EXAM 2: Statistics 100

READ THE DIRECTIONS BELOW TWICE!

Cover Sheet Questions

1) What's your **name**? KEY _____

(Last name) (First name)

2) What's your **net ID** (email)? _____@illinois.edu

3) Which **section** are you in? ***Circle one:***

i) L1 (In Person Section) **ii)** ONL (Online Section)

This test is ALL multiple choice. **Circle all answers on this exam and fill in the corresponding bubble on your orange scantron**. All questions have exactly one answer. If you circle/bubble in more than one answer, you will automatically be marked wrong. Make sure to circle the answers on this test and fill out your scantron. If you don't do both, you will get a 0.

SCANTRON Directions

- Print and bubble in your LAST NAME with **no spaces** starting in the left most column. Print your FIRST INITIAL in the right-most column.
- Print and bubble in your UIN number in the Student Number box.
- Print and bubble in your NET ID with **no spaces** in the NETWORK ID box.
- *Write Stat 100 on the COURSE line.*
- *Write your instructor's name (Karle Flanagan) on the INSTRUCTOR line.*
- *Write your section (L1 or ONL) on the SECTION line.*
- *Sign your name, and right underneath the student signature line PRINT your name.*

READ THIS: Failure to fill out your scantron correctly will result in a loss of 2 points on your exam!

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.

Make sure you have all 8 pages including the normal table (76 questions).

There is NO CLASS on Friday this week!

Scores will be posted on Canvas by Friday at 5pm. Students may pick up their exam in 171 Computer Applications Building during office hours next week.

Questions 1-5 pertain to the following situation: Below is a distribution table for the scores on a Physics exam for a large class. The right-hand column shows the % of scores in each interval. The lowest score was 0, the second lowest was 6, and the highest score was 100. To draw a box plot of the data you'd have to find the Q1, Q2 and Q3.

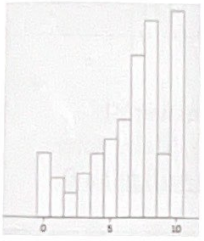
Score	%
0-50	25
50-60	10
60-70	15
70-80	25
80-90	20
90-100	5

1) **What is Q2?** **a)** 25 **b)** 50 **c)** 70 **d)** 80 **e)** 15
2) **What is Q1?** **a)** 25 **b)** 50 **c)** 70 **d)** 80 **e)** 15
3) **What is Q3?** **a)** 25 **b)** 50 **c)** 70 **d)** 80 **e)** 15
4) **What is the IQR?** **a)** 25 **b)** 50 **c)** 80 **d)** 10 **e)** 70
5) **Are there any outliers?**
   - **a)** Yes, there are only high outliers
   - **b)** Yes, there are only low outliers
   - **c)** Yes, there are both high and low outliers
   - **d)** No, there are no outliers

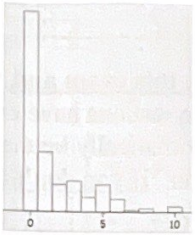
low < Q1 - (1.5)(IQR)

Questions 6-8: Which histograms correspond to which box plots?

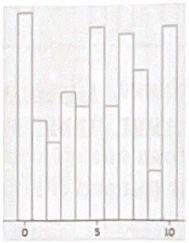
Histogram A



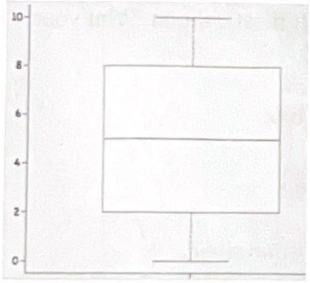
Histogram B



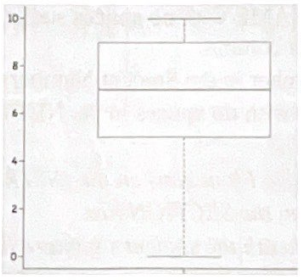
Histogram C



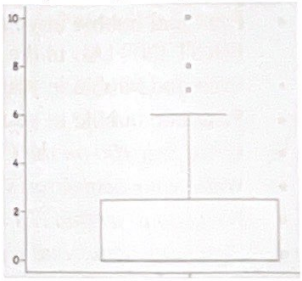
Write the correct letter in each blank below to match the box plots with their corresponding histograms.



skew



leans higher



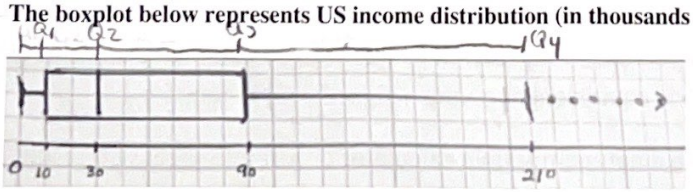
leans lower

6. Histogram _____
a) A **b)** B **c)** C

7. Histogram _____
a) A **b)** B **c)** C

8. Histogram _____
a) A **b)** B **c)** C

The boxplot below represents US income distribution (in thousands of dollars) in 2015.



all quartiles are 25% of the data!

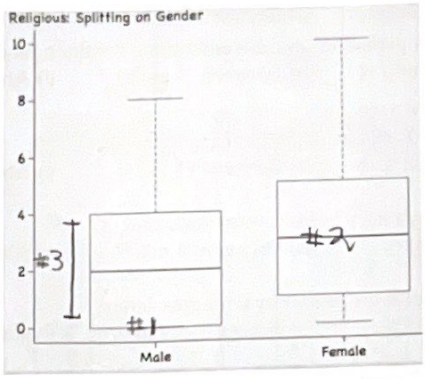
[Q1] **9.** What percent of the population earned less than $10,000?
a) 25% **b)** 50% **c)** 75% **d)** impossible to tell

[Q2] **10.** What percent of the population earned between $10,000 and $30,000?
a) 25% **b)** 50% **c)** 75% **d)** impossible to tell

[Q3] **11.** What percent of the population earned between $30,000 and $90,000?
a) 25% **b)** 50% **c)** 75% **d)** impossible to tell

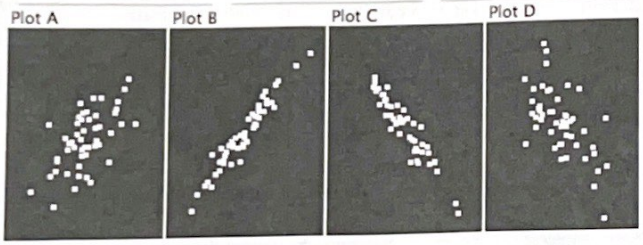
Questions 12-14 pertain to the following situation: The 2 boxplots below depict the Stat 100 survey responses of 140 males and 100 females to the question: "On a scale of 0 to 10, rate how religious you are." (0 is not at all and 10 is extremely). Fill in the table below. All of the answers are whole numbers.

	Males	Females
Average	2.46	3.15
SD	2.38	2.56
Min	Blank 1	
Q1		
Med		Blank 2
Q3		
Max		
IQR	Blank 3	
n	140	100



12. What goes in Blank 1? **a)** 0 **b)** 1 **c)** 2 **d)** 3 **e)** 4
13. What goes in Blank 2? **a)** 0 **b)** 1 **c)** 2 **d)** 3 **e)** 4
14. What goes in Blank 3? **a)** 0 **b)** 1 **c)** 2 **d)** 3 **e)** 4

Questions 15-18 pertain to the following scatterplots.



Write the letter of the plot next to the correlation coefficient that is closest to it.

15. r = - 0.95
a) Plot A **b)** Plot B **c)** Plot C **d)** Plot D

16. r = - 0.62
a) Plot A **b)** Plot B **c)** Plot C **d)** Plot D

17. r = 0.6
a) Plot A **b)** Plot B **c)** Plot C **d)** Plot D

18. r = 0.97 **a)** Plot A **b)** Plot B **c)** Plot C **d)** Plot D

Questions 19-22 pertain to 2 sets of numbers (X and Y) with a correlation coefficient of **r = - 0.75.**
19. If all the original x and y values are multiplied by 4, r would be? **a)** -0.75 **b)** +0.75 **c)** impossible to tell
20. If all the original x values are multiplied by ***negative*** 0.6, r would be? **a)** -0.75 **b)** +0.75 **c)** impossible to tell
21. If 100 is added to only the first x value, r would be? **a)** -0.75 **b)** +0.75 **c)** impossible to tell
22. If all the original X and Y values are switched, r would be? **a)** -0.75 **b)** +0.75 **c)** impossible to tell

Questions 23-25 pertain to this situation: Compute the correlation coefficient (**r**) between X and Y by filling in the table below. A few cells have been done for you. **The average of X = 4 and the SD of X = 2. The average of Y = 3 and the SD of Y = 2.**

X	Y	Z score for X	Z score for Y	Products
1	2	-1.5	-0.5	Blank 1
3	0	-0.5	-1.5	0.75
4	4	0	Blank 2	0
5	3	0.5	0	0
7	6	1.5	1.5	2.25
Totals		**Total *should* =** 0 Check the column sums to what it should	**Total *should* =** 0 Check the column sums to what it should.	**Total =** 3.75

23. What goes in Blank 1?
a) -0.5 **b)** -1.5 **c)** 0.75 **d)** 0

24. What goes in Blank 2?
a) 1 **b)** 0.5 **c)** -0.75 **d)** 0

25. What is the correlation coefficient? **a)** 0.75 **b)** 3.75 **c)** -1 **d)** -0.5

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

$Z = \frac{val - avg}{SD}$

Questions 26-30 pertain to the following pairs of variables. For each of the following pairs of variables, pick the option that best describes its correlation.

26. As X values get larger, the corresponding Y values also get larger.
a) Exactly -1 **b)** Between -1 and 0 **c)** About 0 **(d)** Between 0 and 1 **e)** Exactly 1

27. As X values get larger, the corresponding Y values get smaller.
a) Exactly -1 **(b)** Between -1 and 0 **c)** About 0 **d)** Between 0 and 1 **e)** Exactly 1

28. The sum of X and its corresponding Y is always 100.
(a) Exactly -1 **b)** Between -1 and 0 **c)** About 0 **d)** Between 0 and 1 **e)** Exactly 1

29. Y is always exactly twice its corresponding X.
a) Exactly -1 **b)** Between -1 and 0 **c)** About 0 **d)** Between 0 and 1 **(e)** Exactly 1

30. X and Y values are paired by a random lottery.
a) Exactly -1 **b)** Between -1 and 0 **(c)** About 0 **d)** Between 0 and 1 **e)** Exactly 1

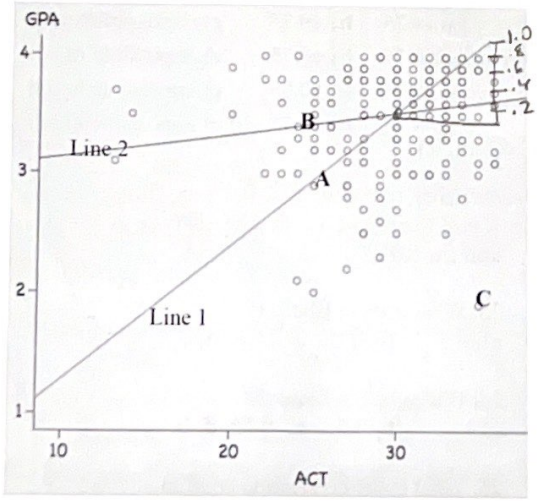
Questions 31-37 pertain to scores on Art and Math Skills tests. These follow the normal curve but have different correlations among different populations. Consider 5 populations where the correlation coefficients between are given in the table below.

Part 1: If someone is in the **84th** percentile in math skills, estimate his art skill percentile for each population.

Percentile in Art Skills	r	Percentile in Math Skills				
31. 84th	**0**	**a)** 16th	**b)** 31st	**(c)** 50th	**d)** 69th	**e)** 84th
32. 84th	**1**	**a)** 16th	**b)** 31st	**c)** 50th	**d)** 69th	**(e)** 84th
33. 84th	**-1**	**(a)** 16th	**b)** 31st	**c)** 50th	**d)** 69th	**e)** 84th
34. 84th	**0.50**	**a)** 16th	**b)** 31st	**c)** 50th	**(d)** 69th	**e)** 84th
35. 84th	**- 0.50**	**a)** 16th	**(b)** 31st	**c)** 50th	**d)** 69th	**e)** 84th

Part 2: Suppose Steve is 2 SD's below average in Art, what percentile would you predict him to be in math if **r = 0.5**?
36. What is Steve's Art Z score? **a)** 0 **b)** -1 **(c)** -2 **d)** -0.5 **e)** impossible to tell
37. What is Steve's Predicted Math Z score? **a)** 0 **(b)** -1 **c)** -2 **d)** -0.5 **e)** impossible to tell

Questions 38-42 pertain to the scatter plot below, which displays the GPA and ACT scores of 240 Stat 100 students.



38. Which line is the regression line? **a)** Line 1 **(b)** Line 2

39. The correlation (r) between ACT and GPA is closest to:
a) 0 **b)** -0.5 **c)** 0. 5 **d)** -0.2 **(e)** 0.2

40. The point of averages lies on…
a) Only Line 1 **b)** Only Line 2 **(c)** Both Lines

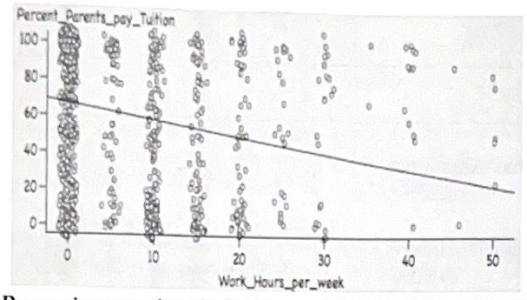
41. Look at Students A and B on the graph. Student A is on Line 1 and Student B is on Line 2. Which student has the same exact Z scores for ACT and GPA? on SD line!
(a) A **b)** B **c)** Both **d)** Neither **e)** Not enough Info

42. Look at Students A and B on the graph. Which student has a positive residual?
a) A **b)** B **c)** Both **(d)** Neither **e)** Not enough Info

residual = actual - predicted
(+) res: actual > predicted

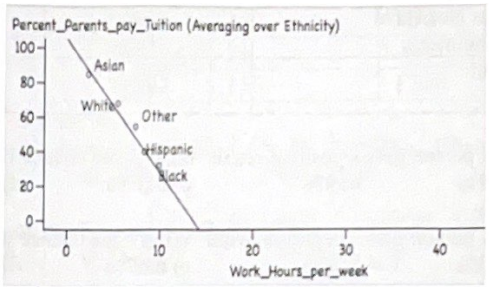
Questions 43-48 pertain to this scenario: The scatter plot on the left depicts the number of hours Stat 100 students reported working per week on the X axis and what percent of their tuition their parents paid for on the Y axis. The students also identified their ethnicity. The graph on the right divides the students into 5 groups by ethnicity and shows the average number of work hours per week and average percent tuition paid by parents for each group. The regression equations are shown below each plot.

Plot A



Regression equation: % Tuition = -0.87 (Work hours) + 67

Plot B



Regression equation: % Tuition = -7.7 (Work hours) + 105

43. Plot _____ has correlation = -0.2 **a)** Plot A **b)** Plot B

44. Plot _____ has correlation = -0.9 **a)** Plot A **b)** Plot B

Here are the RMSE's of each plot in no particular order. Which is which?

45. RMSE = 6.68 **a)** Plot A **b)** Plot B

46. RMSE = 37.82 **a)** Plot A **b)** Plot B

larger RMSE = larger "spread" around reg. line

For the next two questions, make regression estimates (predictions) using the correct equation and plot.

47. Joe works 20 hours per week, we estimate his parents pay _____% of his tuition.

a) 85 **b)** 50 **c)** 70 **d)** 20 **e)** 35

48. 62 Hispanic students responded to the survey. Their average work hours per week= 8 hours. The regression equation predicts the average percent of the tuition their parents pay is?

a) 60 **b)** 80 **c)** 100 **d)** 43 **e)** 18

Questions 49-52: The table below gives the 5 summary statistics of the students this semester who responded to the questions: "How old is your mother?" & "How old is your father?" on Bonus Survey 1.

	Average	*SD*
Mother's Age	*49*	*6*
Father's Age	*51*	*6*

Correlation Coefficient r = 0.8

Make regression estimates for Students A and B by filling in the blanks in the table below.

Mother's Age	Mother's Age Z	r	Father's Age Z	Father's Age
Student A Mother's Age=58	$\frac{58-49}{6} = 1.5$	r=0.8	1.2	Val = (1.2)(6) + 51 = 58.2
Val = (-0.4)(6) + 49 = 46.6	-0.4	r=0.8	-0.5	Student B Father's Age=48 $\frac{48-51}{6} = -0.5$

49. What is Student A's predicted Father's Age? **a)** 58 **b)** 51 **c)** 48 **d)** 62 **e)** 44

50. What is Student B's predicted Mother's Age? **a)** 50 **b)** 62 **c)** 55 **d)** 47 **e)** 42

51. Student C is one SD above average in mother's age and is on the **SD line**. What is his z-score for father's age?

a) -1 **b)** -0.8 **c)** 0 **d)** 0.8 **e)** 1

52. Student D is one SD above average in mother's age and falls on the **regression line**. What is his z-score for father's age? **a)** -1 **b)** -0.8 **c)** 0 **d)** 0.8 **e)** 1

on SD line → same z for both

Mom: z = +1

r = 0.8

dad: z = (.8)(1) = .8

Questions 53-54 pertain to this scenario. Only about 10% of all adults who participate in routine screening for Alzheimer's have the disease. The test is far from perfect. 40% of those without the disease incorrectly test positive while 10% of those with the disease incorrectly test negative. ***Fill in the following table*** for a typical sample of 100 screened adults.

	Tests Positive	Tests Negative	Total
Has Alzheimer's	9	1	10
Does Not Have Alzheimer's	36	54	90
Total	45	55	100

look at (+) column

53. A person gets a positive result. What is the chance the person has Alzheimer's? 9/45 = 0.2 → 20%.
a) 1.8% **b)** 4% **c)** 6.67% **d)** 40% **e)** 20%

look at (-) column

54. A person gets a negative result. What's the chance the adult has Alzheimer's? 1/55 = .018 → 1.8%.
a) 1.8% **b)** 4% **c)** 6.67% **d)** 80% **e)** 90%

Questions 55-60 pertain to the following situation: A large calculus class took two exams- a midterm and a final. The scatter plot of the exam scores was roughly football shaped. Here are the 5 summary statistics.

	Average	*SD*
Midterm Exam	*82*	*10*
Final Exam	*70*	*15*

***Correlation:* r = 0. 6**

$m = r \times \frac{SD_y}{SD_x} = (.6)\left(\frac{15}{10}\right) = 0.9$

55. The slope of the regression equation for ***predicting final exam scores*** from midterm exam scores is?
a) 0 **b)** 0.4 **c)** 0.9 **d)** -0.9 **e)** 1.5

56. The y-intercept of the regression equation for ***predicting final exam scores*** from midterm scores is? $y = mx + b$, $70 = .9(82) + b$, $b = -3.8$
a) 19 **b)** 7.5 **c)** 2 **d)** -3.8 **e)** -10.4

57. Use the regression equation to predict the final score of a student who scored a 92 on the midterm.
a) 79 **b)** 85 **c)** 92 **d)** 70 **e)** 68

58. Calculate the SDerrors (RMSE) when predicting final exam scores from midterm scores below. $\sqrt{1-r^2} \cdot SD_y$
a) 0.8 **b)** 12 **c)** 6 **d)** 8 **e)** -0.6

look up middle area

59. There's about a 95% chance that your estimate in part c) is right to within _____ points.
a) 1 RMSE **b)** 2 RMSEs **c)** 3 RMSEs **d)** Impossible to tell

60. Say that for one specific student, we used the regression line and got a predicted final exam score of 73. If their residual is -8, what is their actual final exam score? res = actual − predicted
a) 100 **b)** 81 **c)** 73 **d)** 65 **e)** 57

The regression equation for predicting midterm exam scores from final exam scores is:
Midterm Score= 0.4(Final Score) + 54

61. How would you interpret the ***slope*** of this regression line?
a) The best estimate of the midterm score for someone who got a 0 on the Final is 54.
b) The best estimate of the final score for someone who got a 0 on the Midterm is 54.
c) On average, for each extra point on the Final, students score about 0.4 extra points on the Midterm.
d) On average, for each extra point on the Midterm, students score about 0.4 extra points on the Final.

62. How would you interpret the ***y intercept***?
a) The best estimate of the midterm score for someone who got a 0 on the Final is 54.
b) The best estimate of the final score for someone who got a 0 on the Midterm is 54.
c) On average, for each extra point on the Final, students score about 0.4 extra points on the midterm.
d) On average, for each extra point on the Midterm, students score about 0.4 extra points on the Final.

Questions 63-70 pertain to the following data: In Fall of 2016, 1240 Stat 100 students identified their ethnicity and answered the question: "If the election were today who would you vote for?" Below are the results:

	Hillary Clinton	Donald Trump	Unsure	Other	Totals
White	234	160	139	85	618
Black	61 ★ #65	4	18	11	94
Hispanic	89	6	33	24	152
Asian	146	30	108	25	309
Other	33	9	14	11	67
Totals	563	209	312	156	1240

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

63. What is the chance that you'll get a Trump voter if you draw only from the whites? look @ white row 160/618
a) 160/209 **b)** 160/618 **c)** 209/1240 **d)** 618/1240 **e)** 234/618

64. What is the chance that you'll get a white if you draw only from the Trump voters? look @ trump column 160/209
a) 160/618 **b)** 618/1240 **c)** 209/1240 **d)** 160/209 **e)** 234/563

65. What is the chance that a randomly selected student will be both black and a Hillary voter? ★ $\frac{563}{1240} * \frac{94}{1240} - \frac{61}{1240}$
a) 94/1240 * 563/1240 **b)** 61/1240 **c)** (563*94)/1240 **d)** (563+ 94)/1240 **e)** 596/1240

66. What is the chance that a randomly selected student will be either black or a Hillary voter?
a) 94/1240 * 563/1240 **b)** 61/1240 **c)** (563*94)/1240 **d)** (563+ 94)/1240 **e)** 596/1240

67. Draw 2 students without replacement. What's the chance that at least one of them is a Trump voter? = 1-P(none)
a) 209/1240 + 208/1239 **b)** 209/1240*208/1239 **c)** 1- 209/1240*208/1239 **d)** 1031/1240*1030/1239
e) 1-1031/1240*1030/1239

68. Draw 2 students without replacement. What's the chance that none of them is a Trump voter? 1240-209= 1031 non-Trump voters
a) 209/1240 + 208/1239 **b)** 209/1240*208/1239 **c)** 1- 209/1240*208/1239 **d)** 1031/1240*1030/1239
e) 1-1031/1240*1030/1239

Questions 69-78 pertains to rolling fair dice.

69. Two dice are rolled. What is the chance that the sum of the spots is 5? 2+3, 3+2, 1+4, 4+1
a) 2/36 **b)** 3/36 **c)** 4/36 **d)** 5/36 **e)** 1/6*1/6

70. Two dice are rolled what's the chance of getting a 2 on the first roll **and** a 3 on the second roll? $\frac{1}{6} \times \frac{1}{6}$
a) 1/36 **b)** 2/36 **c)** 4/36 **d)** 5/36 **e)** 11/36

71. Two dice are rolled what's the chance of getting 2 on the first roll **or** a 3 on the second roll? $\frac{1}{6} + \frac{1}{6} - (\frac{1}{6} \cdot \frac{1}{6})$
a) 1/36 **b)** 2/36 **c)** 4/36 **d)** 11/36 **e)** 12/36

72. One die is rolled 3 times. What is the chance of getting all 6's? $\frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$
a) (5/6)^3 **b)** (1/6)^3 **c)** 1- (5/6)^3 **d)** 1- (1/6)^3 **e)** 3/6

73. One die is rolled 3 times. What is the chance of not getting all 6's? 1-P(all)
a) (5/6)^3 **b)** (1/6)^3 **c)** 1- (5/6)^3 **d)** 1- (1/6)^3 **e)** 3/6

74. One die is rolled 3 times. What is the chance of getting no 6's? $\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}$
a) (5/6)^3 **b)** (1/6)^3 **c)** 1- (5/6)^3 **d)** 1- (1/6)^3 **e)** 3/6

75. One die is rolled 3 times. What is the chance of getting at least one 6? 1 - P(none)
a) (5/6)^3 **b)** (1/6)^3 **c)** 1- (5/6)^3 **d)** 1- (1/6)^3 **e)** 3/6

76. Would the answers to 72-75 change if we replaced "One die is rolled 3 times" with "3 dice are rolled" and kept all else the same?
a) All answers would change. **b)** Some answers would change. **c)** 1 answer would change. **d)** None would change.

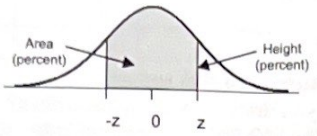
Exam 2 Formulas:

IQR=Q3-Q1
Low outliers < Q1 – 1.5*IQR
High outliers > Q3 + 1.5*IQR

Slope of Regression Line = r*SDy/SDx
RMSE = sqrt(1-r^2) * SDy
P(at least one) = 1 - P(none)

P(not all) = 1 - P(all)
z=(value-average)/SD

STANDARD NORMAL TABLE



Standard Units

z	*Area*		*z*	*Area*		*z*	*Area*
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