Dawson College: **Probability and Statistics**: 201-BZS-05

February 25, 2013

Last Name: SOLUTIONS

First Name: ____________

Student ID: ____________

Test 1

The time allowed for this test is 1 hour and 45 minutes. Please answer all questions in the space provided. A formula sheet is available upon request but if you decide not to use it you will be awarded 3 bonus marks. Remember to write clearly and use correct notation.

Question 1. Listed below are the yearly (absolute) percent changes of the price of a stock over the last seventeen years. In other words, each value is the percent difference in price on Jan 1st from one year to the next (all changes made positive).

6.7	6.1	5.2	8.8	2.6	6.1	1.0	3.8	7.1	3.2
6.4	6.1	1.0	6.2	0.6	5.5	5.3			

(a) *(3 marks)* Sort the data using a stem and leaf display (remember to correctly label your units).

```
0 | 6
1 | 0 0
2 | 6
3 | 8 2
4 |
5 | 2 5 3
6 | 7 1 1 4 1 2
7 | 1
8 | 8
```

STEM: ONES

LEAF: TENTHS

0.6 1.0 1.0 2.6 3.2 3.8 5.2 5.3 5.5 6.1

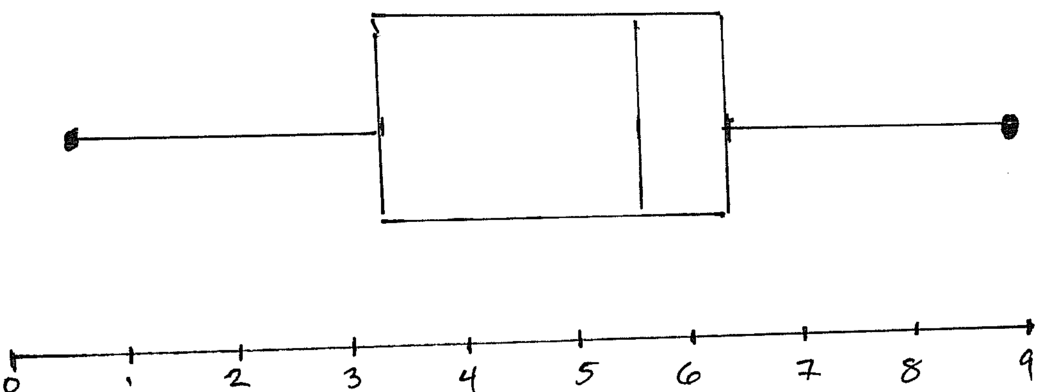
6.1 6.1 6.2 6.4 6.7 7.1 8.8

(b) *(5 marks)* Construct a box and whiskers diagram for this data.

$= 0.6 \cdot H = 8.8$ • $(0.25)17 = 4.25 \Rightarrow d = 5$ $\therefore P_{25} = Q_1 = 3.2$

$(0.5)(17) = 8.5 \Rightarrow d = 9 \Rightarrow Q_2 = 5.5$

$(0.75)(17) = 12.75 \Rightarrow d = 13 \Rightarrow Q_3 = 6.2$



Question 2. *(4 marks)* Scores on the SAT have a mean of 1518 an a standard deviation of 325. Scores on the ACT have a mean of 21.1 and a standard deviation of 4.8. Scores on both tests are approximately normally distributed. Which is better: A score of 1640 on the SAT or a score of 23.0 on the ACT? (Hint, which score is "farter away" from the mean?)

SAT: $z = \dfrac{1640 - 1518}{325} = 0.3754$

ACT: $z = \dfrac{23.0 - 21.1}{4.8} = 0.3958$

23.0 ON THE ACT IS BETTER. (FARTHER FROM MEAN)

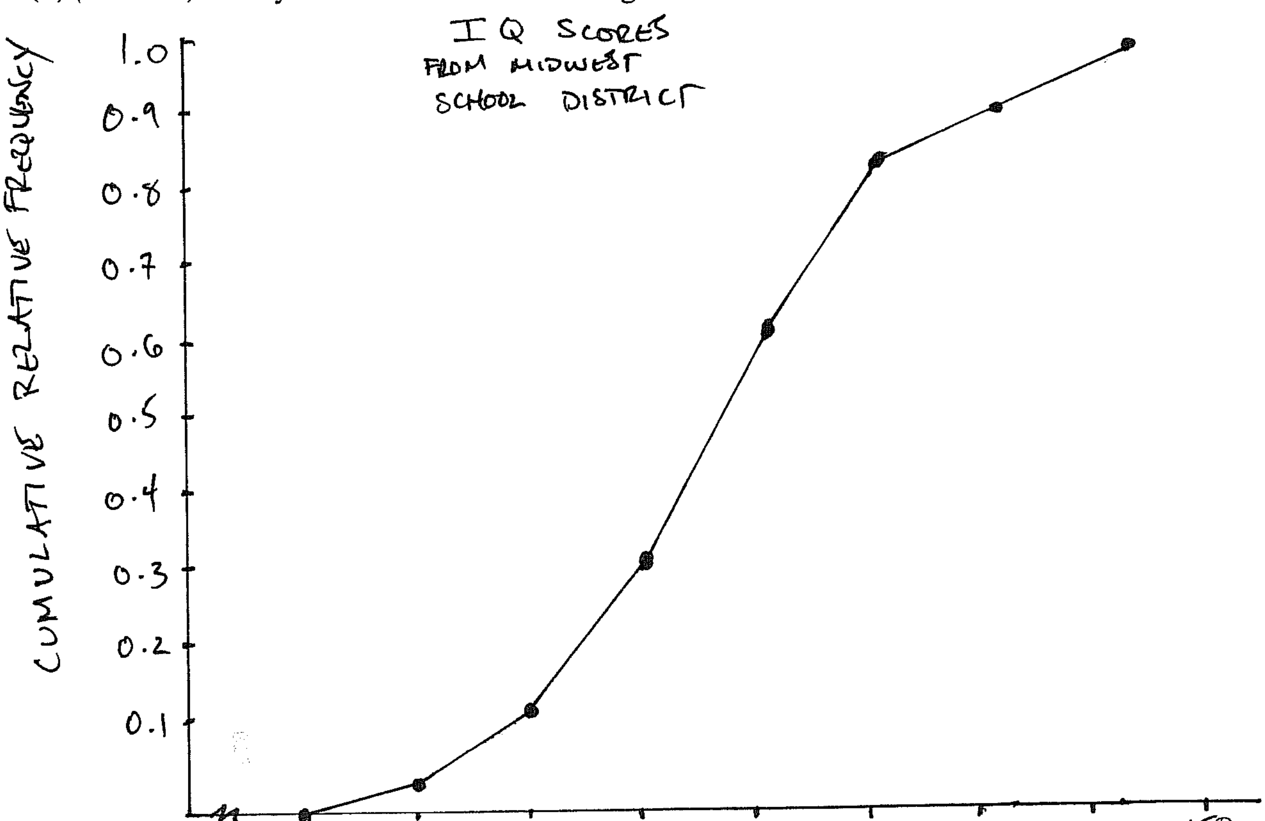
Question 3. The data listed below is a sample of IQ scores from a seventh grade class in a Midwest school district.

114	100	104	89	102	91	114	114	103	105
93	108	130	120	132	121	128	118	92	86
74	98	103	112	107	103	98	96	112	

(a) *(4 marks)* Make a grouped frequeny distribution table with 70 as the lower bound of the first class, and use 10 as the class width. Add to this chart the cumulative frequency and the relative cumulative frequency.

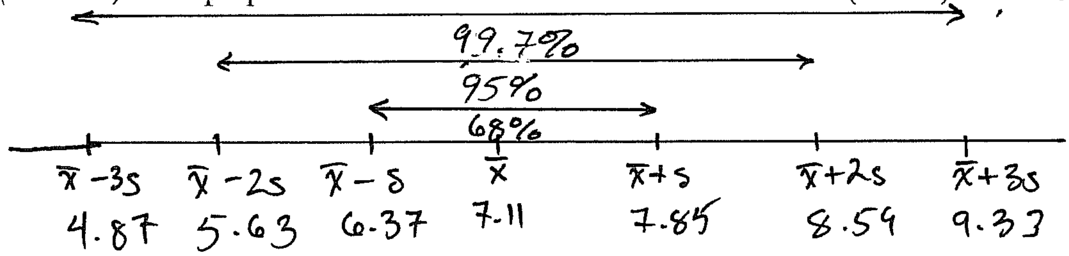
CLASS	f	cf	crf
70 - 79	1	1	$1/29 \approx 0.034$
80 - 89	2	3	$3/29 \approx 0.103$
90 - 99	6	9	$9/29 \approx 0.310$
100 - 109	9	18	$18/29 \approx 0.621$
110 - 119	6	24	$24/29 \approx 0.828$
120 - 129	3	27	$27/29 \approx 0.931$
130 - 139	2	29	$29/29 = 1$

(b) *(3 marks)* Neatly construct a well labeled ogive for this data.



Question 4. During a physical training program, one mile run times for male students at the University of Illinois were collected. The run times turned out to be approximately normally distributed, with a mean of 7.11 minutes and a standard deviation of 0.74 minutes.

(a) *(3 marks)* What proportion of these students had run times of more (slower) than 8.59 minutes?



SLOWER THAN 8.59 MINUTES: $0.50 + \frac{1}{2}(0.95) = 0.975$

$1 - 0.975 = 0.025$

2.5% OF THE STUDENTS

(b) *(3 marks)* What proportion of these students had run times of less (faster) than 7.85 minutes?

LESS THAN 7.85 MINUTES: $0.5 + \frac{1}{2}(0.68) = 0.84$

84% OF THE STUDENTS.

Question 5. Coffee is a leading export from several developing countries. It has been observed that when coffee prices are high, farmers have an incentive to clear forest to plant more coffee trees. Here are six years of data on the average price paid to coffee growers in Indonesia and the percent of forest area lost in a coffee producing region in that country.

	x (Price in Cents per Lb)	y (Forest Area Lost in %)	xy	x^2	y^2
	29	0.49	14.21	841	0.24
	54	1.69	91.26	2916	2.86
	40	1.59	63.6	1600	2.53
	55	1.82	100.1	3025	3.31
	72	3.10	223.2	5184	9.61
	68	2.71	184.3	4624	7.34
sum	318	11.4	676.7	18190	25.9

(a) *(4 marks)* Calculate the coefficient of linear correlation

$$r = \frac{SS(xy)}{\sqrt{SS(x)SS(y)}}$$

Does your result support the idea that higher coffee prices increase forest loss? Is your result strong enough to confirm that the increasing loss in forest area is caused by higher coffee prices?

$$SS(xy) = \sum xy - \frac{(\sum x)(\sum y)}{n} = 676.7 - \frac{(318)(11.4)}{6}$$

$$= 72.5$$

$$SS(x) = \sum x^2 - \frac{(\sum x)^2}{n} = 18190 - \frac{(318)^2}{6} = 1336$$

$$SS(y) = \sum y^2 - \frac{(\sum y)^2}{n} = 25.9 - \frac{(11.4)^2}{6} = 4.24$$

$$r = \frac{72.5}{\sqrt{(1336)(4.24)}} = 0.96$$

THIS SHOWS A STRONG POSITIVE CORRELATION AND DOES SUPPORT THE IDEA THAT HIGHER COFFEE PRICES INCREASE FOREST LOST. HOWEVER, WE CANNOT CONFIRM THAT THE HIGHER PRICES CAUSED THE FOREST LOSS

(b) *(4 marks)* Find the equation of the line of best fit using

$$b_1 = \frac{SS(xy)}{SS(x)} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} \qquad b_0 = \frac{\Sigma y - (b_1 \cdot \Sigma x)}{n} = \bar{y} - (b_1 \cdot \bar{x})$$

How much forrest area would you expect to be lost if the price paid to coffee growers was 75 cents per Lb?

$$b_1 = \frac{72.5}{1336} = 0.054266$$

$$b_0 = \frac{11.4 - (0.054266)(318)}{6} = -0.976$$

$$\therefore \boxed{\hat{y} = 0.0543x - 0.976}$$

$$\hat{y} = 0.0543(75) - 0.976$$
$$= 3.10$$

WE WOULD EXPECT 3.10% FOREST LOSS.

Question 6. *(4 marks)* Suppose that we draw five cards from a standard deck of 52, without replacement. Find the probability that exactly two of those five cards are spades.

OF WAYS TO PICK 2 SPADES

OF WAYS TO PICK OTHER CARDS

$$\frac{{}_{13}C_2 \cdot {}_{39}C_3}{{}_{52}C_5} = \frac{(78)(9139)}{2598960} = 0.2743$$

OF 5 CARD HANDS

Question 7. A study of readers of a popular sports magazine was conducted. It was found that, of there readers

- 45% were hockey fans
- 55% were football fans
- 32% were baseball fans
- 18% were fans of both hockey and football
- 11% were fans of both football and baseball
- 9% were fans of both hockey and baseball
- 5% were fabs of all three sports

(a) *(3 marks)* What is the probabiliy that a reader is both a baseball and football fan given that they are a hockey fan?

SUPPOSE WE SELECT A READER AT RANDOM. LET A BE THE EVENT THAT THE READER IS A HOCKEY FAN, B FOOTBALL FAN, C BASEBALL FAN.

WE WANT $P(B \cap C | A) = \frac{P((B \cap C) \cap A)}{P(A)} = \frac{P(A \cap B \cap C)}{P(A)}$

$= \frac{0.05}{0.45} = 0.\overline{1}$ OR 11.1%

(b) *(4 marks)* What is the probabiliy that a reader is a baseball fan if we know that they are a fan of at least one of these three sports?

$$P(C | A \cup B \cup C) = \frac{P(C \cap (A \cup B \cup C))}{P(A \cup B \cup C)}$$

$$= \frac{P(C)}{P(A)+P(B)+P(C)-P(A \cap B)-P(A \cap C)-P(B \cap C)+P(A \cap B \cap C)}$$

$$= \frac{0.32}{0.45+0.55+0.32-0.18-0.11-0.09+0.05}$$

$$= \frac{0.32}{0.99} = 0.\overline{32}$$ OR 32.3%

Question 8. *(6 marks)* A certain drug test will correctly identify a drug user as testing positive 98% of the time, and will correctly identify a non-user as testing negative 99% of the time. 1% of the employees at a corporation actually use the drug. Given a negative drug test result, what is the probability that an employee is actually a drug user? Clearly defne the events under consideration in your explanation.

LET B BE THE EVENT THAT THE EMPLOYEE TESTED NEGATIVE

A_1 BE THE EVENT THAT THE EMPLOYEE IS A DRUG USER

A_2 " " " " " " IS NOT A DRUG USER.

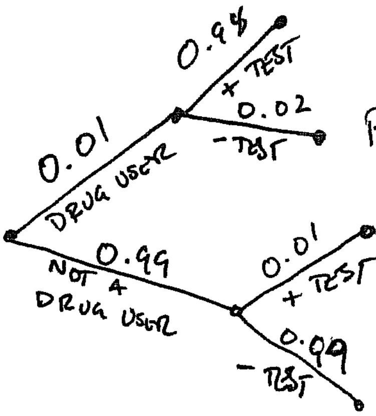
WE WANT $P(A_1|B) = \dfrac{P(B|A_1)\cdot P(A_1)}{P(B|A_1)\cdot P(A_1) + P(B|A_2)\cdot P(A_2)}$

$= \dfrac{(0.02)(0.01)}{(0.02)(0.01) + (0.99)(0.99)}$

$= \dfrac{0.0002}{0.9803}$

$= 0.0002$

OR



$P(A_1 \cap B) = (0.01)(0.02) = 0.0002$

$P(A_2 \cap B) = 0.9801$

$\therefore P(A_1|B) = \dfrac{P(A_1 \cap B)}{P(B)} = \dfrac{P(A_1 \cap B)}{P(A_1 \cap B) + P(A_2 \cap B)} = \dfrac{0.0002}{0.0002 + 0.9801}$

$= 0.0002$