Stat 100 Exam 2
STAT 100 EXAM 2 FALL 2018 17 Key

October 17, 2018

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 (TR at 12:30) **ii)** L2 (MWF at noon) **iii)** ONLINE

Write answers in appropriate blanks. When no blanks are provided **CIRCLE** your answers.

******WARNING: When we say "NO WORK, NO CREDIT", we mean it. You'll get a 0. *****

Do NOT use scrap paper.

Make sure you have all 7 pages including the Normal table (17 problems).

For questions using the normal table, you may "round" z scores and percents to fit the closest line on the normal table and you may round percents on the table to the nearest whole number.

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 _____13

Page 2 _____18

Page 3 _____18

Page 4 _____20

Page 5 _____15

Page 6 _____16

Cover _____1 *for answering cover page questions 1-3 correctly!*

Total Score _____

WARNING- The exams look alike but you are sitting next to people who actually have different version. 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.

There is NO CLASS on Thursday or Friday! Scores will be posted on Compass by Friday morning (the drop deadline) and exams returned in class next week.

Online students may pick up their exam in 23 Illini Hall during office hours next week.

Stat 100 Exam 2 Page 1- 17 October 17, 2018

Question 1 (8 pts.)

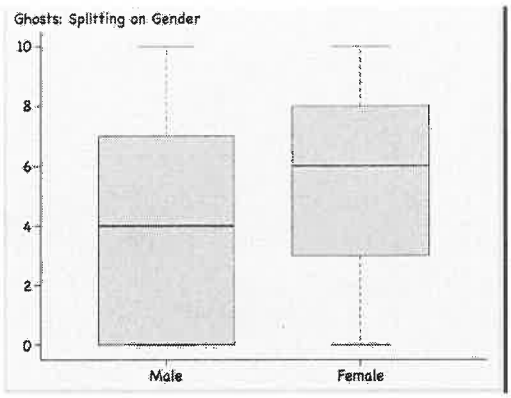
The 2 box plots below depict the survey responses of 365 males and 765 females to the question: "On a scale of 0 to 10, rate how strongly you believe in ghosts. (0 is not at all and 10 is extremely)

a) (6 pts.) Fill in the 12 blanks in the table below. ½ pt for each blank

All answers are whole numbers.

	Males	Females
Average	4.135	5.417
SD	3.515	3.381
Min	0	0
Q1	0	3
Med	4	6
Q3	7	8
Max	10	10
IQR	7	5
n	365	765

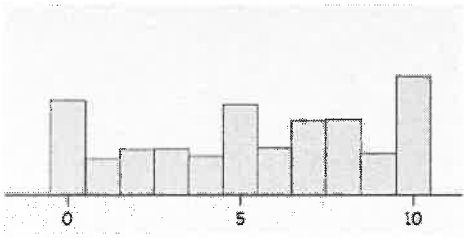
Q3-Q1



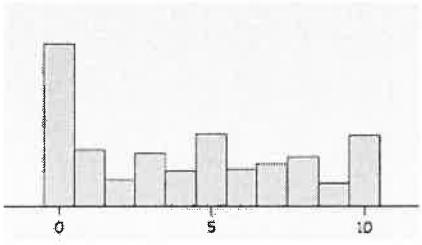
b) (1 pt.) Are there any outliers for males or females? i) Yes, only for males ii) Yes, only for females iii) Yes, for both (iv) No

c) (2 pts.) The histograms below depict the ghost ratings of the 365 males and 765 females. Which histogram depicts the male responses? Choose one: i) A (ii) B

A

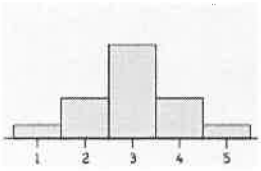


B

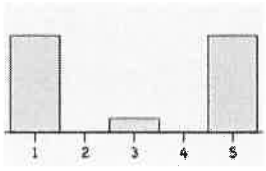


Question 2 (4 pts.) Which histograms correspond to which box plots?

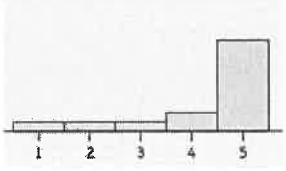
Histogram A



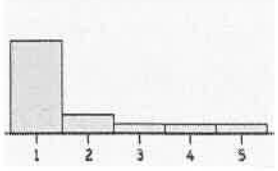
Histogram B



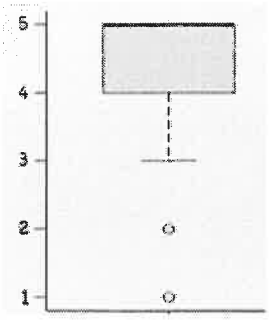
Histogram C



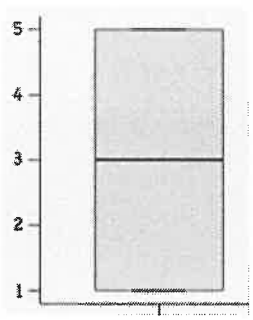
Histogram D



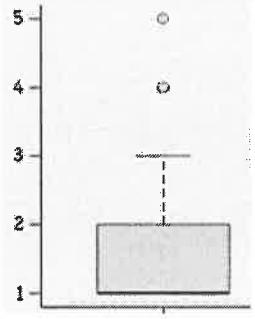
Write the correct letter in each blank below to match the box plots with their corresponding histograms.



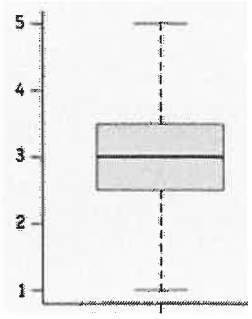
i) C



ii) B



iii) D



iv) A

Stat 100 Exam 2 October 17, 2018

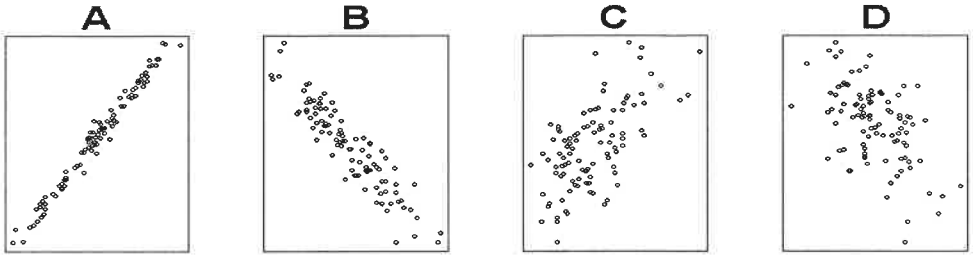
Question 3 (3 pts.) **pertains to the 4 scatter plots below:**

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

i) r = 0.99 A

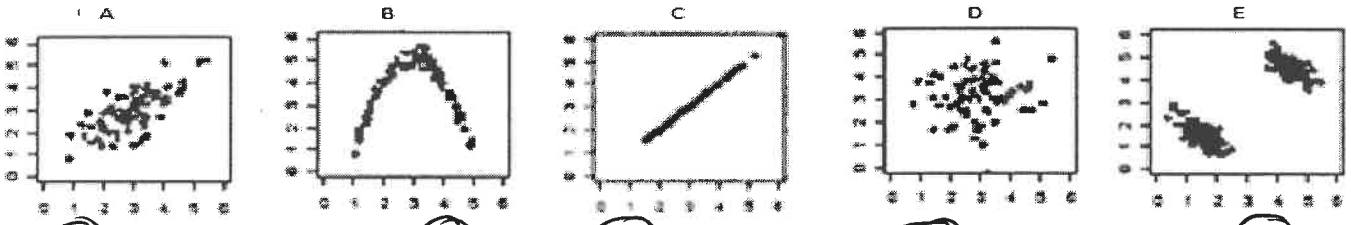
ii) r = -0.9 B

iii) r = 0.48 C



Question 4 (5 pts.)

Is r an appropriate statistic to use in the plots below? Under each plot circle YES if r is appropriate or NO if it's not.



a) (YES) NO b) YES (NO) c) (YES) NO d) (YES) NO e) YES (NO)

Question 5 (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	Exactly 0	Between 0 and -1	Exactly −1	Not Enough Information
X and Y are 2 sets of numbers with the same average and the same SD.	☐	☐	☐	☐	☐	☑
The slope of the regression line for predicting Y from X is 0. (Assume SD's ≠0)	☐	☐	☑	☐	☐	☐
The regression line has a negative slope and a RMSE = 0	☐	☐	☐	☐	☑	☐
Y is ***always exactly*** 0.5 times X	☑	☐	☐	☐	☐	☐
Years of education and income among US adult men.	☐	☑	☐	☐	☐	☐
Temperature and heating bill	☐	☐	☐	☑	☐	☐

Question 6 (4 pts.) X and Y are 2 sets of numbers with a correlation coefficient of **r = 0. 3**

Fill in the 4 blanks below with ***numbers*** (**NOT** with phrases like "increase" or "stay the same".) (4 pts.)

a. If X and Y are converted to Z scores the new correlation coefficient would be = 0.3 (***write a number for r in a-d)***

b. If all the original X values are multiplied by ***negative*** -2, the new correlation coefficient would be = -0.3.

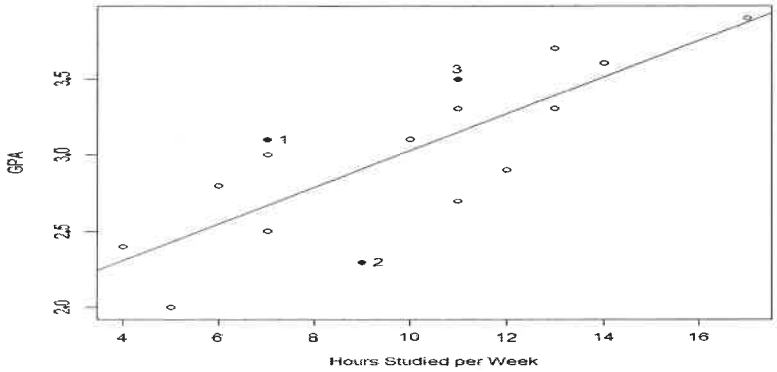
c. If all the original X values are increased by 10%, the new correlation coefficient would be = 0.3.

d. If all the X and Y values are switched, the new correlation coefficient would be = 0.3.

Stat 100 Exam 2 Page 3-17 October 17, 2018

Question 7 (14 pts.)

The scatter plot below shows the GPA and hours studied per week for 16 students. The line shown is the regression line



	Average	*SD*
GPA	*3.0*	*0.5*
Study Hours	*10*	*4*

Correlation: r = 0.8

a) (2 pts.) Look at students **1 and 2** on the graph. How did their actual GPA's compare to their predicted scores? For each student circle whether their actual GPA was better than, worse than, or the same as the regression line predicted from how many hours per week they studied.

i. Student 1 actually did _____than predicted. ***Choose one:*** **a)** Better **b)** Worse **c)** Same as

ii. Student 2 actually did _____than predicted. ***Choose one:*** **a)** Better **b)** Worse **c)** Same as

b) (6 pts.) In the table below you are given the study hours of 2 students. For each, compute the regression estimate for their GPA's, by filling out the chart below. ***No work, no credit and no partial credit for mistakes since you can check your answer with the graph.***

# Hours Studied	Study Z score	r	GPA Z score	GPA
i) 14 hours **Show work to get Z →**	Z= 1 $Z = \frac{14-10}{4} = 1$	× r = 0.8 =	Z = 0.8 **Show work to get GPA →**	GPA = 3.4 val = 3.0 + (0.8)(0.5)
ii) 6 hours **Show work to get Z →**	Z= -1 $Z = \frac{6-10}{4} = -1$	× r = 0.8 =	Z = -0.8 **Show work to get GPA →**	GPA = 2.6 val = 3.0 + (-0.8)(0.5)

c) (4 pts.) **The Regression equation for predicting GPA from hours studied is:** GPA = 0.1 x Study Hours + 2

i) (2 pts.)Fill in the first blank in the equation above with the slope. **Show work.** $m = r \times \frac{SD_y}{SD_x} = 0.8 \times \frac{0.5}{4} = 0.1$

ii) (2 pts.) Fill in the second blank in the equation above with the y- intercept. ***Show your work!***

$y = 0.1x + b$ plug in averages: $3 = 0.1(10) + b$ $b = 2$

(Hint: Check your arithmetic by using the regression equation and make sure it agrees with your estimates in part (b).)

No partial credit will be awarded for any part of this problem since you should be able to catch your own mistakes.

d) (2 pts.) In part (c) above you were predicting GPA from study hours. Would the slope, and y-intercept change if you were predicting study hours from GPA instead?

i) The slope would… ***Choose one:*** change stay the same cannot be determined

ii) The y-intercept would ***Choose one:*** change stay the same cannot be determined

Question 8 (4 pts.) A set of exam scores follows a **normal distribution**. In the table below, you are given either the Z score or the percentile for 4 students scores. Fill in the missing blanks.

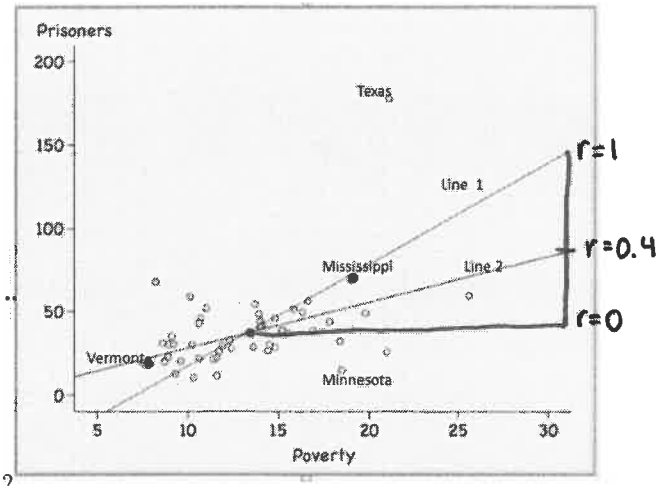
Student	Exam Percentile	Exam Z score
A	50th	0
B	16th	-1

Student	Exam Percentile	Exam Z score
C	4th	-1.75
D	99.5th	2.5

> accept rounded & not rounded answers

Question 9 (14 pts.)

The scatter plot below shows the percent of people living in poverty versus the number of people in prison per 100,000 for the 50 states in the US. **Vermont is on Line 2 and Mississippi is on Line 1.**



a) The average poverty level (in %) is closest to…
i) 10 **ii)** 13 **iii)** 16 **iv)** 18 **v)**20

b) The average number of prisoners (per 100,000) is closet to….
i) 10 **ii)** 40 **iii)** 60 **iv)** 75 **v)**100

c) (2 pts.) Which is the regression line? **i)** Line 1 **ii)** Line 2

d) (2 pts.) The correlation between poverty and prisoners is closest to ..
i) 0 **ii)** -0.4 **iii)** 0.4 **iv)** 0.8 **v)** -0.8 **vi)** 1

e) Which of the labeled states has a residual = 0? on reg line
i) Vermont **ii)** Mississippi **iii)** Texas **iv)** Minnesota

f) (2 pts.)Which state has the same z scores for poverty and prisoners?
i) Vermont **ii)** Mississippi **iii)** Texas **iv)** Minnesota on SD line

g) Which state has the largest prediction error?
i) Vermont **ii)** Mississippi **iii)** Texas **iv)** Minnesota

h) Looking at the scatter plot, about how many prisoners (per 100,000) would we predict for a poverty level of 18?
i) 20 **ii)** 35 **iii)** 50 **iv)** 65 **v)** 100

i) The average of all the residuals is 0 (Fill in blank with a number.) always!

j) (2 pts.) The above graph has 50 points, one for each **individual** state's poverty and prisoner level. If we divided the 50 states into 9 geographical regions and calculated the **average** poverty and prisoner level within each region we'd condense the 50 points into 9 points. Would the correlation coefficient for the 9 points be the same as for the 50 points? ***Choose one:***

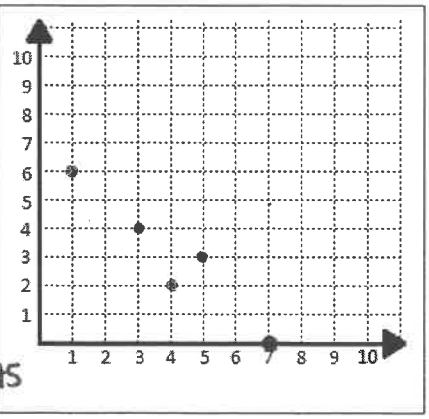
i) Yes, it would be exactly the same since it's the same information.
ii) No, it would probably be lower since we have fewer points.
iii) No, it would probably be higher since the within region scatter would disappear and we'd just see the between region scatter.

Question 10 (6 pts.)

Part A: Fill in the table and plot the points. The **average** of **X = 4** and the average of **Y = 3**. The **SD** of **X** and **Y** are **both 2**. (NOTE: X and Y have DIFFERENT averages)

X	Y	Z-score for X	Z-score for Y	Products
1	6	-1.5	$z = \frac{6-3}{2} = 1.5$	-2.25
3	4	$z = \frac{3-4}{2} = -0.5$	0.5	-0.25
4	2	0	$z = \frac{2-3}{2} = -0.5$	0
5	3	$z = \frac{5-4}{2} = 0.5$	0	0
7	0	1.5	-1.5	-2.25
Totals		Total *should* = 0	Total *should* = 0	Total = -4.75

½ pt for each blank except total for last column (no pts).



Part B: Use the total you got above to find the correlation coefficient. **r** = -0.95 (1pt)

r = avg of products = $\frac{-4.75}{5} = -0.95$

Stat 100 Exam 2 October 17, 2018

Question 11 (7 pts.) Suppose Verbal SAT and Math SAT scores among students who take both exams have the following summary statistics and the scatter plot is football shaped: **Verbal SAT: avg = 500, SD=100 Math SAT: avg = 500, SD=100 r =0.8**

a) (1 pt.) The regression equation when predicting Math scores from Verbal scores is: **Predicted Math = 0.8 * Verbal + 100**
Use the regression equation to predict the Math score of a student who got a 640 on the Verbal. **No work no credit.**

$Y = 0.8(640) + 100 = 612$ Predicted Math SAT= 612

b) (2 pts.) Now, predict the Math score of a student who got a 640 on the Verbal using the 3 step process.
(No partial credit for because you can check your work) **No work no credit.**

verbal 640 $z_v = \frac{640-500}{100} = 1.4$ × r 0.8 = z_m 1.12 math $val = 500 + 1.12(100) = 612$ ✓ Predicted Math SAT= 612

c) (2 pts.) What is the SD of the prediction errors (the RMSE) when predicting Math SAT scores from ACT scores?
i) 10 **ii)** 60 **iii)** 68 **iv)** 80 **v)** 100

$RMSE = \sqrt{1-r^2} \times SD_Y = \sqrt{1-.8^2} \times 100 = 60$

d) (2 pts.) The regression equation predicts Math SAT scores of 500 for those who score 500 on the Verbal SAT. Of course not all will get exactly 500. Instead there's a range of scores, with about **95%** of them scoring between …
Choose one: **i)** 300 and 700 **ii)** 400 and 600 **iii)** 380 and 620 **iv)** 340 and 660

prediction ± 2RMSEs
$500 \pm 2(60) = (380, 620)$

Question 12 (8 pts.)
Suppose blood pressure and temperature follow the normal curve but have different correlations among different populations.
For questions a-e, fill in **each of the 7 blanks** with the correct **numbers**.

a) (1 pt.) Imagine a population where there is a perfect positive correlation (r = 1) between the 2 measurements. If someone's blood pressure is in the 70th percentile then the regression estimate for his temperature would be the 70 th percentile. r=1 ⇒ exactly equal

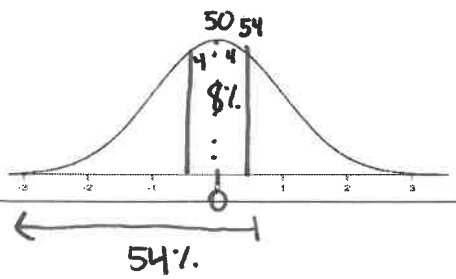
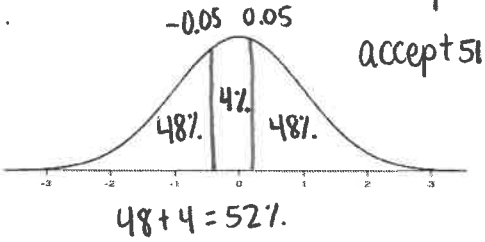
b) (1 pt.) Imagine a population where there is a perfect negative correlation (r = -1) between the 2 measurements. If someone's blood pressure is in the 70th percentile then the regression estimate for his temperature would be the 30 th percentile. r=-1 ⇒ add to 100

c) (1 pt.) Imagine a population where there is a zero correlation (r = 0) between the 2 measurements. If someone's blood pressure is in the 70th percentile then the regression estimate for his temperature would be the 50 th percentile. r=0 ⇒ always 50th

d) (1 pt.) Imagine a population where there is a **positive** correlation between the 2 measurements. If someone's blood pressure is in the 70th percentile then the regression estimate for his temperature would be > 50 th percentile and ≤ 70 th percentile.

e) (1 pt.) Imagine a population where there is a **negative** correlation between the 2 measurements. If someone's blood pressure is in the 70th percentile then the regression estimate for his temperature would be > 30 th percentile and ≤ 50 th percentile.

f) (3 pts.) If someone's blood pressure is in the **54th** percentile where r = 0.5, estimate his temperature percentile ***by filling in the table below.***
You may round areas and z-scores to fit the nearest line on the table.
This problem is either 0 or 3 pts. Must fill in all 4 blanks correctly and mark Z correctly on both graphs.

Blood Pressure Percentile	Blood Pressure Z	r	Temperature Z	Temperature Percentile
Person is in the 54th percentile for blood pressure. What middle area on the Normal Curve does that correspond to? 8 % 1pt accept 7.97%. Correctly mark the graph and write the correct Z in the next column.  54%.	Z = 0.10 1pt	× r=0.5	= Z = 0.05	(1 pt) Temperature Percentile = 52nd Mark the Z score on the graph below. 1pt  accept 51.995th 48+4 = 52%.

Question 15 (8 pts.) pertains to the table below that shows our survey responses for gender and handedness.

	Left-Handed	Ambidextrous	Right-Handed	Totals
Male	30	20	266	316
Female	60	27	560	647
Totals	90	47	826	963

Suppose you randomly draw from these students.

a) What is the chance of getting a female?
i) 647/963 ii) 90/963 iii) 316/963 iv) 737/963 v) 826/963 vi) 1142/963

b) What is the chance of getting someone who is left-handed?
i) 30/90 ii) 90/963 iii) 62/500 iv) 63/500 v) 826/963 vi) 266/826 vii) 37/65

c) What is the chance of drawing once and getting a female **or** someone who is left-handed? $\frac{647}{963} + \frac{90}{963} - \frac{60}{963} = \frac{677}{963}$
i) 60/963 ii) 737/963 iii) 677/963 iv) 30/963 v) 406/826 vi) 376/963 vii) 60/90

d) What is the chance you'll get a female if you draw only from the left-handers?
i) 90/316 ii) 30/90 iii) 30/316 iv) 60/90 v) 60/647 vi) 90/947

e) What is the chance you'll get a left-hander if you draw only from the females?
i) 30/316 ii) 30/90 iii) 90/316 iv) 60/90 v) 60/647 vi) 90/947

f) Draw 3 students ***without*** replacement. What is the chance that ***all*** 3 students are left-handed?
i) $(873/963)^3$ ii) $1-(873/963)^3$ iii) 90/963*89/962*88/961 iv) $1-(90/963)^3$ v) 1- 90/963*89/962*88/961

g) Draw 3 students ***without*** replacement. What's the chance that ***not all*** 3 students are left-handed? = 1-P(all)
i) $(873/963)^3$ ii) $1-(873/963)^3$ iii) 90/963*89/962*88/961 iv) $1-(90/963)^3$ v) 1- 90/963*89/962*88/961

h) Draw 3 students ***with*** replacement. What is the chance that ***at least one*** student is right-handed? = 1- P(none)
i) $(826/963)^3$ ii) $1-(826/963)^3$ iii) $1-(137/963)^3$ **iv)** $1-(90/963)^3$ v) 90/963*89/962*88/961

Question 16 (6 pts.) **pertains to rolling fair dice.**

5,4 6,3 4,5 3,6

a) Two dice are rolled. What is the chance that the sum of the spots is 9?
i) 2/36 ii) 3/36 iii) 4/36 iv) 5/36 v) 1/6*1/6 vi) 7/36

4/36 5,5 4,6 6,4

b) Two dice are rolled. What is the chance that the sum of the spots is 9 **or** 10?
i) 5/36 ii) 6/36 iii) 7/36 iv) 8/36 v) 1/6*1/6 vi) 9/36

c) Two dice are rolled what is the chance the sum of the spots is either 6 **or** doubles (doubles is (1,1), (2,2), (3,3) etc.)?
i) 6/36 ii) 7/36 iii) 8/36 iv) 9/36 v) 10/36 vi) 11/36

1,5 1,1 5,1 2,2 2,4 3,3 4,2 4,4 3,3 5,5 6,6

d) What is the chance of rolling a die 7 times and getting no "3"s?
i) $(1/6)^7$ ii) $1-(1/6)^7$ iii) $1-(5/6)^7$ iv) 7*(1/6) v) $(5/6)^7$

e) What is the chance of rolling a die 7 times and getting all "3"s?
i) $(5/6)^7$ ii) $(1/6)^7$ iii) $1-(5/6)^7$ iv) $1-(1/6)^7$ v) 7*(1/6)

f) What is the chance of rolling a die *7 times and getting at least one "3"*? = 1-P(none)
i) $(5/6)^7$ ii) $1-(1/6)^7$ iii) $1-(5/6)^7$ iv) 7*(1/6) v) $(1/6)^7$

Question 17 (2 pts.) **pertains to tossing fair coins.**

a) What is the chance of tossing a fair coin 3 times and getting all tails?
i) ½ + ½ + ½ ii) ½ * ½* ½ iii) 1- (½ * ½* ½) iv) 3/6

b) What is the chance of tossing a fair coin 3 times and getting this particular sequence: HHT?
i) ½ + ½ + ½ ii) ½ * ½* ½ iii) 1- (½ * ½* ½) iv) 3/6