Statistics 1, Section 7 **Name** Solutions
Exam 1 **September 30, 2009**

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 cars sold by a car salesman on 6 days are:
   2 4 3 1 5 3

   Calculate:

   a. (4 points) Mean $\bar{x} = \frac{2+4+3+1+5+3}{6} = \frac{18}{6} = 3$ cars

   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.

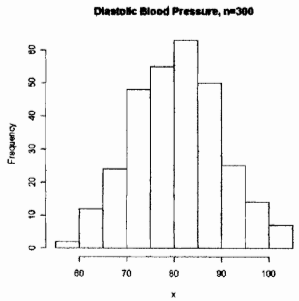
x	$x-\bar{x}$	$(x-\bar{x})^2$
2	$2-3=-1$	$(-1)^2=1$
4	$4-3=1$	$1^2=1$
3	$3-3=0$	$0^2=0$
1	$1-3=-2$	$(-2)^2=4$
5	$5-3=2$	$2^2=4$
3	$3-3=0$	$0^2=0$
	$\emptyset$	10

$$s = \sqrt{\frac{10}{6-1}} = \sqrt{\frac{10}{5}} = \sqrt{2} \approx 1.4$$

   c. (2 pts) What is the deviation from the mean of the data value 1? Indicate whether the deviation is positive or negative.

   $x-\bar{x} = 1-3 = -2$ negative

2. The diastolic blood pressure of 300 subjects has mean 80.7 mm/Hg and standard deviation 9.3 mm/Hg. A histogram of the 300 measured blood pressures is shown below:



$\bar{x} = 80.7$

$s = 9.3$

   a. (2 pts) The mean of this dataset is 80.7 mm/Hg and the standard deviation is 9.3 mm/Hg. Give an interval that contains about 99.7% subjects' diastolic blood pressures. Bell-shaped so Empirical Rule can be used.

$(\bar{x}-3s, \bar{x}+3s) = (80.7-3(9.3), 80.7+3(9.3))$ | $(52.8, 108.6)$

   b. (3 pts) What percent of blood pressures will fall below 71.4 mm/Hg or above 90 mm/Hg?

$(\bar{x}-s, \bar{x}+s) = (80.7-9.3, 80.7+9.3) = (71.4, 90)$ $\approx 68\%$

   c. (2 pts) What percent of the blood pressures will fall above 90 mm/Hg?

32%

Since the bell-shaped distn. is symmetric, 32% outside (71.4, 90)

So $\frac{1}{2}(32\%) = 16\%$

d. (3 pts) What does Chebychev's Rule state about the percent of blood pressures that will fall in the interval calculated in part (a)? Should we apply Chebychev's Rule to this data? Why or why not?

at least $1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} =$ [at least 89%] Don't use Chebychev's.

For bell-shaped data, Empirical Rule is more precise.

3. A data set consisting of 20 values is shown below. The data are in order from lowest to highest.

1, 1, 2, 4, 4 | 5, 8, 8, 8, 9, | 9, 11, 13, 13, 18, | 25, 32, 32, 61, 86

Q_1 m Q_3

a. (2 pts) Calculate the median. $m = \frac{9+9}{2} = 9$

b. (4 pts) Calculate the first and third quartiles. $Q_1 = \frac{4+5}{2} = 4.5$

$Q_3 = \frac{18+25}{2} = \frac{43}{2} = 21.5$

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

upper $= Q_3 + 1.5(Q_3 - Q_1) = 21.5 + 1.5(21.5 - 4.5) = 21.5 + 25.5 = 47$

lower $= Q_1 - 1.5(Q_3 - Q_1) = 4.5 - 25.5 = -21$

d. (1 pts) Using the fences, determine if any of the data values are outliers. If so, indicate which ones.

61 & 86 are outliers

4. A survey is conducted to study the commuters in a large company. The following data are collected for each employee: job classification (clerical, technical, manager), method of commute (bike, walk, auto, public transit, other), miles from home to workplace, amount of time to commute to work, and whether or not the employee is interested in reduced fee bus tickets. A portion of the resulting data frame is shown below.

Name	Job classification	Commute method	Miles from home to work	Amount of time to commute to work (minutes)	Interested in reduced fee bus tickets?
Elmer Fudd	Technical	Auto	3.5	5	Yes
Bugs Bunny	Clerical	Walk	2	15	No
:	:	:	:	:	:
:	:	:	:	:	:
John Smith	Manager	Bus	15	35	Yes

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

i. Job classification: (Q) C D
ii. Commute Method: (Q) C D
iii. Miles from home to work Q (C) D
iv. Amount of time to commute to work Q (C) D
v. Interested in reduced fee bus tickets? (Q) C D

b. (3 pts) You want to compare the miles from home to work across different job classifications.

i. Indicate one type of graph would that help with this comparison:

Side-by-side boxplots or side-by-side histograms

ii. Explain how you would use summary numbers to make this comparison.

	$\bar{x}$	s
Tech	23	10
Cler.	15	7
Manag.	7	5

Calculate the mean, $\bar{x}$, and standard deviation, s, for each job classification, the compare them.

c. (1 pts) What type of graph would help you understand the relationship between miles from home to work and amount of time to commute to work?

scatterplot

d. (1 pts) Suppose you want to study the relationship between commute method and job classification. What type of data summary would you use? contingency table or side-by-side pie charts

5. Consider the bivariate data in the table below for this problem. X = number of pairs of shoes owned by a subject, Y= number of times subject has flown in an airplane.

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$
8	15	$8-5=3$	$15-10=5$	$3(5)=15$
4	10	$4-5=-1$	$10-10=0$	$(-1)(0)=0$
1	13	$1-5=-4$	$13-10=3$	$(-4)(3)=-12$
7	2	$7-5=2$	$2-10=-8$	$2(-8)=-16$
Tot 20	40	0	0	-13

$\bar{x}=\frac{\Sigma x}{n}=\frac{20}{4}=5$ $\bar{y}=\frac{\Sigma y}{n}=\frac{40}{4}=10$

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

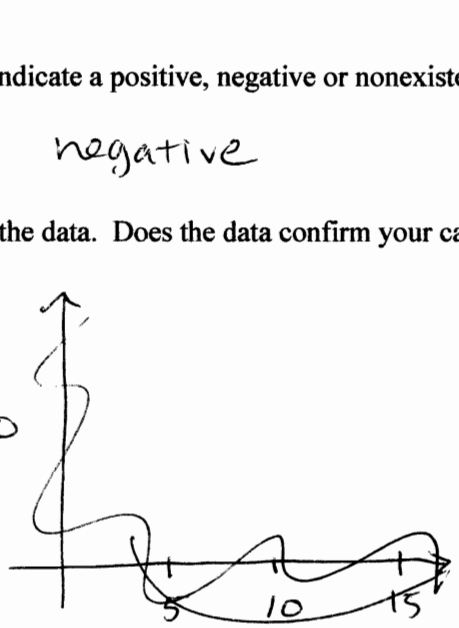
$$s_{xy}=\frac{-13}{4-1}=-\frac{13}{3}=-4.3$$

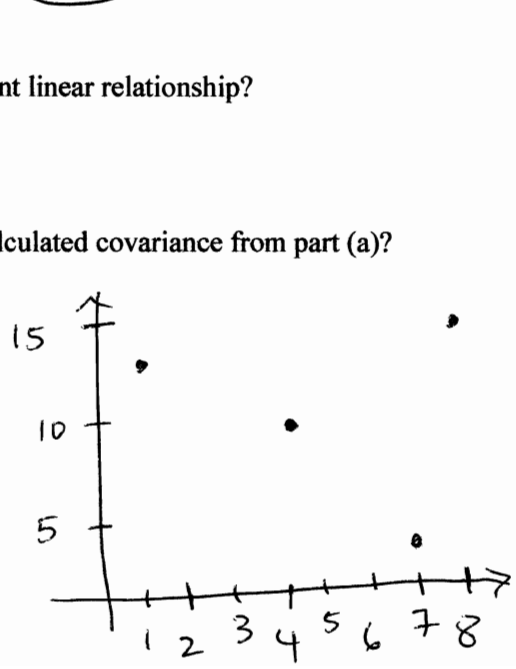
b. (1 pts) Does the covariance indicate a positive, negative or nonexistent linear relationship?

negative

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

Yes, all points except (8,15) ~~seem~~ seem to fit to a line with negative slope.





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

a. x = The amount of sugar in a snack, y = the number of calories in the snack

positive

b. x = The number of games a football team wins in a season, y = The number of games the football team loses in the same season

negative

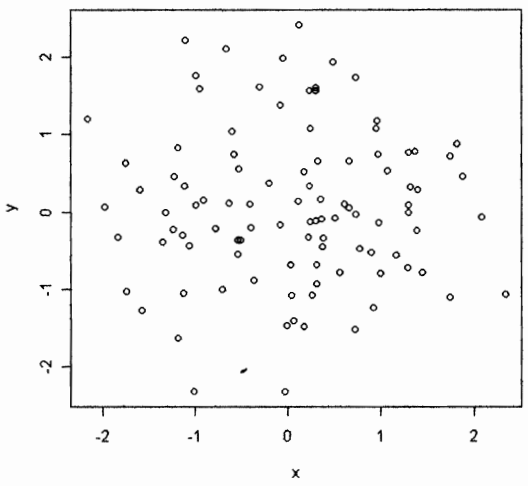
c. x = The last digit in a student's driver's license number, y = the student's score on this exam

near zero

d. x = The annual salary of a employee in the USA (in dollars), y = the amount of taxes paid by the employee (in dollars)

positive

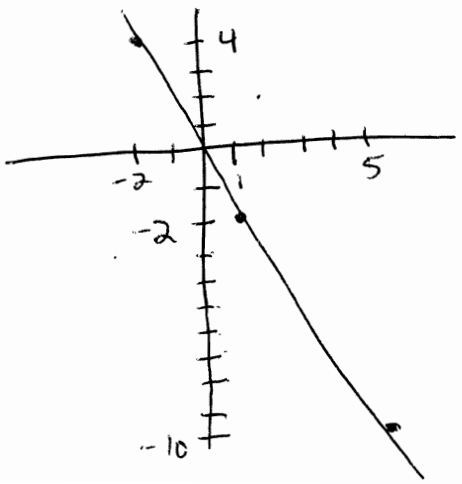
e. For x and y shown in the scatterplot below:



near zero

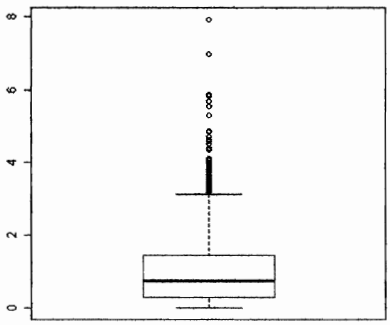
f. What is the correlation coefficient for the following dataset: (1,-2), (-2, 4) and (5,-10)? (You shouldn't need to do any calculations.) r = -1

points fall on a line with negative slope



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 to the left.	T	(F)	C
b. The z-score of the median is 0.	T	(F)	C
c. The mean is greater than the median.	(T)	F	C
d. The lower half of the data are more spread out than the upper half.	T	(F)	C
e. The Empirical rule would apply to this data set.	T	(F)	C



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.	(T)	F	C
b. The mean amount people live on per day is \$2.	T	F	(C)
c. The difference between the median and Q_1 is at most 2.	(T)	F	C
d. The 90th percentile is larger than the mean.	T	F	(C)

9. Suppose k is a fixed but unspecified number.

a. (4 pts) Calculate the mean of each data set. Since k is unknown the mean will be an expression involving k.

Data set 1: $k, k+1, k+2$ $\quad \bar{x}_1 = \frac{k+(k+1)+(k+2)}{3} = \frac{3k+3}{3} = \frac{3k}{3} + \frac{3}{3} = k+1$

Data set 2: $k-1, k, k+1$ $\quad \bar{x}_2 = \frac{(k-1)+k+(k+1)}{3} = \frac{3k}{3} = k$

b. (3 pts) Which data set has a larger mean? Which data set has a larger standard deviation?

Data set (1) $\qquad$ same

c. (3 pts) Calculate the standard deviation of data set 1.

x	$\bar{x}-\bar{x}$	$(x-\bar{x})^2$
k	$k-(k+1) = -1$	$(-1)^2 = 1$
$k+1$	$(k+1)-(k+1) = 0$	$0^2 = 0$
$k+2$	$(k+2)-(k+1) = k+2-k-1 = 1$	$1^2 = 1$

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

$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{2}{3-1}} = \sqrt{\frac{2}{2}} = (1)$