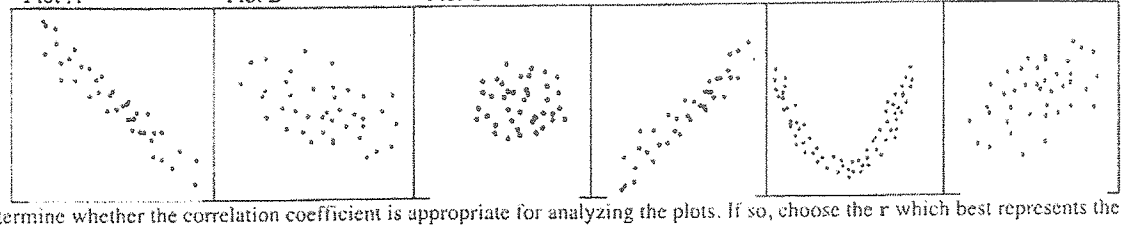
Page 1: Stats

Stats 100 Exam 2 F1 Balcony — Spring 2014

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

Plot A | Plot B | Plot C | Plot D | Plot E | Plot F



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!

	Not appropriate to use r	r = +0.9	r = +0.5	r = 0	r = −0.5	r = −0.9
Plot A	☐	☐	☐	☐	☐	☑
Plot B	☐	☐	☐	☐	☑	☐
Plot C	☐	☐	☐	☑	☐	☐
Plot D	☐	☑	☐	☐	☐	☐
Plot E	☑	☐	☐	☐	☐	☐
Plot F	☐	☐	☑	☐	☐	☐

Question 2 (6 pts.)

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!

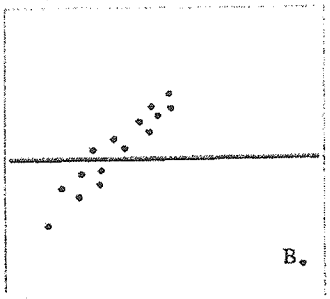
	Exactly +1	Between 0 and +1	About 0	Between 0 and -1	Exactly -1	Not Enough Information
As X ***increases***, Y tends to ***increase***.	☐	☑	☐	☐	☐	☐
As X ***increases***, Y tends to ***decrease***.	☐	☐	☐	☑	☐	☐
X and Y have no relation, they are just randomly paired.	☐	☐	☑	☐	☐	☐
Y is always less than X	☐	☐	☐	☐	☐	☑
Y is always ***exactly*** 1 less than X. $y = x-1$	☑	☐	☐	☐	☐	☐
X and Y ***always*** add up to 100. $x+y=100$, $y=100-x$	☐	☐	☐	☐	☑	☐

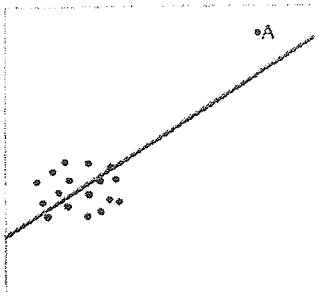
Question 3 pertains to the 2 scatter plots below. The regression line is shown for each plot. (4 pts.)

2 pts each

a) If we removed point B the correlation coefficient (r) would.... ***Choose one:***
(i) Increase ii) Decrease
iii) Stay the Same iv) Not enough info

b) If we removed point A the correlation coefficient (r) would.... ***Choose one:***
i) Increase (ii) Decrease
iii) Stay the Same iv) Not enough info





16 points total

Page 2 : Stats

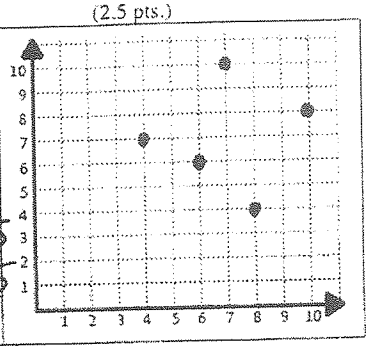
Stats 100 Exam 2 F1 Balcony — Spring 2014

Question 4 (18 pts.)
Part A: Calculate the correlation coefficient (r) by filling in the table. The averages of X and Y are both 7. The SD of X and Y are both 2. (13 pts.)

Plot the points on the graph below. (2.5 pts.)

X	Y	Z-score for X	Z-score for Y	Products
4	7	$z=\frac{4-7}{2}=-1.5$	$z=\frac{7-7}{2}=0$	$(-1.5)(0)=0$
6	6	$z=\frac{6-7}{2}=-0.5$	$z=\frac{6-7}{2}=-0.5$	$(-.5)(-.5)=0.25$
7	10	$z=\frac{7-7}{2}=0$	$z=\frac{10-7}{2}=1.5$	$(0)(1.5)=0$
8	4	$z=\frac{8-7}{2}=0.5$	$z=\frac{4-7}{2}=-1.5$	$(.5)(-1.5)=-0.75$
10	8	$z=\frac{10-7}{2}=1.5$	$z=\frac{8-7}{2}=0.5$	$(1.5)(.5)=0.75$
Totals		Total *should* = 0	Total *should* = 0	Total = 0.25

3pts



Part B: From above, find the correlation coefficient r. r = 0.05 (1 pt) $\frac{0.25}{5}=0.05$

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 pts.)

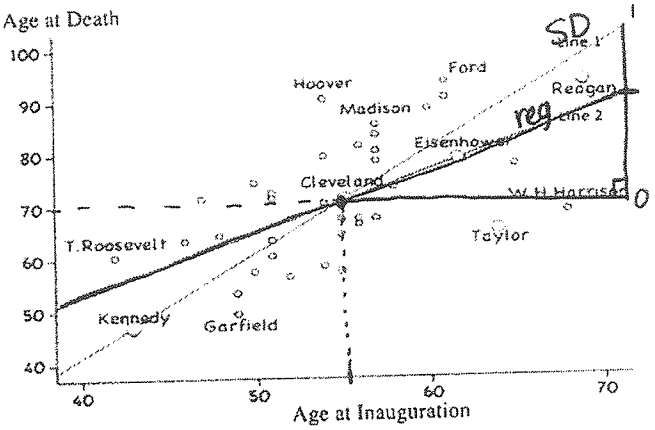
i) We subtract 2 from each Y value. r = 0.05

ii) We multiply each Y value by -3. r = -0.05

iii) We divide each Y value by 2. r = 0.05

iv) We swap the X and Y columns r = 0.05

Question 5: The graph below plots the age at inauguration vs the age at death of 39 US Presidents. (14 pts.)



2pts a) The SD Line and the Regression Line are shown. Which is the SD Line? Steeper
Choose One: (i) Line 1 ii) Line 2

2pts b) The correlation is closest to … *Choose One:*
i) 0.2 (ii) 0.6 ii) 0.9 iv) -0.2 v)-0.6

1pt c) About what is the average age at Death?
Choose One: i) 50 ii) 55 iii) 60 iv) 65 (v)70

1pt d) About what is the average age at Inauguration?
Choose One: i) 50 (ii) 55 iii) 60 iv) 65 v)70

2pts e) Taylor is an outlier that ____ the correlation.
Choose One:
i) strengthens (ii) weakens iii) does not affect

4 pts total, 1 for each

f) Which president(s) died exactly when predicted by the regression estimate? on reg line
Circle all that died when predicted: i) Kennedy (ii) Cleveland (iii) Eisenhower iv) Reagan

g) Which president(s) have the same z-scores for age at Inauguration and age at Death? on SD line
Circle all that have the same z-scores: (i) Kennedy (ii) Cleveland iii) Eisenhower iv) Reagan

h) The regression equation for predicting Age at Death from Age at Inauguration is: $Age_{death} = 1.2\,(Age_{Inauguration}) + 5.4$
Fill in the 2 blanks below with numbers.

1pt i) Ford was inaugurated at age 61, what is the regression estimate for his age at death? 78.6 years (Don't round.)
$y=1.2(61)+5.4=78.6$

1pt ii) Ford actually died at age 93. What is his residual (also called his prediction error)? 14.4 years (Don't round.)

residual = actual age − predicted age
= 93 − 78.6 = 14.4

32 points total

Page 3: Stats

Stats 100 Exam 2 F1 Balcony — Spring 2014

Question 6 (8 pts.)
Suppose the correlation between IQ scores and ACT scores have the following summary statistics among the population of students who take both exams and that the scatter plot is football-shaped:

	Average	SD
IQ	100	15
ACT	20	5

Correlation: r = 0.8

a) What is the slope of the regression equation when predicting ACT scores from IQ scores? *Choose one:*
i) 3 ii) 2.4 iii) 0.8 iv) 0.333 (v) 0.267
$m = r \times \frac{SD_y}{SD_x} = 0.8 \times \frac{5}{15} = .267$

b) What is the SD of the prediction errors (the RMSE) when predicting ACT scores from IQ scores?
Choose one: $\sqrt{1-r^2} \times SD_y$
i) 5 ii) 15 iii) $\sqrt{1-0.8^2} \times 15$ (iv) $\sqrt{1-0.8^2} \times 5 = 3$ v) 4

2 pts each

c) The regression equation predicts an ACT score of 24 for those who score 115 on the IQ test. Of course not everyone with an IQ of 115 will score 24 on the ACT test. Instead there's a range of ACT scores, with about 68% of them scoring between ... $24 \pm 1\,RMSE = 24 \pm 3$ (24−3 = 21, 24+3 = 27)
Choose one:
i) 15 and 25 (ii) 21 and 27 iii) 23 and 25 iv) 24 and 33

d) The regression equation for predicting IQ scores from ACT scores is: IQ = 2.4 x (ACT) + _____. What is the y-intercept? (the y-intercept is the blank in the above equation.)
Choose one: (i) 52 ii) -52 iii) -21.6 iv) 21.6 v) 0
plug in averages: $100 = 2.4(20) + b$, $b = 52$

Question 7 (12 pts.)
To the right are the 5 summary statistics for the ages of fathers and mothers of a large group of students (roughly based on our Survey data).

	Average	SD
Father's Age	52	6
Mother's Age	48	5

Correlation: r = 0.8

a) In the table below you are given the age of the fathers for 2 students. For each, compute the regression estimate for the age of the student's mother. **Show work, where indicated.**

Father's Age	Father's Z-score	r	Mother's Z-score	Mother's Age
Father is 64 years. *Show work for full credit.*	Z = 2 (2pt) $Z = \frac{64-52}{6} = \frac{12}{6} = 2$	r = 0.8	Z = 1.6 (1pt) *Show work for full credit.* $48 + 1.6(5) = 56$	Mother's Age = 56 (1pt)
Father is 46 years. *Show work for full credit.*	Z = −1 (2pt) $Z = \frac{46-52}{6} = \frac{-6}{6} = -1$	r = 0.8	Z = −0.8 (1pt) *Show work for full credit.* $48 + (-0.8)(5) = 44$	Mother's Age = 44 (1pt)

2 pts each

b) One student has a father who is 1.5 SD's above average in age. The regression estimate for his mother's age is 1.2 SD's above average. ***Fill in the blank in the sentence with the correct number.*** (2 pt.)
z_{dad} r z_{mom}: $1.5 \quad 0.8 = 1.5 \times 0.8 = 1.2$

c) Students with 58 year-old fathers have mothers who are 52 years old on the average. Students with 52 year-old mothers have fathers who are about _____ years old on the average. (2 pt.)
Choose the closest answer: i) 52 (ii) 56 iii) 58 iv) 60 v) not enough info given.
mom: 52 → z: $\frac{52-48}{5} = 0.8$ → r: 0.8 → z: 0.64 → dad: $52 + 0.64(6) = 55.84$

2 pts

Question 8 (2 pts)
The **average** yearly wine consumption and **average** life span in 50 countries was computed. The correlation for the **50 pairs of averages was 0.8**. Can you conclude that the correlation between wine consumption and life span for the millions of **individuals** in these 50 countries is also 0.8? ***Choose one:***
i) Yes, since it's based on the same info.
ii) No, it's likely to be higher than 0.8 since it's based off millions of data points instead of just 50.
(iii) No, it's likely to be lower than 0.8 since the millions of data points will produce scatter within each country.

22 points total

Page 4: Stats

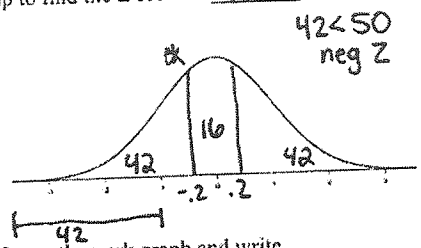
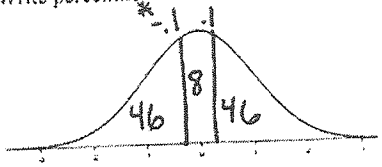
Stats 100 Exam 2 F1 Balcony — Spring 2014

Question 9 (9 pts.)
Suppose reading skills and math skills of grade school children follow the normal curve but have different correlations among different populations.

a) Consider 5 populations where the correlation coefficients between reading and math skills are as given in the table below. If a child is in the 75th percentile in reading skills, estimate her percentile in math skills in each population. (5 points)

Reading Skills Percentile	r	Math Skills Percentile
75th	0	Choose One: 20th 25th 30th **50th** 70th 75th 80th
75th	1	Choose One: 20th 25th 30th 50th 70th **75th** 80th
75th	-1	Choose One: 20th **25th** 30th 50th 70th 75th 80th
75th	0.8	Choose One: 20th 25th 30th 50th **70th** 75th 80th
75th	-0.8	Choose One: 20th 25th **30th** 50th 70th 75th 80th

b) If a student is in the 42nd percentile in reading skills where r = 0.5, estimate her math skills percentile *by filling in the table below.* ***You may round areas and z-scores to fit the nearest line on the table.*** Include ***negative*** sign for ***negative*** Z scores! (4 points)

Reading Percentile	Reading Z	r = 0.5	Math Z	Math Percentile
Student is in the 42nd percentile for reading skills. (That means 42% scored **lower.**) What middle area on the table should you look up to find the Z score? 16 % (1 pt) 42<50 neg Z  42 Correctly mark graph and write the correct Z in the next column. (1pt.)	Z = -0.2	r = 0.5	Z = -0.10 (1 pt)	Math Skills Percentile = 46 (1 pt) (Percentile is --What % scored *lower)* Mark the Z score on the graph below. Use the normal table to find the middle area. Write percentile in blank above.  **Round the middle area given in the NormalTable to the nearest WHOLE number andthen calculate the percentile.**

Question 10 pertains to a drug screening test for bus drivers. Suppose only 5% of bus drivers who take the test are really using drugs. Suppose **80%** of those using drugs will correctly test positive, but **20%** of those ***NOT*** using drugs will also test positive. ***Fill in the table*** for a typical sample of 100 people who get tested. (6 pts)

	Tests Positive	Tests Negative	Total
Using Drugs	(0.80)(5) = 4 (1pt)	1	*Hint: Fill in this blank first using the underlined info above.* (1pt) 5
Not Using Drugs	(0.20)(95) = 19 (1pt)	76	95
Total	23	77	100

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

a) A person tests positive, what's the chance (s)he is really using drugs? 4/23 (leave answer as a fraction) (1pt)

b) A person tests negative, what's the chance (s)he is really using drugs? 1/77 (leave answer as a fraction) (1pt)

15 points total

Page 5: Stats

Question 11 pertains to the table below that shows last semester's survey responses to 2 questions: "Have you ever tried marijuana?" and "Should the regulations on marijuana be more strict than those for alcohol, the same as those for alcohol, or less strict as those for alcohol?

	More Strict	Same	Less Strict	Totals
Tried Marijuana	59	188	162	409
Never Tried Marijuana	178	173	39	390
Totals	237	361	201	799

Suppose you randomly draw a student from those who answered this survey: *(4 pts)*

a) What is the chance that you'll get someone who answered "Less Strict"?
i) 162/201 ii) 162/409 iii) 409/799 (iv) 201/799 v) 162/799

b) What is the chance you'll get someone who answered "Less Strict" if you draw only from those who tried marijuana?
i) 162/201 (ii) 162/409 iii) 409/799 iv) 201/799 v) 162/799

c) What is the chance that you'll get someone who tried marijuana if you draw only from those who answered "Less Strict"?
(i) 162/201 ii) 162/409 iii) 409/799 iv) 201/799 v) 162/799
409/799 + 201/799 − 162/799

d) What is the chance that a randomly drawn student will either have tried marijuana *or* have answered "Less Strict"?
i) 162/201+162/409 ii) 162/201+162/409-162/799 (iii) (409 + 201-162)/799 iv) (409 + 201)/799 v) 162/799

Question 12 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. (So there are 4 Kings, 4 Queens and 13 Hearts.) (4 pts.)

a) Draw 2 cards **with replacement**. What is the chance that the first card is an King and the second is a Heart? $\frac{4}{52} \times \frac{13}{52}$
i) 4/52 + 13/52 ii) 4/52 + 13/52 -1/52 iii) 4/52*3/51 (iv) 4/52*13/52 v) 4/52*13/51

b) Draw 2 cards **without replacement**. What is the chance that the both cards are Queens? $\frac{4}{52} \times \frac{3}{51}$
i) 4/52 + 4/52 ii) 4/52 + 4/52 -1/52 iii) 4/52*4/52 (iv) 4/52*3/51 v) 4/52*4/51

c) Draw *one* card. What is the chance that it's **either** a King **or** a Queen? $\frac{4}{52} + \frac{4}{52}$
i) 4/52 + 4/52 -1/52 (ii) 4/52 + 4/52 iii) 4/52*4/52 iv) 4/52*3/51 v) 4/52*4/51

d) Draw *one* card. What is the chance that it's **either** a King **or** a Heart? $\frac{4}{52} + \frac{13}{52} - \frac{1}{52}$
(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

Question 13 pertains to rolling fair dice. (7 pts.)
1,4 3,2
4,1 2,3

a) Two dice are rolled. What is the chance that the sum of the spots is 5?
i) 2/36 ii) 3/36 (iii) 4/36 iv) 5/36 v) 1/6*1/6 vi) 1/6 + 1/6

b) One die is rolled 4 times. What is the chance of getting *no* 4'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* 4? P(at least 1) = 1 − P(none)
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) $(1/6)^3$ iii) $1-(5/6)^3$ iv) $1-(1/6)^3$ v) 3/6

e) One die is rolled 3 times. What is the chance of getting *not all* 2's? P(not all) = 1 − P(all)
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 2 *or* the second roll is a 3? $\frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{6}{36} + \frac{6}{36} - \frac{1}{36}$
i) 1/36 ii) 2/36 iii) 6/36 (iv) 11/36 v) 12/36

g) One die is rolled twice. What is the chance that the first roll is a 2 ***and*** the second roll is a 3? $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$
(i) 1/36 ii) 2/36 iii) 6/36 iv) 11/36 v) 12/36

15 points total : Each question worth 1 point