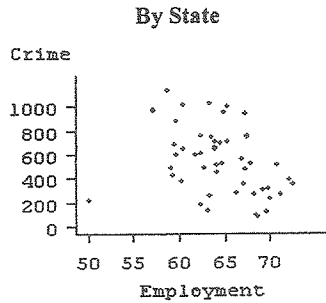
	Exactly +1	Between 0 and +1	About 0	Between 0 and -1	Exactly -1
How much you exercise and percent body fat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
As X gets larger, corresponding Y values get larger	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
X and Y values are paired randomly	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Y is always exactly twice it's corresponding X.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The sum of X and its corresponding Y value is always 10.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Page 2 Statistics

Statistics 100 Exam 2

March 11th, 2015

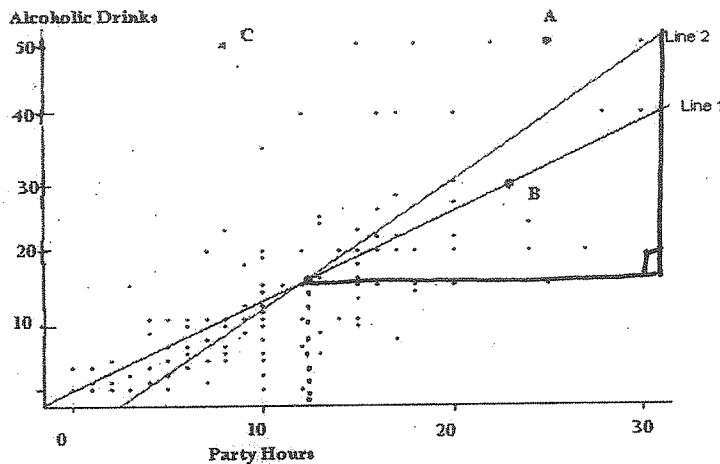
Question 8: Below are 2 scatter plots depicting the relationship between employment rate and crime rate by state and by geographical region. The graph on the left has 50 points: one for each *individual* state's employment and crime rate. The graph on the right has the same information condensed into 9 points; one for each geographical region in the country. In other words, each point in the geographical region plot represents the average employment rate and the average crime rate of the states in that region.



Which statement is best? (2 points)

- ☒ i) The graph of the 9 regions has a stronger correlation since it has eliminated the within region scatter.
- ☐ ii) The two graphs must have the same correlation coefficient since they're based on the same exact data.
- ☐ iii) The graph of the 50 states has a stronger correlation since it contains more data.

Question 9 pertains to the scatter diagram below which shows party hours per week on the X axis and the number of alcoholic drinks per week on the Y axis for the 148 Greek members who answered Survey 2 this semester.



- a) (1 point) The average number of party hours is around... Choose one: i) 10 ☒ ii) 12 iii) 18 iv) 25
- b) (2 points) Which is the regression line? Choose one: ☒ i) Line 1 ii) Line 2 iii) impossible to tell
- c) (2 points) The correlation between partying and drinking is closest to... Choose one: i) -0.5 ii) 0.3 ☒ iii) 0.7 iv) 1
- d) (1 point) The residual for person A is closest to... Choose one: i) 0 ii) -10 iii) -20 iv) 10 ☒ v) 20
- e) (1 point) The residual for person B is closest to... Choose one: ☒ i) 0 ii) -10 iii) -20 iv) 10 v) 20
- f) (1 point) If a new scatter plot was drawn with partying measured in minutes instead of hours then the correlation between party hours and drinks would.... Choose one: i) increase ii) decrease ☒ iii) stay the same
- g) (2 points) One student is exactly average in both drinks and party hours. Which line does he fall on? Choose one: i) Regression line only ii) SD line only ☒ iii) Both iv) Neither
- h) (2 points) One student is exactly 1.5 SDs below average in both drinks and party hours. Which line does he fall on? Choose one: i) Regression line only ☒ ii) SD line only iii) Both iv) Neither

2 of 7 pages (17 problems)

Page 3 Statistics

Statistics 100 Exam 2

March 11th, 2015

Question 10 pertains to our last survey

The table below gives the 5 summary statistics of the 290 students who responded to the questions: "How many drinks do you typically consume per week?" and "How many drinks do you think you'll consume on Unofficial St. Patrick's Day?" (10 points total)

	Average	SD
Drinks per Week	12	12
Drinks on Unofficial	15	16

Correlation Coefficient $r = 0.5$

a) Make regression estimates for Kira and Kevin by filling in the blanks in the table below. For Kira, you're given drinks per week and asked to predict drinks on Unofficial. For Kevin you're given drinks on Unofficial and asked to predict drinks per week. *Don't round your answers.*

Drinks per Week	Drinks per Week z-score	r	Drinks on Unofficial z-score	Drinks on Unofficial
Kira: 6 drinks per week Show work for full credit. $z = \frac{6-12}{12}$	$z = \frac{-6}{12} = -0.5$ (2 points)	$r = 0.5$	$z = \frac{-0.25}{1} = -0.25$ (1 point) $val = 15 + (-0.25)(16)$	drinks on unofficial = <u>11</u> (2 points) Show work for full credit.
<u>13.5</u> drinks per week (2 points) Show work for full credit. $value = 12 + (0.125)(12)$	$z = \frac{0.125}{1} = 0.125$ (1 point)	$r = 0.5$	$z = \frac{0.25}{2} = 0.25$ (2 points) $z = \frac{19-15}{16} = 0.25$	Kevin: 19 drinks on Unofficial 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) (2 points) Say we made a prediction of drinks on Unofficial for another student, Chauncey. We predict that Chauncey will drink 21 drinks on Unofficial. Knowing that his residual is 5, how many drinks did Chauncey actually have on Unofficial? Show work below and fill in the blank with a number.

Chauncey had 26 drinks.

$$\text{residual} = \text{actual} - \text{predicted}$$

$$5 = \text{actual} - 21$$

Question 11 (5 points total)

On April 15th, 1912 the Titanic sank into the North Atlantic Ocean. Only 15% of passengers were in first class. Suppose 60% of passengers in first class will survive. Also, suppose that 70% of passengers in second class, third class, or crew will die. Fill in the following table for a typical sample of 1,000 passengers.

	Survived	Died	Total
First Class	$.6(150) = 90$ (1 point)	60	150 (1 point)
Second Class, Third Class, or Crew	255 (1 point)	$.7(850) = 595$	850
Total	345	655	1,000

- a) (1 point) Given that a passenger died, what's the chance that they were in first class? 60/655
(leave your answer as a fraction)
- b) (1 point) Suppose a passenger was in second class, third class, or crew, what is the chance that they survived?
255/850
(leave your answer as a fraction)

Page 4 Statistics

Statistics 100 Exam 2

March 11th, 2015

Question 12 (15 points total)

A large class took 2 exams. The scatter plot of the exam scores was roughly football shaped. Below are the 5 summary statistics.

	Average	SD
Exam 1	70	20
Exam 2	80	10

Correlation: $r = 0.6$

- a) Find the regression equation for predicting Exam 2 from Exam 1. $Y = 0.3X + 59$

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

$$m = r \times \frac{SD_y}{SD_x} = 0.6 \times \frac{10}{20} = 0.3$$

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

$$y = 0.3x + b$$

$$80 = 0.3(70) + b \quad b = 59$$

- b) (2 points) Using the regression equation from part a, predict the exam 2 score of a student whose exam 1 score is 92. Round your answer to the nearest whole number.

$$y = 0.3(92) + 59 \approx 86 \text{ or } 87$$

- c) (2 points) Suppose the exam 1 score of the person above is actually an 88. What is his *residual*? Show work below:

$$\text{residual} = \text{actual} - \text{predicted}$$

$$= 88 - 87 = 1 \text{ or } 2$$

- d) (2 points) What is the SD of the prediction errors (the RMSE) when exam 2 scores from exam 1? Show work below and round your answer to 2 decimal places.

$$RMSE = \sqrt{1 - r^2} \times SD_y = \sqrt{1 - 0.6^2} \times 10 = 8$$

- e) (2 points) Of course, the prediction we made in part b isn't going to be exactly correct. Instead there's a **range** of exam 2 scores, with about 92% of them falling within ...

Choose one:

i) 1 SDerrors

☒ ii) 1.75 SDerrors

iii) 2.35 SDerrors

iv) Not enough info

- f) (2 points) The regression equation for estimating exam 1 scores from exam 2 scores is: Exam 1 score = $1.2 \times (\text{exam 2 score}) - 26$. How would you interpret the slope for this equation?

Choose one:

i) If you get a 0 on exam 2, on average you will get a 1.2 on exam 1

ii) If you get a 0 on exam 2, on average you will get a -26 on exam 1

☒ iii) On average, you will get about 1.2 points on exam 1 for every one point you get on exam 2.

iv) On average, you will lose about 26 points on exam 1 for every one point you get on exam 2.

- g) (1 point) Say that a student named Brooke happens to fall on the SD line. If her z-score for exam 1 is -0.3, what is her z-score for exam 2?

Choose one:

☒ i) -0.3

ii) 1

iii) 0

iv) 0.18

v) Not enough info to answer

Page 5 Statistics

Statistics 100 Exam 2

March 11th, 2015

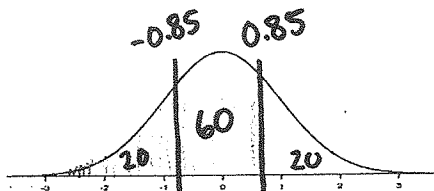
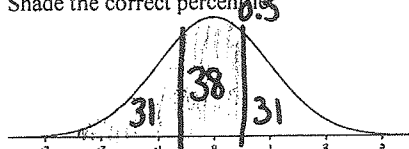
Question 13 (10 points total)

Suppose husbands' and wives' heights follow the normal curve but have different correlations in different countries.

a) Consider 5 countries where the correlation coefficients between the heights of husbands and wives are as given in the table below. If a husband is in the 15th percentile in height, estimate his wife's percentile in height for each country.

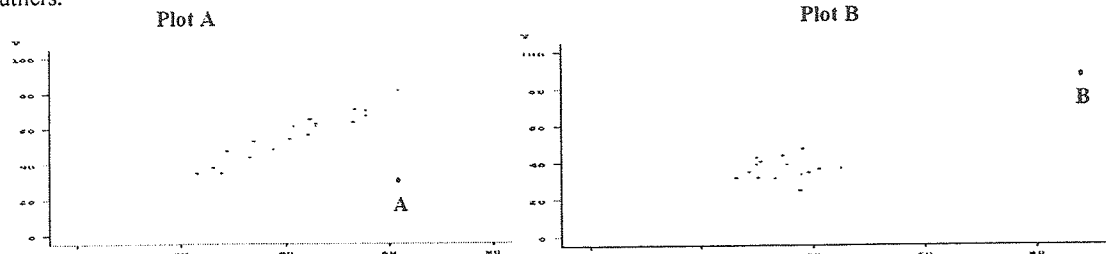
Husband's Percentile for Height	r	Wife's Percentile for Height
15 th	-1	Choose One: 10 th 15 th 34 th 50 th 66 th <u>85th</u> 90 th
15 th	1	Choose One: 10 th <u>15th</u> 34 th 50 th 66 th 85 th 90 th
15 th	0	Choose One: 10 th 15 th 34 th <u>50th</u> 66 th 85 th 90 th
15 th	0.4	Choose One: 10 th 15 th <u>34th</u> 50 th 66 th 85 th 90 th
15 th	-0.4	Choose One: 10 th 15 th 34 th 50 th <u>66th</u> 85 th 90 th

b) If a husband is in the 80th percentile for height where $r = 0.6$, what percentile would you estimate for his wife's height? Solve by filling in the table below. (You may round areas and z-scores to fit the nearest line on the table.)

Husband's Percentile for Height	Husband Z	r = 0.6	Wife Z	Wife's Percentile for Height
80 th Percentile	Z = <u>0.85</u> (1 point)	r = 0.6	Z = <u>0.5</u> (1 point)	Wife's Height Percentile = <u>69th</u> (1 point)
What middle area on the table should you look up to find the Z score? <u>60</u> % (1 point)			Round your z-score to the nearest tenth.	Mark the Z score on the graph below. Shade the correct percentile.
				
Correctly mark the z-score and shade the area corresponding to the 80 th percentile. (1/2 point for shading correctly)				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 14

Use the scatterplots below to answer the following questions. Both scatter plot A and scatter plot B pictured below have outliers.



- a) (1 point) If Point A was removed from Plot A, the correlation coefficient (r) would....
Choose one:
i) decrease ii) increase iii) stay the same iv) not enough information
- b) (1 point) If Point B was removed from Plot B, the correlation coefficient (r) would....
Choose one:
i) decrease ii) increase iii) stay the same iv) not enough information

Page 6 Statistics

Statistics 100 Exam 2

March 11th, 2015

Question 15 pertains to the table below which shows the survey responses of the 290 males and the 474 females to the question on our last survey: "Will you go to any of your classes drunk on Unofficial?" (7 points total)

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

Suppose you draw randomly from the students who answered this survey:

- What is the chance of getting a student certain to attend a class drunk?
 i) $46/764$ ii) $21/290$ **iii) $25/474$** iv) $25/46$ v) $100/764$ vi) $336/764$ vii) $315/764$
- What is the chance of getting a student certain to attend a class drunk if you draw only from the females?
 i) $46/764$ ii) $21/290$ **iii) $25/474$** iv) $25/46$ v) $100/764$ vi) $336/764$ vii) $315/764$
- What is the chance of getting a student certain to attend a class drunk if you draw only from the males?
 i) $46/764$ **ii) $21/290$** iii) $25/474$ iv) $25/46$ v) $100/764$ vi) $336/764$ vii) $315/764$
- What is the chance of getting a male student if you draw only from those certain to attend a class drunk?
 i) $21/46$ ii) $21/290$ iii) $25/474$ iv) $25/46$ v) $100/764$ vi) $336/764$ vii) $315/764$
- What is the chance that a randomly selected student will either probably or certainly come to a class drunk?
 i) $46/764$ ii) $21/290$ iii) $25/474$ iv) $25/46$ **v) $100/764$** vi) $336/764$ vii) $315/764$
- What is the chance that a randomly selected student is either female or certain to come to a class drunk?
 i) $46/764$ ii) $520/764$ iii) $25/474$ **iv) $495/764$** v) $100/764$ vi) $336/764$ vii) $315/764$
- Draw 3 students *without replacement*. What is the chance that not all of them are certain to attend a class drunk?
 i) $46/764 * 45/763 * 44/762$ ii) $46/764 + 45/763 + 44/762$ **iii) $1 - 46/764 * 45/763 * 44/762$** iv) $1 - (46/764)^3$

Question 16 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 *with replacement*. What is the chance that both cards are Diamonds?
 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 2 cards *without replacement*. What is the chance that both cards are Queens
 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'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 17 pertains to rolling fair dice. (4 points total)

1,5 5,1 2,4 4,2 3,3

- Two dice are rolled. What is the chance that the sum of the spots is 6?
 i) $2/36$ ii) $3/36$ iii) $4/36$ **iv) $5/36$** v) $1/6 * 1/6$ vi) $1/6 + 1/6$
- One die is rolled 3 times. What is the chance of getting no 6's?
i) $(5/6)^3$ ii) $(1/6)^3$ iii) $1 - (5/6)^3$ iv) $1 - (1/6)^3$ v) $3/6$
- One die is rolled 3 times. What is the chance of getting at least one 6?
 i) $(5/6)^3$ ii) $(1/6)^3$ **iii) $1 - (5/6)^3$** iv) $1 - (1/6)^3$ v) $3/6$
- One die is rolled 3 times. What is the chance of not getting all 6's?
 i) $(5/6)^3$ ii) $(1/6)^3$ iii) $1 - (5/6)^3$ **iv) $1 - (1/6)^3$** v) $3/6$

6 of 7 pages (17 problems)