

**EXAM 2: Statistics 100**

Spring 2019 Key

**Cover Sheet Questions (1 pt.)**

1) What's your name? \_\_\_\_\_

(Last name)

(First name)

2) What's your net ID (email)? \_\_\_\_\_

3) Which Section are you in?

**Circle one:**    i) L1 (MWF at 12pm)            ii) L2 (Tues/Thurs 11am)            iii) ONLINE

Write answers in appropriate blanks. All multiple-choice questions have exactly one answer. If you circle more than one answer you will automatically be marked wrong.

**\*SHOW WORK** when requested, otherwise no credit. Do NOT use scrap paper.

**Make sure you have all 6 pages including the normal table (15 questions).**

DO NOT WRITE BELOW THIS LINE

**The numbers written in each blank below indicate how many points you missed on each page.**  
The numbers printed to the right of each blank indicate how many points each page is worth.

Page 1 \_\_\_\_\_ 17

Page 2 \_\_\_\_\_ 25

Page 3 \_\_\_\_\_ 18

Page 4 \_\_\_\_\_ 19

Page 5 \_\_\_\_\_ 20

Cover Page \_\_\_\_\_ 1

**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.

1 pt. for answering cover page questions 1-3 correctly!

**Total Score** \_\_\_\_\_

**There is NO CLASS on Thursday or Friday!**

**Scores will be posted on Compass by Friday morning and exams will be returned in class next week. Online students may pick up their exam in 23 Illini Hall during office hours next week.**

**Question 1 (7 points)**

You guys took exam 1 exactly a month ago on February 6<sup>th</sup>! Here is a random sample of 10 scores:  
22, 69, 75, 80, 81, 88, 90, 92, 97, 100

med

To draw a boxplot of the data, you'd need to calculate Q1, Q2, Q3 and the IQR. You'd also need to check for outliers. Fill in the blanks below:

Step 1: Find Q1, Q2, Q3, and the IQR	Step 2: Check for outliers (no work, no credit)
Q1= <u>75</u> Q2= <u>84.5</u> Q3= <u>92</u> IQR= <u>17</u>	Are there any low outliers? Choose one: <input checked="" type="radio"/> a) yes    b) no low outliers $< Q1 - 1.5(IQR) = 75 - 1.5(17) = 49.5$ Are there any high outliers? Choose one: a) yes <input checked="" type="radio"/> b) no high outliers $> Q3 + 1.5(IQR) = 92 + 1.5(17) = 117.5$ List the outliers here: <u>22</u> *if there are none, write "none" in the blank.

continued error  
Q3-Q1

**Question 2 (7 points)**

Instead of just looking at 10 people like we did in the previous question, let's look at the entire class. To save paper, instead of giving you every student's score, here is a distribution table showing what percent of the class got the following scores:

Score	Percent (%)
0-62	2%
62-80	23%
80-85	10%
85-89	15%
89-92	25%
92-100	25%

- a) Q1= 80      b) Q2= 89      c) Q3= 92
- d) The middle 50% of the data lies between 80 and 92. Fill in the blanks with numbers.  
cont error from Q1 + Q3
- e) What percentage of the students are low outliers?  
Choose one: i) 0%     ii) 2%    iii) 25%    iv) 10%  
low outliers  $< Q1 - 1.5(IQR) = 80 - 1.5(12) = 62$
- f) What percentage of the students are high outliers?  
Choose one:  i) 0%    ii) 2%    iii) 25%    iv) 10%  
high outliers  $> Q3 + 1.5(IQR) = 92 + 1.5(12) = 110$

**Question 3 (3 points)**

Let's draw a boxplot! Last semester, students just like you took exam 2 in October. Here's how they did. Draw a vertical boxplot on the graph below:

Info Needed to Draw Box Plot	Draw the boxplot on the graph below. Make sure to include the outliers!
1. Q1= 60, Q2= 80, Q3=90 2. There were 2 low outliers: 0 and 10 3. The lowest score that wasn't an outlier was 30. 4. There were no high outliers. 5. The highest score was 100.	<p>3 points for perfect 1 point if they drew something that looks like a boxplot or go it partially correct otherwise 0</p>

**Question 4 (5 points)**

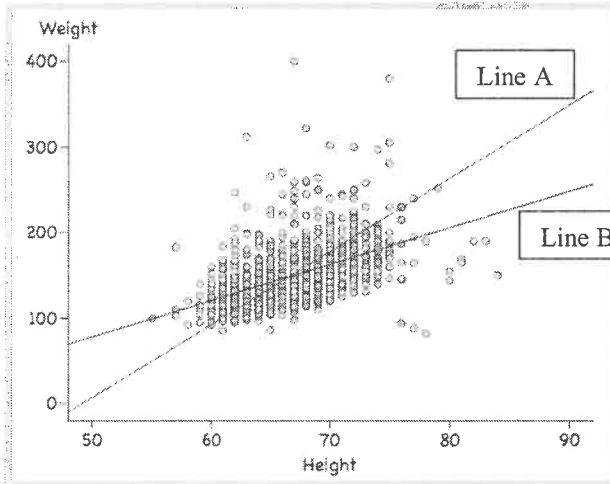
X and Y are two lists of numbers with a negative correlation of  $r = -0.3$ . Answer the following questions:

- a) Does the scatterplot of X and Y slope up to the right or down to the right? **Circle one:**  
 i) Up **ii) Down** iii) It could slope up or down depending on whether X and Y values are negative or positive.

For parts b-e, still assume X and Y are 2 lists of numbers with  $r = -0.3$ . Fill in the blanks with **NUMBERS**, not words.

- b) If 5 is subtracted from all of the X values,  $r = -0.3$   
 c) If all the original X values are doubled,  $r = -0.3$   
 d) If all the original Y values are multiplied by -2,  $r = 0.3$   
 e) If all the original X and Y values were switched,  $r = -0.3$

**Question 5 pertains to the scatterplot below of weight in pounds and height in inches. (20 points)**



- a) (2 pts.) Look at Line A and Line B on the scatter plot. One is the regression line and one is the SD line. Which is the SD line? **Choose one:** **i) Line A** ii) Line B  
 b) (2 pts.) The correlation (r) between height and weight is closest to? **Choose one:** i) -0.2 ii) 0.2 **iii) 0.5** iv) 0.95  
 c) (2 pts.) The average height is closest to ...  
**Choose one:** i) 60 **ii) 67** iii) 72 iv) 148  
 d) (2 pts.) The largest prediction error is 85 & the smallest is -61. What is the average all the prediction errors (residuals)?  
0

Fill in the above blank with a number.

e) (2 pts.) If a new scatter plot was drawn with height measured in centimeters (2.54 cm/inch) and weight measured in kilograms (2.2 lbs/kg) the correlation coefficient (r) between height and weight would ...

- Choose one:** i) increase ii) decrease **iii) stay the same**

f) (2 pts.) One student (not labeled) falls exactly on the regression line. What is his prediction error (residual)? 0  
**Fill in the above blank with a number.**

g) (2 pts.) One student (not labeled) falls exactly on the SD line. If he is 2 SD's above average in height, then how many SD's above average is he in weight? 2 **Fill in the blank with a number.**

h) (2 pts.) The regression equation for predicting weight from height is:  $\text{Weight} = 4.3 * (\text{Height}) - 134.4$   
 Use the equation to predict the weight of someone who is 73 inches tall. 179.5 (No work, no credit. Round your answer to two decimal places.)  
 $\text{weight} = 4.3(73) - 134.4 = 179.5$  1pt work  
1pt answer

i) (2 pts.) The SD of the prediction errors (RMSE) is 30 lbs., so your prediction for weight +/- 30 lbs. will be right about 68 % of the time and your prediction for weight +/- 60 lbs. will be right about 95 % of the time. **Fill in the blanks with numbers.** 1 pt for each blank  
 accept 68.27 + 95.45

- j) (2 pts.) Which of the following statements are true about the scatter plot above? **Choose one:**  
 i) All points on the SD line have positive prediction errors.  
 ii) All points on the SD line have negative prediction errors.  
**iii) Points on the SD line with above average X values have positive prediction errors and points on the SD line with below average X values have negative prediction errors.**

**Question 6 (8 points)**

This question pertains to the height and shoe size of the 250 male students who responded to Survey 1 this semester. Here are the 5 rounded summary statistics:

	Average	SD
Shoe Size	11	1.5
Height	70	3
$r = 0.75$		

a) (6 points) In the table below, you're either given a student's height or his shoe size. Make regression estimates (predictions) for shoe size or height by filling in the blanks in the table below and showing work where indicated. Showing no work where indicated will result in no credit.

1 pt for each blank →

Height	Height Z Score	r	Shoe Size Z Score	Shoe Size
76"	$Z = 2$ Show work below. $Z = \frac{76-70}{3} = 2$	$r=0.75$	$Z = 1.5$	Shoe Size = <u>13.25</u> Show work below. $val = 11 + (1.5)(1.5) = 13.25$
Height = <u>65.5</u> Show work below. $val = 70 + (-1.5)(3) = 65.5$	$Z = -1.5$	$r=0.75$	$Z = -2$ Show work below. $Z = \frac{8-11}{1.5} = -2$	Shoe Size = 8

b) (2 points) A student is 1 SD above average in shoe size and falls exactly on the SD line. How tall are they? Height = 73. Fill in the blank with a number.

**Question 7 (10 points)**

Suppose a large public health survey including thousands of participants yielded the following results on the heights of father's and their daughters:

	Average	SD
Fathers' Heights	70	3
Daughters' Heights	64	2
$r=0.5$		

a) (2 pts.) The slope of the regression equation for predicting fathers' heights from daughters' heights is...  
 Choose one: i) 0.5    ii) 3/2    iii) 2/3    **iv) 0.5\*3/2**    v) 0.5\*2/3

b) (2 pts.) The SD of the prediction errors (RMSE) for predicting fathers' heights from daughters' heights is 2.598. Show work below. Round your answer to 3 decimal places. No work, no credit. 1pt work 1pt answer  
 $RMSE = \sqrt{1-r^2} \times SD_y = \sqrt{1-0.5^2} \times 3 = 2.598$  or 2.6

c) (2 pts.) The regression line predicts that if a daughter is 60 inches, the father is 67 inches. One father in the study has a daughter who is 60 inches, but he's actually only 63 inches himself. What is his prediction error or residual? -4. Show work below. No work, no credit. 1pt work 1pt answer  
 $residual = actual - predicted = 63 - 67 = -4$

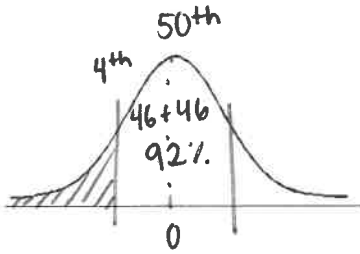
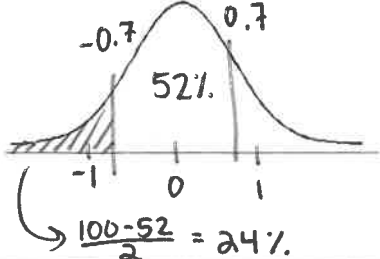
d) (2 pts.) The slope of the regression equation for predicting daughters' heights from fathers' heights is 0.33. What is the y-intercept? Fill in the blank below with the y-intercept. Round your answer to 2 decimal places. No work, no credit. 1pt work 1pt answer  
 Daughter Height =  $0.33 * \text{Father Height} + 40.9$   
 $64 = 0.33(70) + b$      $b = 40.9$

e) (2 pts.) This survey included thousands of fathers and daughters and had a correlation coefficient of 0.5. If we took these participants and looked at the average father and daughter heights for each of the 50 states, r would....  
 Choose one: **i) increase**    ii) decrease    iii) stay the same

**Question 8 (9 points)**

Suppose scores on physical fitness and math skills tests follow the normal curve but have different correlations at different schools. Answer the following questions below. **Round all answer to 2 decimal places when necessary.**

- a) (4 pts.) At School A, the correlation between fitness and math skills is 0.4. If someone is in the 4<sup>th</sup> percentile in fitness, estimate their percentile in math. **Show work when requested, otherwise no credit.**

Percentile in Physical Fitness	Fitness Z-Score	r	Math Z-Score	Percentile in Math Skills
4 <sup>th</sup> Percentile What middle area corresponds to the 4 <sup>th</sup> percentile? <u>92</u> 	$Z = -1.75$	$r = 0.4$	$Z = -0.7$ Cont error	Math Percentile = <u>24<sup>th</sup></u> <i>accept between 24 + 24.5</i> Show work by marking your z-score and shading the percentile. 

- b) (2 pts.) At School B, the correlation between physical fitness and math skills is -0.4. If a student is in the 4<sup>th</sup> percentile for physical fitness, their predicted percentile for math would be 76<sup>th</sup>. Fill in the blank with the correct percentile. **No work necessary! Use part a. Cont error**  
*should be 100-24 or whatever percentile from a*
- c) (3 pts.) The correlations between physical fitness and math for 3 other schools are given in the table below. If a student scores in the 85<sup>th</sup> percentile for physical fitness, estimate his math percentile at each school.

Physical Fitness Percentile	r	Math Percentile
85 <sup>th</sup>	1	85
85 <sup>th</sup>	-1	15
85 <sup>th</sup>	0	50

**Question 9 (3 points)**

Circle True or False for the following statements about the correlation coefficient, r.

- a) r always has to be between -1 and 1 Choose one:  True  False
- b) If I delete an outlier, r will always get stronger (bigger in absolute value) Choose one:  True  False
- c) If I have two sets of points with a negative correlation (negative r), the slope of the regression line will always be negative. Choose one:  True  False

**Question 10 (7 points)**

Fill in the following table and calculate the correlation coefficient, r. The mean of X is 1 and the mean of Y is 2.  $SD_X = 1, SD_Y = 2$ . Round the correlation coefficient to 2 decimal places.

X	Y	Z score for X	Z Score for Y	Products
0	2	-1	0	0
1	0	0	-1	0
3	6	2	2	4
0	0	-1	-1	1
1	2	0	0	0
1	2	0	0	0

$r = 0.83$   
 1 pt

$$r = \frac{0+0+4+1+0+0}{6} = \frac{5}{6} = 0.83$$

*1/2 pt for each blank*

**Question 11 (8 points)**

Let's say that 50% of women who take pregnancy tests are actually pregnant. Suppose 90% of women who are pregnant will correctly get a positive result. 5% of women who are not pregnant will also get a positive result (false positives- scary!). Fill in the following table for a sample of 10,000 women and answer the questions below. *1/2 pt for each blank*

	Tests Positive	Tests Negative	Total
Pregnant	$.90(5000) = 4500$	500	5,000
Not Pregnant	$.05(5000) = 250$	4750	5,000
Total	4750	5250	10,000

- a) A woman gets a positive test result, what's the chance she's actually pregnant?  $\frac{4500}{4750}$   
 Leave your answer as a fraction. **2 pts**
- b) Given that a woman is not pregnant, what's the chance she'll get a negative result?  $\frac{4750}{5000}$   
 Leave your answer as a fraction. **2 pts**

**Question 12 (4 points)** The question, "Do you have special abilities?" was asked of 100 superheroes from either the Marvel or DC universe. Results are shown in the table.

	Yes	No	Total
Marvel	25	35	60
DC	10	30	40
Total	35	65	100

- a) What is the probability that a randomly chosen superhero claims to have special abilities?  
 Choose one:  i) 35/100    ii) 65/100    iii) 25/100    iv) 60/100    v) 100/100
- b) What is the probability that a randomly chosen superhero is created by Marvel?  
 Choose one: i) 35/100    ii) 65/100    iii) 25/100     iv) 60/100    v) 100/100
- c) What is the probability of a superhero claiming to have special abilities given that they are from the DC universe?  
 Choose one: i) 10/35    ii) 35/65    iii) 30/40    iv) 10/100     v) 10/40
- d) Given that a superhero does not claim to have special abilities, what is the probability that they were created by Marvel? Choose one: i) 10/35     ii) 35/65    iii) 30/40    iv) 10/100    v) 10/40

**Question 13 (3 points)** A jar contains 3 red marbles, 7 white marbles and 10 blue marbles. If 2 marbles are drawn from the jar at random **without replacement**, find the probability that:

- a) The first marble is white, and the second marble is blue:  
 Choose one: i)  $7/20 * 10/20$     ii)  $7/20 * 9/19$      iii)  $7/20 * 10/19$     iv)  $7/20 * 3/19$
- b) Both marbles are red: Choose one:  i)  $3/20 * 2/19$     ii)  $3/20 * 3/19$     iii)  $3/20 * 3/20$     iv)  $3/20 * 3/20$
- c) Neither of the marbles is red: Choose one: i)  $3/20 * 2/19$     ii)  $7/20 * 6/19$      iii)  $17/20 * 16/19$     iv)  $17/20 * 17/20$

**Here's all of the info that you will need to know to answer Questions 14 and 15:** There are 52 cards in a deck. There are 4 queens, 13 diamonds, 13 clubs, and 13 spades. There is 1 ace of spades.

**Question 14 (4 points)** Draw 4 cards **with replacement** out of a standard deck of 52 cards:

- a) What is the probability never drawing the Ace of Spades?  
 Choose one:  i)  $(51/52)^4$     ii)  $(4/52)^4$     iii)  $1 - (51/52)^4$     iv)  $(1/52)^4$     ~~v)  $(13/52)^4$~~
- b) What is the probability of drawing at least one Queen? =  $1 - P(\text{no Queens})$   
 Choose one: i)  $1 - (39/52)^4$     ii)  $1 - (4/52)^4$     iii)  $1 - (13/52)^4$      iv)  $1 - (48/52)^4$     v)  $(13/52)^4$
- c) What is the probability of drawing all diamonds?  
 Choose one: i)  $1 - (39/52)^4$     ii)  $1 - (4/52)^4$     iii)  $1 - (13/52)^4$     iv)  $(4/52)^4$      v)  $(13/52)^4$
- d) What is the probability of not getting all clubs? =  $1 - P(\text{all clubs})$   
 Choose one: i)  $1 - (39/52)^4$     ii)  $1 - (4/52)^4$      iii)  $1 - (13/52)^4$     iv)  $1 - (48/52)^4$     v)  $(13/52)^4$

**Question 15 (1 point)**

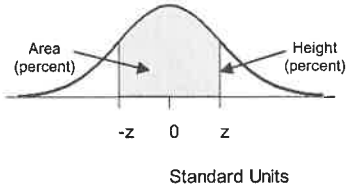
Draw one card at random from a deck. What is the probability of drawing either a queen or a spade?

- Choose one: a) 17/52     b) 16/52    c) 4/52    d) 13/52    e) 26/52    f) 8/52

$$P(\text{queen or spade}) = P(\text{queen}) + P(\text{spade}) - P(\text{both})$$

$$= 4/52 + 13/52 - 1/52 = 16/52$$

**STANDARD NORMAL TABLE**



<i>z</i>	<i>Area</i>		<i>z</i>	<i>Area</i>		<i>z</i>	<i>Area</i>
0.00	0.00		1.50	86.64		3.00	99.730
0.05	3.99		1.55	87.89		3.05	99.771
0.10	7.97		1.60	89.04		3.10	99.806
0.15	11.92		1.65	90.11		3.15	99.837
0.20	15.85		1.70	91.09		3.20	99.863
0.25	19.74		1.75	91.99		3.25	99.885
0.30	23.58		1.80	92.81		3.30	99.903
0.35	27.37		1.85	93.57		3.35	99.919
0.40	31.08		1.90	94.26		3.40	99.933
0.45	34.73		1.95	94.88		3.45	99.944
0.50	38.29		2.00	95.45		3.50	99.953
0.55	41.77		2.05	95.96		3.55	99.961
0.60	45.15		2.10	96.43		3.60	99.968
0.65	48.43		2.15	96.84		3.65	99.974
0.70	51.61		2.20	97.22		3.70	99.978
0.75	54.67		2.25	97.56		3.75	99.982
0.80	57.63		2.30	97.86		3.80	99.986
0.85	60.47		2.35	98.12		3.85	99.988
0.90	63.19		2.40	98.36		3.90	99.990
0.95	65.79		2.45	98.57		3.95	99.992
1.00	68.27		2.50	98.76		4.00	99.9937
1.05	70.63		2.55	98.92		4.05	99.9949
1.10	72.87		2.60	99.07		4.10	99.9959
1.15	74.99		2.65	99.20		4.15	99.9967
1.20	76.99		2.70	99.31		4.20	99.9973
1.25	78.87		2.75	99.40		4.25	99.9979
1.30	80.64		2.80	99.49		4.30	99.9983
1.35	82.30		2.85	99.56		4.35	99.9986
1.40	83.85		2.90	99.63		4.40	99.9989
1.45	85.29		2.95	99.68		4.45	99.9991