Stat 100 Exam 2 March 7, 2018

**STAT 100 EXAM 2 SPRING 2018**

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

**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 noon) **ii)** L2 (TR at 11am) **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 (16 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 _____17

Page 2 _____16

Page 3 _____12

Page 4 _____25

Page 5 _____15

Page 6 _____14

Cover _____1

**Total Score** _____

**There is NO CLASS tomorrow 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.**

Comma - Page 1

Stat 100 Exam 2 March 7, 2018

**Question 1** (6 pts.) Below is a distribution table for US income (in thousands of dollars) in 2016. The right-hand column shows the % of the population in each interval.

| Income (in thousands of $) | % |
|---|---|
| 0-10 | 25 |
| 10-30 | 25 |
| 30-50 | 15 |
| 50-90 | 10 |
| 90-210 | 24 |
| 210 and up | 1 |

a) Median = 30

b) Q1= 10 and Q3= 90

IQR = Q3 - Q1
= 90 - 10 = 80

c) What percent of the population are low outliers?
i) 0% ii) 1% iii) 2% iv) 25% v) not enough info

low outliers < Q1 - 1.5(IQR)
< 10 - 1.5(80)
= -105

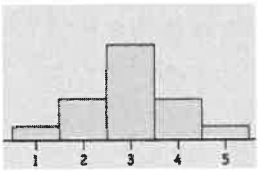
d) What percent of the population are high outliers?
i) 0% ii) 1% iii) 2% iv) 25% v) not enough info

high outliers > Q3 + 1.5(IQR)
> 90 + 1.5(80)
= 210

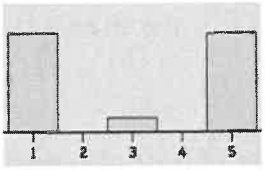
e) If we drew a histogram from the table it would …
i) be symmetrical ii) have a long left-hand tail iii) have a long right-hand tail

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

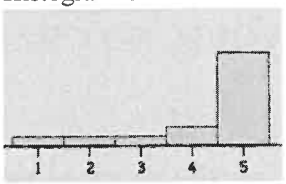
Histogram A



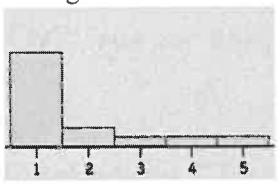

Histogram B



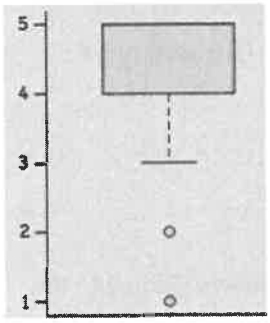

Histogram C



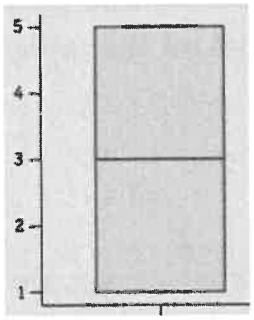

Histogram D



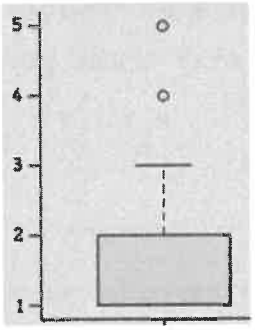

**Write the correct letter in each blank below to match the box plots with their corresponding histograms.**



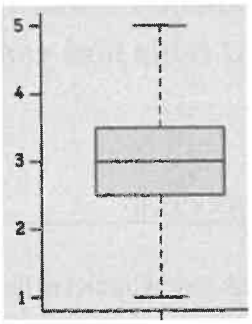

i) C




ii) B



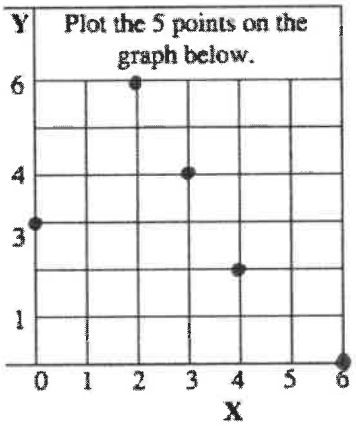

iii) D




iv) A

**Question 3** (7 pts.) Compute the correlation coefficient ( r ) between X and Y by filling in the table below. Then check that your answer makes sense by graphing the points in the box provided. *The average of X and Y is 3 and the SD of X and Y is 2.

(1pt) (1/2 pt for each blank in table below)




| X | Y | X in Standard Units (Zx) | Y in Standard Units (Zy) | Products |
|---|---|---|---|---|
| 0 | 3 | -1.5 | 0 | 0 |
| 2 | 6 | -0.5 | 1.5 | -0.75 |
| 3 | 4 | 0 | 0.5 | 0 |
| 4 | 2 | 0.5 | -0.5 | -0.25 |
| 6 | 0 | 1.5 | -1.5 | -2.25 |

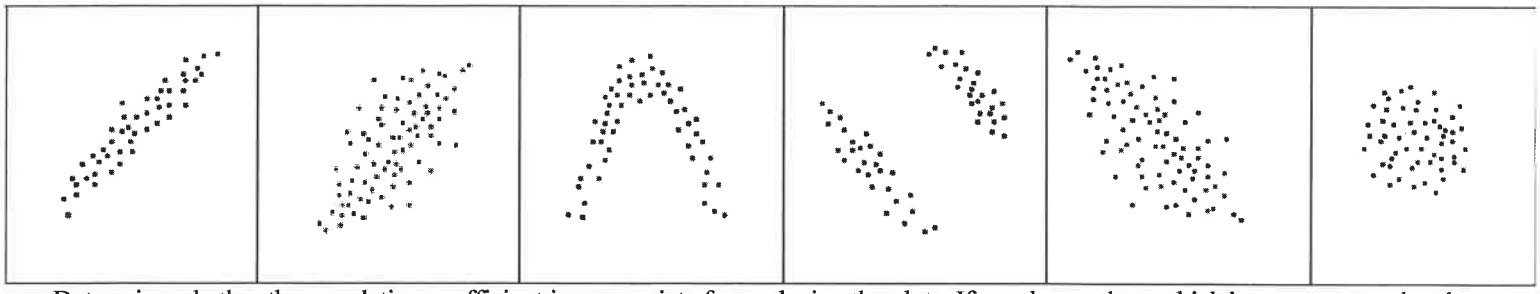
cont error on products

r = avg of products = $\frac{0+(-0.75)+0+(-0.25)+(-2.25)}{5}$

What is r? -0.65 (1 pt)

**Question 4 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- (Hint: One column will be used twice, the rest will be used exactly once.)***

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

**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 | About 0 | Between −1 and 0 | Exactly −1 | Not Enough Information |
|---|---|---|---|---|---|---|
| As X increases, Y usually decreases | ☐ | ☐ | ☐ | ☑ | ☐ | ☐ |
| As X increases, Y usually increases | ☐ | ☑ | ☐ | ☐ | ☐ | ☐ |
| X and Y are randomly paired numbers 1 to 100. | ☐ | ☐ | ☑ | ☐ | ☐ | ☐ |
| X and Y are always negative | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ |
| Y is always ***exactly*** 5 more than X | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ |
| X and Y ***always*** add up to 10. | ☐ | ☐ | ☐ | ☐ | ☑ | ☐ |

**Question 6** (4 pts.) X and Y are 2 sets of numbers with r = 0.4. How would r change if ....
(Assume all questions refer to the original X and Y values.)

**a)** All X values were multiplied by 2. The new r = 0.4

**b)** All Y values were multiplied by -0.5. The new r = -0.4

**c)** All X and Y values were increased by 0.3. The new r = 0.4

**d)** All X and Y values were changed to Z scores. The new r = 0.4

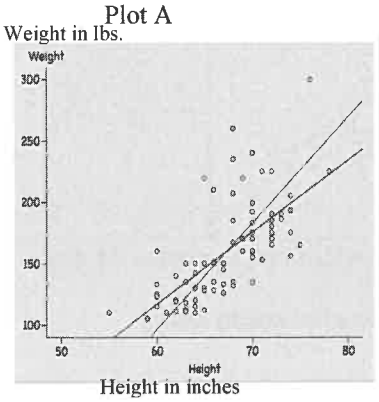
←Fill in all 4 blanks with numbers, NOT words.

Comma- Page 3

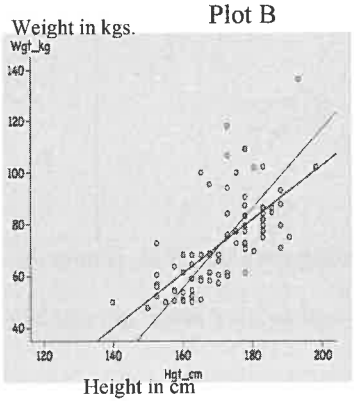


**Question 7** (4 pts.) The 4 scatter plots below depict the height and weight of 126 Stat 100 students. Plot A displays **height in inches** on the X axis and **weight in lbs** on the Y axis, while Plot B display the same data in **centimeters and kilograms**. Plots C and D show the same data but switch the X and Y axis, so that weight is on the X and height is on the Y. (The heights were converted to cm by multiplying them all by 2.54 cm/inch and the weights were converted to kgs by dividing them all by 2.2 lbs/kg.)

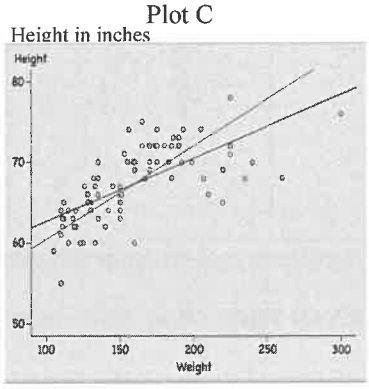
Plot A



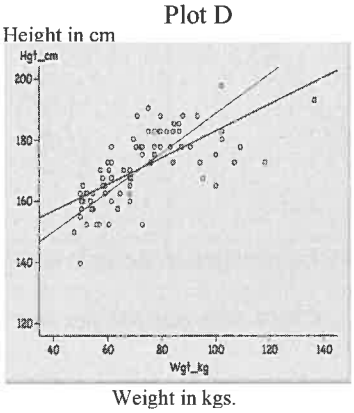

Plot B




Plot C




Plot D




One **or more** of the plots above has a correlation coefficient **r= 0.7**. For each plot circle the correct r. (**Hint:** You don't need to look at the graphs to answer this question. Just ***read*** the descriptions of the plots or focus on the relation of the regression line to the SD line for each plot.)

| | | | | | |
|---|---|---|---|---|---|
| **a)** | Plot A | ***Choose one:*** | (i) $r = 0.7$ | ii) $r > 0.7$ | iii) $r < 0.7$ |
| **b)** | Plot B | ***Choose one:*** | (i) $r = 0.7$ | ii) $r > 0.7$ | iii) $r < 0.7$ |
| **c)** | Plot C | ***Choose one:*** | (i) $r = 0.7$ | ii) $r > 0.7$ | iii) $r < 0.7$ |
| **d)** | Plot D | ***Choose one:*** | (i) $r = 0.7$ | ii) $r > 0.7$ | iii) $r < 0.7$ |

**Question 8** (2 pts.)

In 2004, millions of students nationwide took the Math and Verbal SAT tests. Within each state the average Verbal and the average Math SAT scores were calculated. The correlation between these 50 pairs of averages was 0.97. Does that mean that the correlation between the millions of individual students' Verbal and Math SAT scores would also be 0.97?

**Choose one:**

**i)** Yes, since the state averages are computed from the individual scores the correlation for individual Verbal and Math scores must also be 0.97.

**ii)** No, the correlation for individuals is likely to be higher than 0.97 since it's based on millions of data points instead of just 50.

**(iii)** No, the correlation for individuals is likely to be lower than 0.97 since the individuals within each state would add more scatter so the points wouldn't follow a straight line as closely.

**Question 9** (6 pts) pertains to a screening test for prostate cancer. Suppose **10%** of people who get tested have prostate cancer. If someone has cancer the test will correctly give a positive result 90% of the time and if they don't have cancer the test will correctly give a negative result 80% of the time.

***Fill in the following table*** for 100 people who get tested. (4 pts.- ½ pt for each blank)

| | **Positive result** | **Negative result** | **Total** |
|---|---|---|---|
| **Has Cancer** | 9 | 1 | ***(Hint: Fill in this blank first using underlined info above).*** 10 |
| **Does Not have Cancer** | 18 | 72 | 90 |
| **Total** | 27 | 73 | **100** |

cont error for totaling

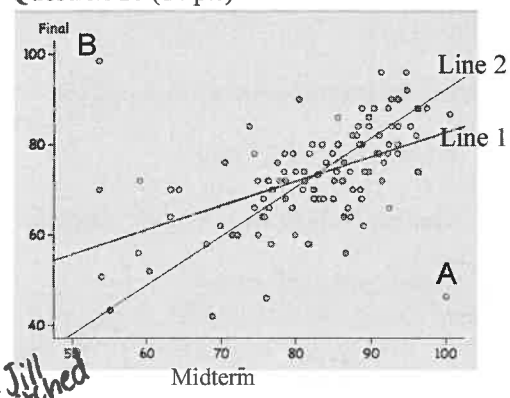
**a)** (1 pt.) If someone tests positive what's the chance he really has cancer? 9/27 or 1/3 (leave answer as a fraction)

**b)** (1 pt.) If someone tests negative what's the chance he really has cancer? 1/73 (leave answer as a fraction)

continued error from table

Stat 100 Exam 2 — Comma - Page 4 — March 7, 2018

**Question 10** (14 pts)




The scatter plot to the left shows the Final and midterm Exam scores of a group of students. **Here are the 5 summary stats: *r = 0.5***

| | *Average* | *SD* |
|---|---|---|
| *Midterm* | *82* | *10* |
| *Final* | *73* | *12* |

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

**b)** (2 pts.) A student scored an 82 on the Midterm and a 73 on the Final. Which line does he lie on? point of averages
i) The regression line only ii) The SD line only (iii) Both iv) Neither

-2 if Jack & Jill are switched

**c)** (4 pts.) Jack and Jill both scored exactly 1 SD above average on their midterms. Jack falls on the SD line and Jill falls on the regression line. What are their Final scores? Jack's Final = 85 Jill's Final= 79 *Show work below.*

* *No work no credit.* Jack midterm $z=1$ Jack final $z=1$ Jack final $= 73+(1)(12)$, $=85$ Jill midterm $z=1$ Jill final $z=0.5$

**d)** (2 pts.) If a student scored a 62 on the midterm, what's the regression estimate for their Final Score? ***Use the 3 step process.***

* *No work no credit* midterm 62 $z_{mid} = \frac{62-82}{10} = -2$ r $\times 0.5 =$ $z_{final} = -1$ final val $= 73+(-1)(12) = 61$ ½ cred for 85 (1pt) $73+(.5)(12) = 79$ Jill Final =

**e)** (1 pt.) If student A was removed the **correlation coefficient** would (i) increase ii) decrease iii) stay the same.

**f)** (1 pt.) If student B was removed the **RMSE** (SD of the prediction errors) would i) increase (ii) decrease iii) stay the same.

**g)** (2 pts.) Student B scored a 99 on the Final. The regression line predicts that he will score a 58 on the Final. What is his residual?

* *No work no credit.* residual = actual − predicted $= 99-58 = 41$

residual= 41 -1 for -41

**Question 11** (11 pts.) Suppose ACT and Math SAT scores among students who take both exams have the following summary statistics and the scatter plot is football shaped: **ACT: avg = 22 , SD=4 Math SAT: avg = 500, SD=100 r =0.8.**

**a)** (2 pts.) What is the **slope** of the regression equation when predicting Math SAT scores from ACT scores?
***Choose one:*** i) 0.024 ii) 0.032 iii) .05 iv) 15 v) 20

$m = r \times \frac{SD_y}{SD_x} = 0.8 \times \frac{100}{4} = 20$

**b)** (2 pts.) What is the **y-intercept** of the regression equation when predicting Math SAT scores from ACT scores?
***Choose one:*** i) - 60 ii) 0 iii) 40 (iv) 60 v) 170

$y = 20x + b$ $500 = 20(22) + b$ $b = 60$

**c)** (2 pts.) Using the slope and y-intercept you got in parts a and b, predict the Math SAT score of a student who got a 26 on the ACT. Make sure to use the regression equation. **No work no credit.** *

$y = 20x + 60$
$y = 20(26) + 60 = 580$

Predicted Math SAT= 580

**d)** (1 pt) Now, predict the Math SAT of a student who got a 26 on the ACT using the 3 step process. **No work no credit.** *

Math SAT = ? ACT = 26 ACT 26 $z_{ACT} = \frac{26-22}{4} = 1$ r $\times 0.8 = 0.8$ $z_{math}$ Math SAT $500 + (.8)(100) = 580$ ✓

Predicted Math SAT= 580

**e)** (2 pts.) What is the SD of the prediction errors (the RMSE) when predicting Math SAT scores from ACT scores?
(i) $\sqrt{1-0.8^2} * 100 = 60$ ii) $\sqrt{1-0.8^2} * 4$ iii) $\sqrt{1-0.6^2} * 4$ iv) $\sqrt{1-0.6^2} * 100$ v) 25

*f)* (2 pts.) The regression equation predicts Math SAT scores of 500 for those who score 22 on the ACT. 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

$500 \pm 2\text{RMSEs} \Rightarrow 500 \pm 2(60)$

*Check that you showed work for each problem that said "No work, no credit", otherwise you'll get a 0.

**Question 12** (11 pts.)

Suppose blood pressure and temperature follow the normal curve but have different correlations among different populations.

**a)** (2 pts.) Imagine a population where there is a perfect positive correlation (r = 1) between the 2 measurements, then everyone's blood pressure percentile would ____ their temperature percentile. ***Choose one:***
**i)** approximately equal **ii)** exactly equal **iii)** be closer to the 50th percentile than **iv)** be exactly 100 minus

**b)** (2 pts.) Imagine a population where there is a perfect negative correlation (r = -1) between the 2 measurements, then everyone's blood pressure percentile would ____ their temperature percentile. ***Choose one:***
**i)** approximately equal **ii)** exactly equal **iii)** be closer to the 50th percentile than **iv)** be exactly 100 minus

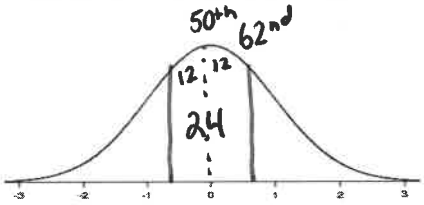
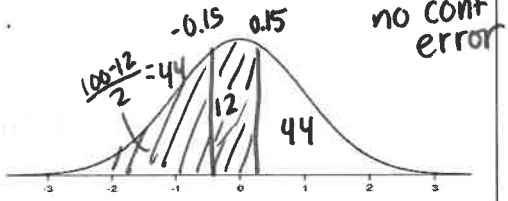
**c)** (2 pts.) Imagine a population where there is zero correlation (r = 0) between the 2 measurements, then no matter what someone's blood pressure percentile is, the regression estimate for his temperature would always be the __50__th percentile.
***Fill in the blank with a number***

**d)** (1 pt.) Imagine a population where the correlation between the 2 measurements is 0.5. If someone's blood pressure is in the 30th percentile, then the regression estimate for his temperature would be closest to the _____ percentile. ***Choose one:***
a) 20th b) 30th c) 40th d) 50th e) 60th f) 70th g) 80th

**e)** (1 pt.) Imagine a population where the correlation between the 2 measurements is - 0.5. If someone's blood pressure is in the 30th percentile then the regression estimate for his temperature would be closest to the _____ percentile. ***Choose one:***
a) 20th b) 30th c) 40th d) 50th e) 60th f) 70th g) 80th

**f)** (3 pts.) If a person's blood pressure is in the **62nd** percentile where r = 0.5, estimate her temperature percentile ***by filling in the table below.***
***You may round areas and z-scores to fit the nearest line on the table.***

accept exact or rounded answers

| Blood Pressure Percentile | Blood Pressure Z | r | Temperature Z | Temperature Percentile |
|---|---|---|---|---|
| Person is in the 62nd percentile for blood pressure. Correctly mark the graph and write the correct Z in the next column.   | Z = 0.3 | × r = 0.5 | = Z = 0.15 cont error from 1st z-score | Temperature Percentile = 56 Mark the Z score on the graph below.   |

**Question 13** (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 | 48th | -0.05 |
| C | 2.5th | -2 |
| D | 84th | 1 |

accept rounded + exact answers

Comma - Page 6

Stat 100 Exam 2 March 7, 2018

**Question 14** pertains to the table below which shows the survey responses of the 381 students who identified themselves as "white" and the 223 students who identified themselves as "non-white" to the question "Do you believe sex before marriage is immoral?"

| | Yes, immoral | Depends on type of sex | No, not immoral | Totals |
|---|---|---|---|---|
| White | 49 | 28 | 304 | 381 |
| Non-White | 63 | 29 | 131 | 223 |
| Totals | 112 | 57 | 435 | 604 |

*(4 pts)* Suppose you draw **randomly from the students who answered this survey**:

*a)* What is the chance that you'll get a student who answered "Yes, immoral"?
i) 112/604 ii) 49/112 iii) 49/381 iv) 63/223 v) 63/112 vi) 381/604

*b)* What is the chance that you'll get a student who answered "Yes" if you draw only from the white students?
i) 112/604 ii) 49/112 iii) 49/381 iv) 63/223 v) 63/112 vi) 381/604

*c)* What is the chance that you'll get a student who answered "Yes" if you draw only from the non-white students?
i) 112/604 ii) 49/112 iii) 49/381 iv) 63/223 v) 63/112 vi) 381/604

*d)* What is the chance that you'll get a non-white student if you draw only from those who answered "Yes"?
i) 112/604 ii) 49/112 iii) 49/381 iv) 63/223 v) 63/112 vi) 381/604

**Question 15 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. For example, there are 4 Kings and 13 Hearts.)** (4 pts.)

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

**b)** Draw 2 cards **without replacement**. What is the chance that the first card is a King and the second is a Queen?
**i)** 8/52 **ii)** 7/52 **iii)** 4/52*3/51 **iv)** 4/52*4/51 **v)** 4/52*4/52

**c)** Draw 2 cards **with replacement**. What is the chance that both cards are Kings?
**i)** 8/52 **ii)** 7/52 **iii)** 4/52*3/51 **iv)** 4/52*4/51 **v)** 4/52*4/52

**d)** Draw ***one*** card. What is the chance that it's either a King or a Heart? P(A or B) = P(A) + P(B) - P(A and B)
**i)** 4/52 + 13/52 **ii)** 4/52 + 13/52 -1/52 **iii)** 4/52*3/51 **iv)** 4/52*13/51 **v)** 4/52*13/52

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

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

**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) 1/6 + 1/6

**b)** One die is rolled 4 times. What is the chance of getting ***all*** 1'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 ***not*** getting all 1's?
i) $(5/6)^4$ ii) $(1/6)^4$ iii) $1-(5/6)^4$ iv) $1-(1/6)^4$ v) 4/6

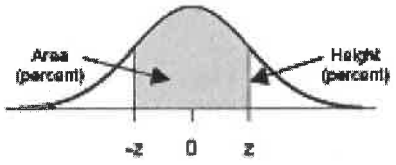
**d)** One die is rolled 4 times. What is the chance of getting ***no*** 1's?
i) $(5/6)^4$ ii) $(1/6)^4$ iii) $1-(5/6)^4$ iv) $1-(1/6)^4$ v) 4/6

**e)** One die is rolled 4 times. What is the chance of getting ***at least one*** 1?
i) $(5/6)^4$ ii) $(1/6)^4$ iii) $1-(5/6)^4$ iv) $1-(1/6)^4$ v) 4/6

**f)** Two dice are rolled. What is the chance that the sum of the spots is odd? (In other words, what is the chance that the sum of the spots is either 3, 5, 7, 9, or 11)? i) 14/36 ii) 15/36 iii) 16/36 iv) 17/36 v) 18/36

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

## 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 |