

Calculators and one 8.5" by 11" sheet of handwritten notes allowed. Show all work and answers clearly in the space provided. There are 80 points possible.

1. The number of eggs laid by 4 hens in a month are:

28 29 24 31

Calculate:

a. (4 pts) Mean $\bar{x} = \frac{28+29+24+31}{4} = 28$

- b. (4 pts) Standard deviation using the formula $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$. Calculate by hand and show all work.

x	$x - \bar{x}$	$(x - \bar{x})^2$
28	$28 - 28 = 0$	0
29	$29 - 28 = 1$	1
24	$24 - 28 = -4$	16
31	$31 - 28 = 3$	9

Total

$26 = \sum(x - \bar{x})^2$

$s = \sqrt{\frac{26}{4-1}}$

$= \sqrt{\frac{26}{3}} = \sqrt{8.6}$

$= 2.944$

- c. (2 pts) What is the deviation from the mean of the data value 24? Indicate whether the deviation is positive or negative.

$x - \bar{x} = 24 - 28 = -4$ negative

2. The number of miles until the first major repair for a new model of electric car has a bell-shaped distribution with mean 80,000 miles and standard deviation 7000 miles.

$\bar{x} = 80,000$ $s = 7000$

- a. (2 pts) Give an interval where the number of miles until first major repair will fall about 95% of the time.

$(\bar{x} - 2s, \bar{x} + 2s) = (80,000 - 2(7000), 80,000 + 2(7000))$
 $= (66,000, 94,000)$

- b. (3 pts) What percent of cars will go less than 59,000 or more than 101,000 miles before needing the first major repair?

$(\bar{x} - 3s, \bar{x} + 3s) = (80,000 - 3(7000), 80,000 + 3(7000))$

$\approx (100 - 99.7)\% = 0.3\% = (59,000, 101,000)$ ~~to 99.7%~~

- c. (2 pts) What percent of cars will go over 101,000 miles before needing the first major repair?

$\frac{0.3\%}{2} = 0.15\%$

- d. (3 pts) For this part only, assume the distribution is extremely right skewed. At least what percent of the cars will go between 59,000 and 101,000 miles until the first major repair?

$(\bar{x} - 3s, \bar{x} + 3s)$ Chebychev's Rule
At least $1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} \approx 89\%$

3. The pulse rate in beats per minute for 20 patients in a hospital emergency room are shown below:

84, 86, 90, 90, 91, 94, 96, 96, 98, 99, 100, 102, 103, 103, 104, 106, 107, 109, 111, 121

Q_1 m Q_3

- a. (2 pts) Calculate the median.

$$m = \frac{99 + 100}{2} = 99.5$$

- b. (4 pts) Calculate the first and third quartiles.

$$Q_1 = \frac{91 + 94}{2} = 92.5$$

$$Q_3 = \frac{104 + 106}{2} = 105$$

- c. (3 pts) Calculate the upper and lower fences.

$$\text{lower} = Q_1 - 1.5(Q_3 - Q_1) = 92.5 - 1.5(105 - 92.5) = 92.5 - 18.75 = 73.75$$

$$\text{upper} = Q_3 + 1.5(Q_3 - Q_1) = 105 + 18.75 = 123.75$$

- d. (1 pts) Suppose a data set has 1023 values. What is the position of the first quartile? Once you calculate the position, explain how you use it to calculate the first quartile.

$$\text{position of } Q_1 = \frac{1}{4}(1024) = 256^{\text{th}}$$

- ① put data in order from lowest to highest value
- ② Q_1 is the value in the 256th position

4.

- a. (5 pts) Indicate whether each variable is qualitative (Q), quantitative-continuous (C) or quantitative-discrete (D).

i. Ethnicity

☒ Q C D

ii. Amount of time studying yesterday

☒ Q ☒ C D

iii. Blood type

☒ Q C D

iv. Favorite TV channel

☒ Q C D

v. Number of cars a person has owned

☒ Q C ☒ D

- b. (1 pts) Suppose you want to determine if there is a link between video games and aggressive behavior in children. For a sample of 100 children, you collect data on hours spent playing video games in a week and a numeric score on a test which measures aggressiveness (test scored from 1= no aggression to 10=high aggression). Indicate one type of graph would that help with this comparison.

Scatterplot

- c. (2 pts) Suppose you want to study the relationship between gender and amount of time spent playing video games. Name two types of graphs that would be appropriate for these data.

Side-by-side boxplots, side-by-side histograms

- d. (2 pts) The contingency table below summarizes data on gender and favorite movie genre for a group of students. Is there a difference in movie genre preferences by gender? Give reasons based on the data for your conclusion.

	Drama	Action	Comedy	Total
Male	5	30	15	50
Female	40	40	20	100

Conditional Probability
 Drama Act Comedy
 10% 60% 30%
 40% 40% 20% or percentages

Comparing the conditional probabilities, we see that there are differences.

5. Consider the bivariate data in the table below for this problem. X = number of alcoholic beverages consumed by subject, Y = amount of time for subject to push a button after buzzer sounds (in milliseconds).

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
1	15	$1 - 2.5 = -1.5$	$15 - 25 = -10$	15
2	13	$2 - 2.5 = -0.5$	$13 - 25 = -12$	$(-0.5)(-12) = +6$
3	32	$3 - 2.5 = 0.5$	$32 - 25 = 7$	$0.5(7) = 3.5$
4	40	$4 - 2.5 = 1.5$	$40 - 25 = 15$	$1.5(15) = 22.5$
		0	0	47

$$\bar{x} = \frac{10}{4} = \frac{5}{2} = 2.5$$

$$\bar{y} = \frac{100}{4} = 25$$

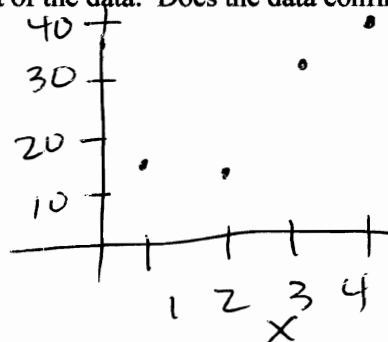
- a. (7 pts) Calculate the covariance for this data by hand. Use the formula $s_{xy} = \frac{\sum(x - \bar{x})(y - \bar{y})}{n - 1}$.

$$s_{xy} = \frac{47}{4 - 1} = \frac{47}{3} = 15.67$$

- b. (1 pts) Does the covariance indicate a positive, negative or nonexistent linear relationship?

positive

- c. (2 pts) Draw a scatterplot of the data. Does the data confirm your calculated covariance from part (a)? Explain why or why not.



The data confirm a positive covariance. as x increases, y also increases

6. (6 pts) Indicate whether the correlation coefficient is likely to be negative, positive or near zero for the following bivariate data.

a. x = area of a home in square feet, y = price of the home

positive

b. x = The number of units of coursework a Sacramento State student has completed, y = The number of remaining units of coursework the student has to finish to earn his/her bachelor's degree

negative

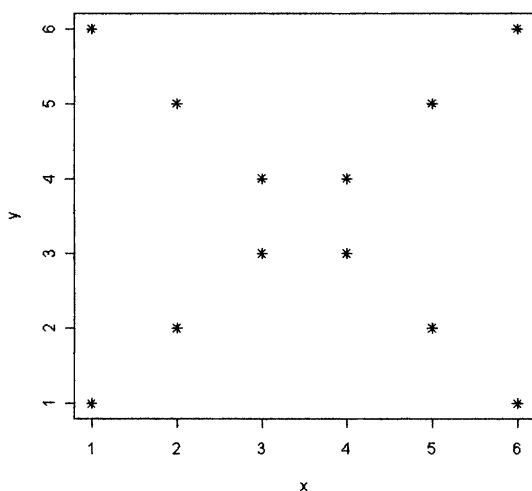
c. x = The number of letters in a student's first name, y = the student's grade point average

near zero

d. x = A dining group's total bill (in dollars) for a meal at a restaurant, y = the tip (in dollars) left for the server by the dining group

positive

e. For x and y shown in the scatterplot below:



near zero

f. What is the value of y that will make the correlation coefficient, r , equal to 1 for the following data set? $(1,2)$, $(2,4)$ and $(-5,y)$. (You shouldn't have to calculate r .) $y = \underline{-10}$

7. (10 pts) Indicate whether each statement is true (T) or false (F) or can't tell from the given information (C) by circling the appropriate letter. The given information about the data is the boxplot shown below.

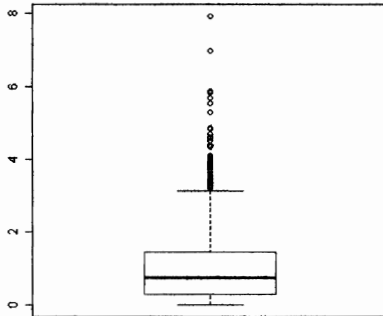
- a. The data are skewed right.
b. The upper fence is less than 4.
c. The mean is greater than the median.

<input checked="" type="radio"/> T	F	C
<input checked="" type="radio"/> T	F	C
<input checked="" type="radio"/> T	F	C

- d. The data between Q_1 and the median are more spread out than the data between the median and Q_3 .

T	<input checked="" type="radio"/> F	C
T	<input checked="" type="radio"/> F	C

- e. The Empirical rule would apply to this data set.



8. (4 pts) Economic data indicate that 50% of the world's population lives on \$2 or less per day (a sobering thought). Indicate whether each statement is true (T) or false (F) or can't tell from the given information (C). Read carefully.

- a. The median amount people live on per day is \$2.

<input checked="" type="radio"/> T	F	C
------------------------------------	---	---

- b. The mean amount people live on per day is \$2.

T	F	<input checked="" type="radio"/> C
---	---	------------------------------------

- c. The first quartile is less than \$2.

<input checked="" type="radio"/> T	F	C
------------------------------------	---	---

- d. The third quartile is larger than the mean.

T	F	<input checked="" type="radio"/> C
---	---	------------------------------------

9. Suppose k is a fixed but unspecified positive number.

- a. (4 pts) Show the two data sets have the same mean.

Data set 1: $-k, 0, k$

Data set 2: $-10k, 0, 10k$

$$\bar{x}_1 = \frac{-k + 0 + k}{3} = \frac{0}{3} = 0$$

$$\bar{x}_2 = \frac{-10k + 0 + 10k}{3} = \frac{0}{3} = 0$$

- b. (3 pts) Which data set has a larger standard deviation? Give a reason for your choice.

Data set 2 The values will be more spread out around 0 since k is multiplied by 10.

- c. (3 pts) Calculate the standard deviation of data set 2.

$$s^2 = \frac{(-10k - 0)^2 + (0 - 0)^2 + (10k - 0)^2}{3 - 1} = \frac{100k^2 + 100k^2}{2} = 100k^2$$

$$s = \sqrt{100k^2} = 10k$$