Statistics 1, Section 6	Name Key
Exam 1	October 6, 2008

Calculators and one 8.5" by 11" sheet of handwritten notes allowed. Show all work and answers clearly in the space provided. Questions 1-6 are worth 15points each; question 7 is 10 points. There are 100 points possible.

1. The values below are snow depths (in centimeters) measured at different locations.

8 8 12 16 18 19 22 25
a. Find the mean and median.
-5x 128 1/
$\chi = \frac{2}{10} = \frac{1}{2} = \frac{10}{10}$
$median = \frac{16718}{16718} = 17$
b. Calculate the standard deviation using the formula $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{2}}$. Show work.
$X \mid x - \overline{x} \mid (x - \overline{x})^2$
8 2-11 = -2 1.4
2 0 16 0 161
8 -8 64 (274 '
$ 2 - 4 6 \leq 1 = 1$
11.000
$= 10^{-2}$
193 1 - 0.00
23/10 36
25 q 81
77.6
c. Calculate Q_1 $(2+9)$
$a_{1}(1) = \frac{1}{2}(1) = \frac{1}{2}(0) = 227$
post(10) - 4(11) - 4(9) - 2(2)

between
$$8 \neq 12$$

 $Q_1 = 8 + 0.25(12 - 8) = 8 + 1 = 9$

2. An experiment is conducted to determine which of two diets is more effective. The subjects are randomly assigned to Diet 1 or Diet 2. After six months on the assigned diet, the weight loss of each subject is measured. The boxplots below summarize the resulting data.



- 4 a. Use the boxplots to estimate the median weight loss for Diet 1 and Diet 2.
 - $m_1 = Q$ $m_2 = 12$ b. Which diet group had the largest interquartile range (IQR)? Estimate the IQR for this group. Explain what this IQR means in simple English (pretend you are explaining it to someone who has never taken statistics).

Piet Z has largest IQR.
IQR=Q3-Q1
$$\approx$$
 17-5=12

- The IQR is the spread of the middle 50% of the data. c. Calculate the upper fence for the diet group from part (b). What is the upper
- 5 fence used for?

3. The histogram and summary statistics below are for a data set consisting of the wealth in billions of dollars of 233 billionaires.

Column	n	Mean	Variance	Std. Dev.	Max	Q1	Q3
wealth	233	2.681545	11.014701	3.3188403	37	1.3	3



- a. Is there skewness in this data set? If so, toward what values? kewed toward high values (right)
- b. The empirical rule states that approximately 68% of data fall within one standard deviation of the mean.
 - i. Calculate mean +/- one standard deviation for this data. (OK to round to the nearest tenths.) $2,7\pm3,3$ (-0, 6, 6, 0)
 - ii. Approximately what percent of the data points fall in this interval? (Hint: The height of the bar at 37 represents one person.)

$$\frac{125+70+25}{222} = 94.40/0$$

iii. Is your answer to (ii) consistent with the empirical rule? If not, does this mean the empirical rule is not valid? Explain.

No. Empirical Rule Says about 68°/0 of the data are within one standard dev. of the mean, we have about 94°/0 within one stand dov. c. This data set contains two outliers - one person at 24 and one at 37 billion of the man. dollars. Suppose these two individuals had a fit of generosity and each gave away

1

all but 15 billion dollars of their wealth. Would this change the standard deviation? If so, higher or lower?

4. There are three finalists in a beauty pageant: Flora, Sabrina and Janet. Assume the contestants are equally beautiful and talented so the judges randomly choose who will win first, second and third place.

 ζ_{2} a. List the sample space for this experiment. (Hint: Tree diagram)

B. First prize goes to Flora, second to Sabrina and third to Janet.
B. FISTLETSZ VGT G V6
C. Flora wins first prize. UNDE $\{FSJ,FJS\}$ 2/6 = 1/33 d. Flora wins second or third prize.

$$\gamma$$
 e. Janet wins third prize.

5. Workers at a business are classified according to religion and gender. The results are shown in the table below.

	Male	Female
Protestant	30	20
Catholic	45	30
Jewish	45	30
Other	7	8

A worker is randomly selected from this business. What is the probability the worker is

9 a. Catholic
$$\frac{75}{215} = \frac{15}{43}$$

3 b. Catholic, given the worker is a female
30/88
Ac. Jewish and Male
(15/215
3 d. Jewish or Male
 $\frac{127+30}{215} = \frac{157}{215}$

- Le. Are the events Jewish and Male mutually exclusive? Why or why not? No. There are 45 subjects that are Jewish of Male so these events can occur at the 2 f. Male, given the worker is Jewish Same time 3 45175
- 6. Consider the bivariate data in the table below for this problem. X = price of a widget in dollars, Y= number of widgets sold (in thousands).



 l_{ℓ} a. Calculate the correlation coefficient for this data. Hint: The standard deviations of x and y are, 1.29 and 1.41, respectively.

$$Sxy = \frac{2xy - h \sum x \sum y}{n-1} = \frac{25 - \frac{1}{4}(10)(12)}{4 - 1} = \frac{25 - 30}{3} = -\frac{5}{3}$$

$$r = \frac{5xy}{5x5y} = \frac{-5/3}{1.29(1.41)} = \frac{-0.92}{1.29(1.41)}$$

b. Calculate the best fit regression line. 5 $b = \frac{5xy}{5x^2} = \frac{-5/3}{1.29^2} = -1.00$ $b = \frac{5xy}{5x^2} = \frac{-5/3}{1.29^2} = -1.00$ y = 5.5 - x $a = \overline{y} - b\overline{x} = (\frac{12}{4}) - (-1.00)(\frac{10}{4}) = 3 + 2.5 = 5.5$ c. Use the line in part (b) to predict the number of widgets sold when the price is \$2.50. y = 5.5 - 2.50 = 3 + 3d. Interpret the slope calculated in part (b). Give a specific interpretation involving price of widgets and sales 2

Э price of widgets and sales.

a. Match the correlation coefficient with the appropriate plot. Note there are more correlation coefficients than plots so some will not be used: -1.00, - 0.70, 0, 0.90, 1.00, 1.10.



b. Which of the above correlation coefficients would you prefer if you were interested in accurate prediction of y for a given value of x?

Plot 4