Statistics 1, Section 4 **Name** HLY

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.

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

   19 18 12 25 22 8 8 16

   8 8 12 16 18 19 22 25

   Calculate:

   3 a. Mean

   $\bar{X} = \frac{128}{8} = 16$

   3 b. Median

   position $\frac{1}{2}(n+1) = \frac{1}{2}(8+1) = 4.5$ $m = \frac{16+18}{2} = 17$

   5 c. Standard deviation using the formula $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$. Show work.

x	$x - \bar{x}$	$(x-\bar{x})^2$
8	$8-16=-8$	64
8	-8	64
12	$12-16=-4$	16
16	$16-16=0$	0
18	2	4
19	3	9
22	$22-16=6$	36
25	$25-16=9$	81
		274

   $s = \sqrt{\frac{274}{8-1}} = \boxed{6.3}$

   4 d. Calculate the first quartile, Q_1.

   position $\frac{1}{4}(8+1) = \frac{9}{4} = 2.25$

   $Q_1 = 8 + 0.25(12-8) = 8 + 0.25(4) = \boxed{9 = Q_1}$

2. A portion of an ***ordered*** data set is 0.4, 1.7, 4.3, 4.3,..., 364.4, 418.1, 420.7, 450.4, 460.9, 679.8. There are 100 observations in the data set and the five-number summary is: 0.4, 25.7, 81.4, 166.9, 679.8

   4 a. Calculate the interquartile range.

   $Q_3 - Q_1 = 166.9 - 25.7 = 141.2$

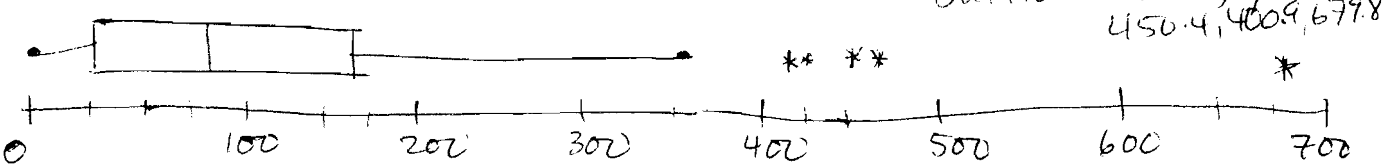
   5 b. Identify outliers, if they exist.

   lower fence $= 25.7 - 1.5(141.2) = -186.1$

   upper $= 166.9 + 1.5(141.2) = 378.7$

   outliers: 418.1, 420.7, 450.4, 460.9, 679.8

   5 c. Construct a boxplot for the data.

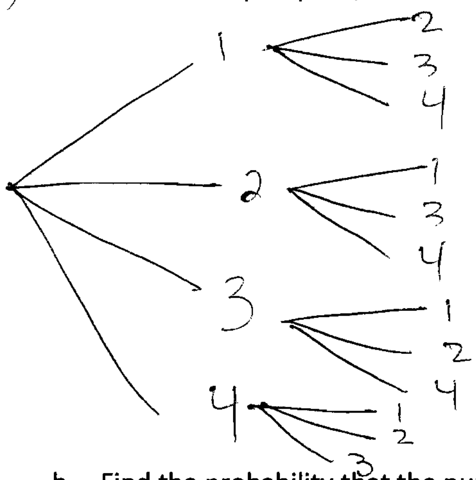


   1 d. Are the data skewed or symmetric?

   skewed

3. The numbers 1,2,3 and 4 are written on four slips of paper. Suppose two of the slips of paper are randomly selected from a hat without replacement.

a. List the sample space, S.



$S = \{(1,2), (1,3), (1,4), (2,1), (2,3), (2,4), (3,1), (3,2), (3,4), (4,1), (4,2), (4,3)\}$

12 outcomes

b. Find the probability that the numbers on the two randomly selected slips are:

i. 1 and 2 $P(\{(1,2),(2,1)\}) = 2/12 = 1/6$

ii. Both even

$P(\{(2,4),(4,2)\}) = 2/12 = 1/6$

iii. Sum to 5

$P(\{(1,4),(2,3),(3,2),(4,1)\}) = \frac{4}{12} = 1/3$

iv. Sum to a number greater than 3

$1 - P(\text{sum} \le 3) = 1 - P(\{(1,2),(2,1)\}) = 1 - \frac{2}{12} = \frac{10}{12} = 5/6$

4. Data is collected for a random sample of subjects on their favorite music genre and their income level. Income is classified as "high" if over \$100,000 and "low," otherwise. The results are summarized in the table below.

	Rock	Country	Classical	
High income	30	20	50	100
Low income	55	120	25	200
				300

If a person from this group is randomly selected, what is the probability the person

a. Has a high income? $\frac{100}{300} = \frac{1}{3}$

b. Has a high income AND likes classical music?

$50/300 = 1/6$

c. Has a high income OR likes classical music?

$$\frac{100+25}{300} = \frac{125}{300} = \frac{5}{12}$$

d. Given the person has high income, what is the probability he/she prefers country music?

$20/100 = 1/5$

e. Given the person likes rock music, what is the probability he/she has a low income?

$55/85 = 0.65$

f. Are the events the person has a high income and prefers country music mutually exclusive? Why or why not? No. There are 20 subjects who had high income & prefer country music so these events can happen at the same time.

5. 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).

x	y	xy
3	2	6
2	3	6
4	2	8
1	5	5
10	12	$\Sigma xy = 25$

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

$$s_{xy} = \frac{\Sigma xy - \frac{1}{n}\Sigma x \Sigma y}{n-1} = \frac{25 - \frac{1}{4}(10)(12)}{4-1} = \frac{25-30}{3} = \frac{-5}{3}$$

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

b. Calculate the best fit regression line.

$$b = \frac{s_{xy}}{s_x^2} = \frac{-5/3}{(1.29)^2} = -1.00$$

$$\boxed{y = 5.5 - x}$$

$$a = \bar{y} - b\bar{x} = \left(\frac{12}{4}\right) - (-1.00)\left(\frac{10}{4}\right) = 3 + 2.5 = 5.5$$

c. Use the line in part (b) to predict the number of widgets sold when the price is $2.50.

$y = 5.25 - 2.50 = 2.75$ thousand

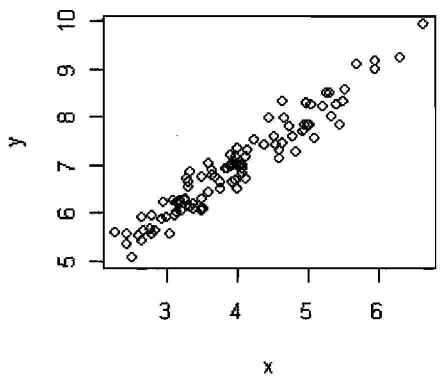
d. Interpret the slope calculated in part (b). Give a specific interpretation involving price of widgets and sales.

for every dollar increase in widget price, sales decline by about 1 thousand units

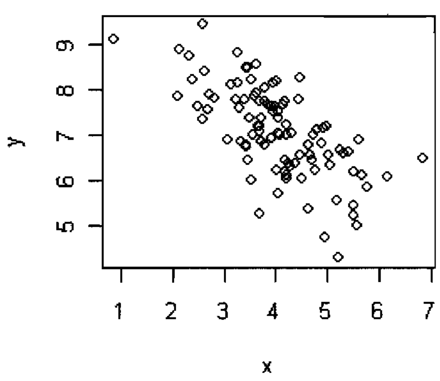
6.

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.

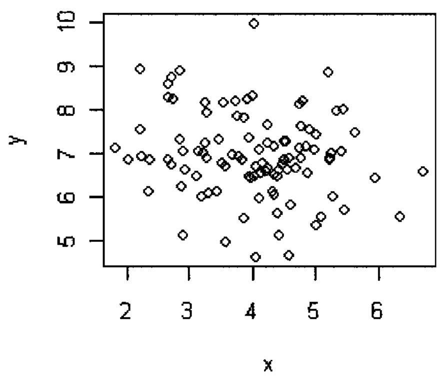
Plot 1 : r= 0.90



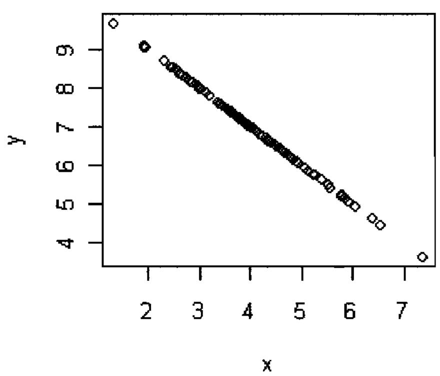
Plot 2 : r= -0.70



Plot 3 : r= 0



Plot 4 : r= -1.00

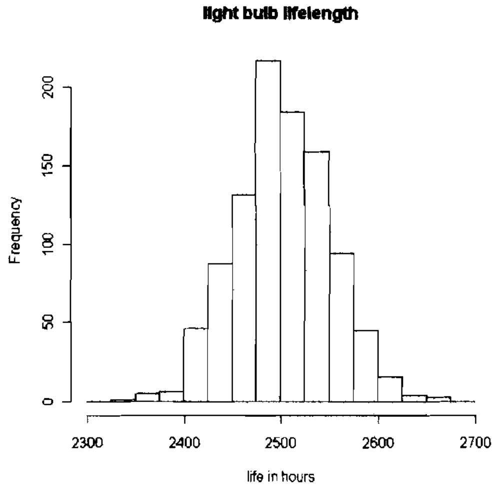


b. Which of the above correlation coefficients would you prefer if you were interested in accurate prediction of y for a given value of x?

r = -1 or r = 1

Prob	Time
1	4
2	3
3	4
[illegible]	1

7. Brand X light bulbs have a mean life of 2500 hours with a standard deviation of 50 hours. A large sample of Brand X light bulb lifelengths yields the histogram shown below.



a. Would the Empirical Rule or Chebychev's Rule or both apply to this data set? Explain why. Both would apply. Chebychev's applies to ANY dataset. The Empirical Rule applies to mound-shaped data and the above graph looks mound-shaped.

b. Use the best rule from part (a) to estimate the percent of light bulbs that will last between 2400 and 2600 hours. The Empirical rule gives more accurate estimates.

$\bar{x} \pm 2s = 2500 \pm 2(50) = 2500 \pm 100$

$(2400, 2600)$ Empirical Rule: $\approx 95\%$ within 2 st. dev.

c. What can you say about the percent of light bulbs that will last over 2650 hours?

note: $\bar{x} + 3s = 2500 + 3(50) = 2500 + 150 = 2650$

$\approx 99.7\%$ of data in $\bar{x} \pm 3s$ so about 0.3% of data outside this interval. So about $\frac{0.3}{2} = 0.15\%$ above 2650

d. Which rule, Chebychev's or the Empirical Rule, would you use for a data set on the salaries of actors and actresses? Give your reasons.

Salaries of actors are likely to be right skewed since a few make extremely high salaries. Therefore, the ~~Empeer~~ Empirical Rule would not apply. Use Chebychev's.