Stat 100 Exam 2 Key #1 - Fall 2016 Fall 2016

PRINT NAME _____ _____ _____
(Last name) (First name) (netID)

CIRCLE SECTION L1 12:30pm L2 3:30 pm MWF at noon Online

Write answers in appropriate blanks. When no blanks are provided **CIRCLE** your answers.
SHOW WORK when requested.

No notes or books are allowed. Calculators (including graphing ones) are allowed.
Do not use your own scrap paper. If you need some, ask a proctor.

Make sure you have all 8 pages including the normal table (16 problems).

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 _____ 12 points

Page 2 _____ 14 points

Page 3 _____ 22 points

Page 4 _____ 14 points

Page 5 _____ 14 points

Page 6 _____ 10 points

Page 7 _____ 14 points

L1 (T/R 12:30 Class)
Last names: A-L
Foellinger Floor
(157 exams)

Total Score _____

There is NO class on Thursday!
Scores will be posted on Compass Thurs. night and available in 23 Illini Hall during lab hours on Monday.
Remember: The deadline to drop a class is this Friday, Oct 14 at 5pm

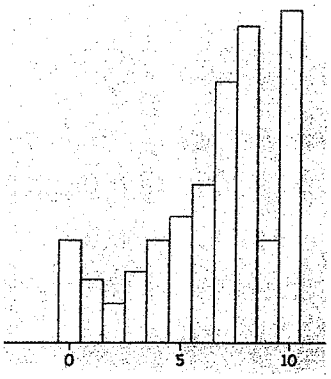
Question 1 *(6 pts)*
Below is a distribution table for the scores on a Physics exam for a large class. The right-hand column shows the % of scores in each interval. **The lowest score was 0, the second lowest was 6, and the highest score was 100.** To draw a box plot of the data you'd have to find the median, Q1 and Q3.

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

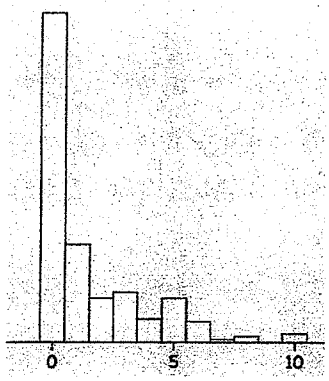
a) **Median=** 70
b) **Q1 =** 50
c) **Q3=** 80
d) **The middle 50% of the scores lie between** 50 **and** 80. ← continued error give credit for whatever they put as Q1 + Q3
e) **List all outliers** 0

Question 2 *(3 pts)* Which histograms correspond to which box plots?

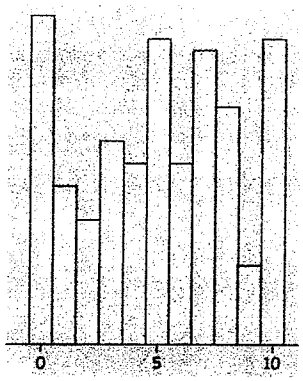
Histogram A



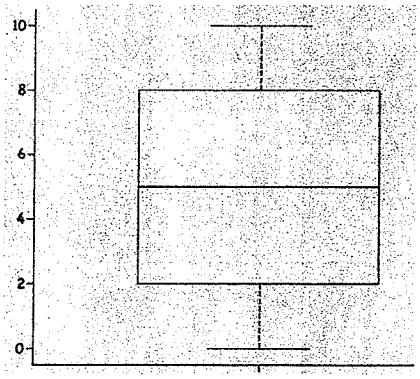
Histogram B



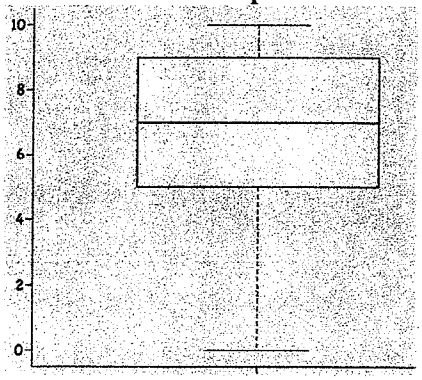
Histogram C



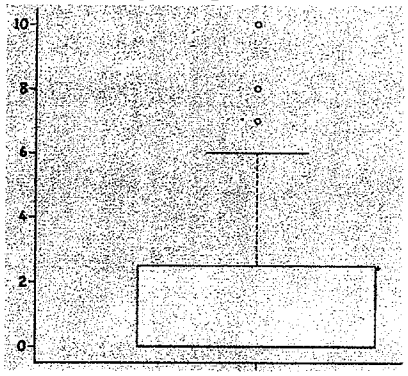
Write the correct letter in each blank below to match the box plots with their corresponding histograms.



i) C

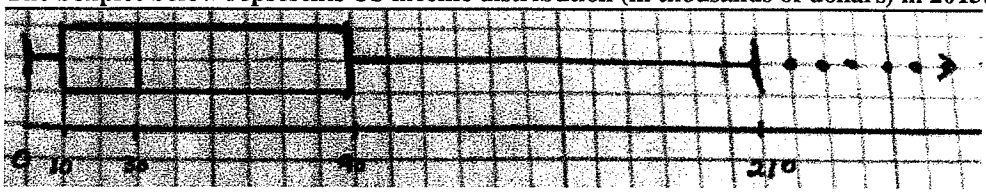


ii) A



iii) B

Question 3 *(3 pts)*
The boxplot below represents US income distribution (in thousands of dollars) in 2015.



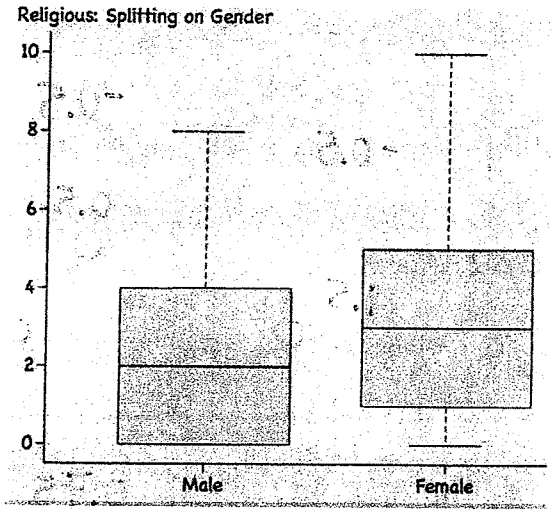
a) What percent of the population earned less than $10,000? 25%
b) What percent of the population earned between $10,000 and $30,000? 25%
c) What percent of the population earned between $30,000 and $90,000? 25%

Question 4 *(6 pts)*
The 2 boxplots below depict the Stat 200 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)

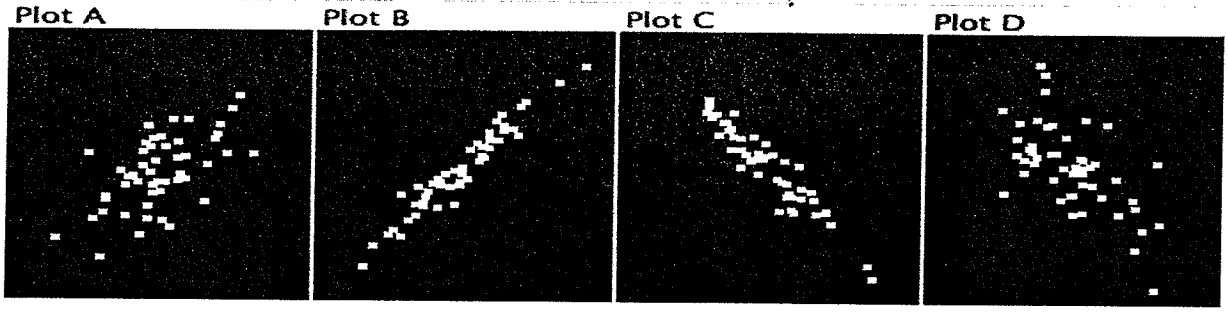
a) Fill in the 12 blanks in the table below.
All answers are whole numbers.

	Males	Females
Average	2.46	3.15
SD	2.38	2.56
Min	0	0
Q1	0	1
Med	2	3
Q3	4	5
Max	8	10
IQR	4	4
n	140	100

½ pt for each cell



Question 5 *(4 pts)* **pertains to the 4 scatter plots below:**



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

i) r = - 0.95 C *ii)* r = - 0.62 D *iii)* r = 0.6 A *iv)* r = 0.97 B

Question 6 *(4 pts)* X and Y are 2 sets of numbers with a correlation coefficient of **r = - 0.75.** Circle the correct answer.

a.	If all the original x and y values are multiplied by 4, r would be?	**i)** -0.75	**ii)** +0.75	**iii)** impossible to tell
b.	If all the original x values are multiplied by ***negative*** 0.6, r would be?	**i)** -0.75	**ii)** +0.75	**iii)** impossible to tell
c.	If 100 is added to only the first x value, r would be?	**i)** -0.75	**ii)** +0.75	**iii)** impossible to tell
d.	If all the original X and Y values are switched, r would be?	**i)** -0.75	**ii)** +0.75	**iii)** impossible to tell

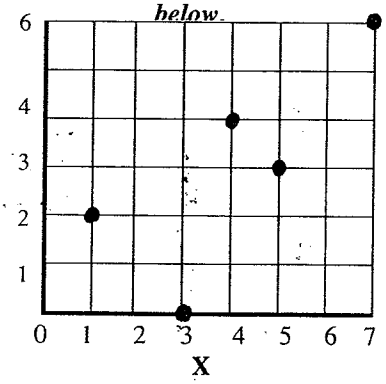
Question 7 *(8 pts)*

a) Compute the correlation coefficient (**r**) between X and Y by ***filling in the table below***. A few cells have been done for you.
The <u>average of X = 4</u> and the SD of X = 2. The <u>average of Y = 3</u> and the SD of Y = 2.

don't forget

X	Y	Z score for X	Z score for Y	Products
1	2	-1.5	-0.5	0.75
3	0	-0.5	-1.5	0.75
4	4	0	0.5	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

(1 pt) Plot the 5 points (<u>NOT</u> the Z scores) below.



b) (0.5 pt) The correlation coefficient **r** = 0.75 *(Fill in the blank.)*

$r = 3.75/5 = 0.75$

Question 8 (6 pts.) For each of the following pairs of variables. Check the box under the column heading that best describes its correlation. (**Hint:** Every column should have exactly one box checked.)

	Correlation	Exactly -1	Between -1 and 0	About 0	Between 0 and 1	Exactly +1	Not enough Info
a)	As X values get larger, the corresponding Y values also get larger.	☐	☐	☐	☑	☐	☐
b)	As X values get larger, the corresponding Y values get smaller.	☐	☑	☐	☐	☐	☐
c)	The sum of X and its corresponding Y is always 100.	☑	☐	☐	☐	☐	☐
d)	Y is always exactly twice its corresponding X.	☐	☐	☐	☐	☑	☐
e)	X and Y values are paired by a random lottery.	☐	☐	☑	☐	☐	☐
f)	Y is always smaller than its corresponding X.	☐	☐	☐	☐	☐	☑

Question 9 *(8 pts)* Suppose scores on Art and Math Skills tests follow the normal curve but have different correlations among different populations.

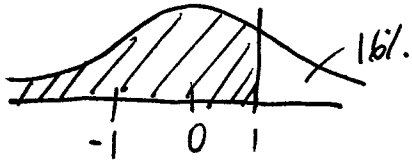
a) Consider 5 populations where the correlation coefficients between are given in the table below. If someone is in the **84th** percentile in math skills, estimate his art skill percentile for each population. (5 pts.)

Percentile in Art Skills	r	Percentile in Math Skills
84th	**0**	*Choose One:* 8th 16th 31st **50th** 69th 84th 92nd
84th	**1**	*Choose One:* 8th 16th 31st 50th 69th **84th** 92nd
84th	**-1**	*Choose One:* 8th **16th** 31st 50th 69th 84th 92nd
84th	**0.50**	*Choose One:* 8th 16th 31st 50th **69th** 84th 92nd
84th	**- 0.50**	*Choose One:* 8th 16th **31st** 50th 69th 84th 92nd

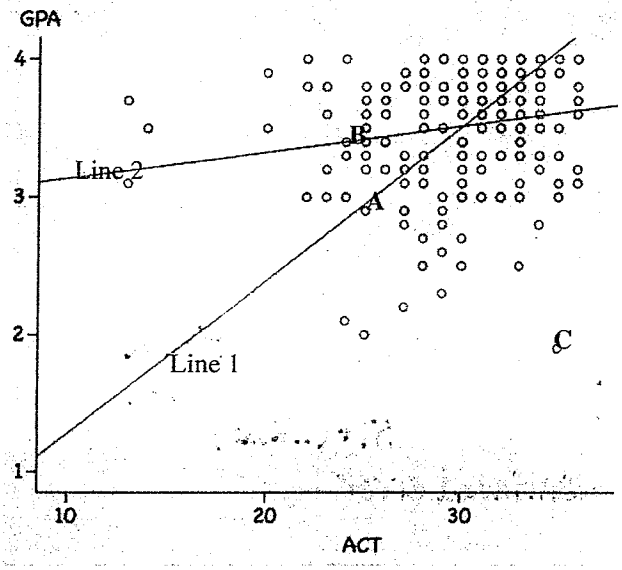
b) *(3 pts)* Suppose someone is 2 SD's above average in Art, what percentile would you predict them to be in math if **r = 0.5**?

check for cont. error

i) Art Z score= 2 ii) Predicted Math Z score = 1 iii) Predicted Math percentile= 84th (Hint: Draw Normal Curve)



Question 10 *(6 pts)* pertains to the scatter plot below, which displays the GPA and ACT scores of 240 Stat 200 students.



a) 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) The correlation (r) between ACT and GPA is closest to: *Choose one:*
i) 0 ii) -0.5 iii) 0. 5 iv) -0.2 (v) 0.2

c) The point of averages lies on…*Choose one:*
i) Only Line 1 ii) Only Line 2 (iii) Both Lines iv) Neither Line

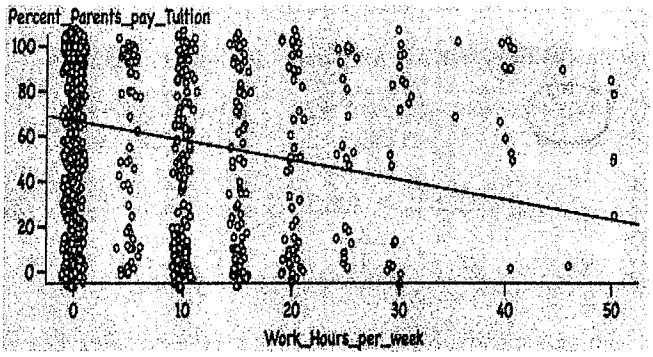
d) Look at Students A and B on the graph. Student A is on Line 1 and Student B is on Line 2. Which student has the same exact Z scores for ACT and GPA? *Choose one:*
(i) A ii) B iii) Both iv) Neither v) Not enough Info

e) Look at Students A and B on the graph. Which student has a positive residual? *Choose one:*
i) A ii) B iii) Both (iv) Neither v) Not enough Info

f) If student C was removed r would …
(i) increase ii) decrease iii) stay the same

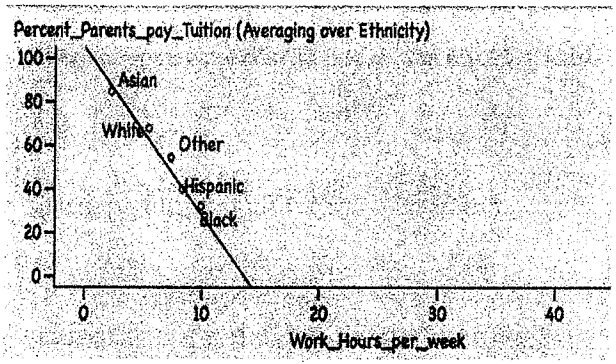
Question 11 *(8 pts)*
The scatter plot on the left depicts the number of hours 1138 Stat 100 students reported working per week on the X axis and what percen of their tuition their parents paid for on the Y axis. The students also identified their ethnicity. The graph on the right divides the 1138 students 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.

Plot A



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

Plot B



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

For a) and b) identify the correct plot by writing either A or B in the 4 blanks below.

a) (2 pts) Plot A has correlation = -0.2 and Plot B has correlation = 0.9

b) (2 pts) Here are the RMSE's of each plot in no particular order. Which is which?
i) RMSE = 6.68 **Circle one:** Plot A (Plot B)
ii) RMSE = 37.82 **Circle one:** (Plot A) Plot B

For c-e, make regression estimates (predictions) using the correct equation (given below each plot). Fill in the first blank with the prediction and the second with the correct RMSE (6.68 or 37.82). (Round prediction to 2 decimal places.)

c) (1 pt) Joe works 20 hours per week, we estimate his parents pay 49.6 % of his tuition, give or take about 37.82 %.
½ pt ½ pt

d) (1 pt) Michael works 30 hours per week and is Hispanic, we estimate that his parents pay 40.9 % of his tuition, give or take about 37.82 %.
½ pt ½ pt

e) (2pts.) 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 43.4 %, give or take about 6.68 %
1 pt 1 pt

Question 12 *(8 pts)*
The table below gives the 5 summary statistics of the students this semester who responded to the questions: "How old is your mother?" and "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

a) *(6 pts)* Make regression estimates for Students A and B by filling in the blanks in the table below. For Student A, you're asked to predict father's age and for Student B, you're asked to predict mother's age.
Show work converting values to z scores and z scores to values as indicated. Don't round answers.

Mother's Age	**Mother's Age Z Score**	**r**	**Father's Age Z Score**	**Father's Age**
Student A Mother's Age= 58 $\frac{58-49}{6} = 1.5$	Z= 1.5 ✗ Show work.	r = 0.8 =	Z = 1.2 continued error	Father's Age = 58.2 Show work. 51 + (1.2)(6)
Mother's Age= 46.6 Show work. 49 + (-0.4)(6) continued error	Z= -0.4 =	r = 0.8 ×	Z= -0.5 Show work.	*Student B* Father's Age= 48 $\frac{48-51}{6} = -0.5$

b) *(1 pt) 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?

i) -1 **ii)** -0.8 **iii)** -0.6 **iv)** 0 **v)** 0.6 **vi)** 0.8 **(vii) 1** **viii)** $\sqrt{1-0.6^2} * 6$ **ix)** $\sqrt{1-0.8^2} * 6$

c) *(1 pt) 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?

i) -1 **ii)** -0.8 **iii)** -0.6 **iv)** 0 **v)** 0.6 **(vi) 0.8** **vii)** 1 **viii)** $\sqrt{1-0.6^2} * 6$ **ix)** $\sqrt{1-0.8^2} * 6$

Question 13 *(6 pts)* 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. (4 pts) ½ pt for each cell

	Tests Positive	**Tests Negative**	**Total**
Has Alzheimer's	9	1	10
Does NOT have Alzheimer's	36	54	90
Total	45	55	**100**

a) (1 pt) A person gets a positive result. What is the chance the person has Alzheimer's? _____% $\frac{9}{45}$
Choose one: **i)** 1.8% **ii)** 4% **iii)** 6.67% **(iv) 20%** **v)** 40% **vi)** 80% **vii)** 90%

b) (1 pt) A person gets a negative result. What's the chance the adult has Alzheimer's?_____% $\frac{1}{55}$
Choose one: **(i) 1.8%** **ii)** 4% **iii)** 6.67% **iv)** 20% **v)** 40% **vi)** 80% **vii)** 90%

Question 14 *(10 pts)* 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**

a) (2 pts) The slope of the regression equation for **predicting *final exam scores*** from midterm exam scores is? ***Show work.***

$m = r \times \frac{SD_y}{SD_x} = 0.6 \times \frac{15}{10} = 0.9$

b) (1 pt) The y-intercept of the regression equation for ***predicting final exam scores*** from midterm scores is? ***Show work.***

$y = 0.9x + b$

$70 = 0.9(82) + b$

$b = -3.8$

c) (1 pt) Use the regression equation to predict the final score of a student who scored a 92 on the midterm.

$y = 0.9(92) - 3.8 = 79$

d) (1 pt) Check to make sure your regression equation is correct by repeating part c using the 3 step process like we did in 12a (converting to z scores). Start with a midterm score of 92 and make a prediction for their final. You should get the same answer as in part c. **Show work.**

$92 \quad \frac{92-82}{10} = 1 \times 0.6 = 0.6 \quad 70 + (0.6)(15) = 79$ ✓

e) (1 pt) Calculate the SDerrors (RMSE) when predicting final exam scores from midterm scores below. **Show work.**

$RMSE = \sqrt{1-r^2} \times SD_y = \sqrt{1-0.6^2} \times 15 = 12$

f) (1 pt) There's about a 95% chance that your estimate in part c) is right to within _____ points. ***Circle one:***

i) 1 RMSE **ii)** 2 RMSEs **iii)** 3 RMSEs **iv)** Impossible to tell

g) (1 pt) 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? ***Show work below!***

$-8 = \text{actual} - 73 \qquad \text{actual} = 65$

h) (2 pts) The regression equation for predicting midterm exam scores from final exam scores is:
Midterm Score= 0.4(Final Score) + 54

Part 1: (1 pt) How would you interpret the ***slope*** of this regression line?

i) The best estimate of the midterm score for someone who got a 0 on the Final is 54.
ii) The best estimate of the final score for someone who got a 0 on the Midterm is 54.
iii) On average, for each extra point on the Final, students score about 0.4 extra points on the Midterm.
iv) On average, for each extra point on the Midterm, students score about 0.4 extra points on the Final.

Part 2: (1 pt) How would you interpret the ***y intercept***?

i) The best estimate of the midterm score for someone who got a 0 on the Final is 54.
ii) The best estimate of the final score for someone who got a 0 on the Midterm is 54.
iii) On average, for each extra point on the Final, students score about 0.4 extra points on the midterm.
iv) On average, for each extra point on the Midterm, students score about 0.4 extra points on the Final.

Question 15 *(6 pts)* On our survey, 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	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**:

a) What is the chance that you'll get a Trump voter if you draw only from the whites? (leave answer as a fraction.)

160/618

b) What is the chance that you'll get a white if you draw only from the Trump voters? (leave answer as a fraction.)

160/209

c) What is the chance that a randomly selected student will be both black and a Hillary voter?
i) 94/1240 * 563/1240 ii) 61/1240 iii) (563*94)/1240 iv) (563+ 94)/1240 v) 596/1240 vi) (563+ 61)/1240

d) What is the chance that a randomly selected student will be either black or a Hillary voter?
i) 94/1240 * 563/1240 ii) 61/1240 iii) (563*94)/1240 iv) (563+ 94)/1240 v) 596/1240 vi) (563+ 61)/1240

e) Draw 2 students without replacement. What's the chance that at least one of them is a Trump voter?

i) 209/1240 + 208/1239 **ii)** 209/1240*208/1239 **iii)** 1- 209/1240*208/1239 **iv)** 1031/1240*1030/1239 **v)** 1-1031/1240*1030/1239

f) Draw 2 students without replacement. What's the chance that none of them is a Trump voter?

i) 209/1240 + 208/1239 **ii)** 209/1240*208/1239 **iii)** 1- 209/1240*208/1239 **iv)** 1031/1240*1030/1239 **v)** 1-1031/1240*1030/1239

Question 16 *(8 pts)* **pertains to rolling fair dice.**

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

b) Two dice are rolled what's the chance of getting a 2 on the first roll **and** a 3 on the second roll?
i) 1/36 ii) 2/36 iii) 4/36 iv) 5/36 v) 11/36 v) 12/36

c) Two dice are rolled what's the chance of getting 2 on the first roll **or** a 3 on the second roll?
i) 1/36 ii) 2/36 iii) 4/36 iv) 5/36 v) 11/36 v) 12/36

d) One die is rolled 3 times. What is the chance of 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

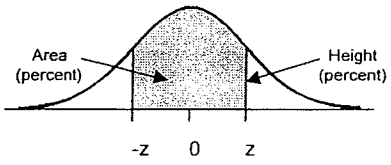
e) One die is rolled 3 times. What is the chance of not getting all 6's?
i) $(5/6)^3$ ii) $(1/6)^3$ iii) $1-(5/6)^3$ iv) $1-(1/6)^3$ v) 3/6

f) 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

g) 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

h) Would the answers to d-e change if we replaced "One die is rolled 3 times." with "3 dice are rolled" and kept all else the same? ***Choose one.***
i) All answers would change. ii) Some answers would change iii) 1 answer would change. iv) None would change.

STANDARD NORMAL TABLE



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