Statistics 1, Section 6 **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 library books checked out by 5 students at the Sac State library are:
   5 1 4 7 3

   Calculate:

   a. (4 pts) Mean $\bar{x} = \frac{5+1+4+7+3}{5} = \frac{20}{5} =$ 4 books

   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	$x-\bar{x}$	$(x-\bar{x})^2$
5	$5-4=1$	1
1	$1-4=-3$	9
4	$4-4=0$	0
7	$7-4=3$	9
3	$3-4=-1$	1
		$20 = \sum(x-\bar{x})^2$

$$s = \sqrt{\frac{20}{5-1}} = \sqrt{\frac{20}{4}} = \sqrt{5} = 2.236 \text{ books}$$

   c. (2 pts) What is the deviation from the mean of the data value 1? Indicate whether the deviation is positive or negative.
   $x - \bar{x} = 1-4 =$ (−3)

2. In Clark County, the mean monthly electricity bill for a household is \$85 with a standard deviation of \$30. Suppose the distribution of monthly electricity bills is bell-shaped.

   a. (2 pts) Give an interval that contains about 68% of the household's electricity bills in Clark County.
   $(\bar{x}-s, \bar{x}+s) = (85-30, 85+30)$
   $= (\$55, \$115)$

   b. (3 pts) What percent of electricity bills will be ***below*** \$25 or ***above*** \$145?
   $(\bar{x}-2s, \bar{x}+2s) = (85-2(30), 85+2(30))$
   $= (\$25, \$145)$ By the Empirical Rule, about 95% are in (25, 145)
   Thus, 100% − 95% = 5% below 25 or above 145

   c. (2 pts) What percent of bills will be above \$145?
   By symmetry of the bell-shaped distribution $\frac{1}{2}(5\%) =$ 2.5%

   d. (3 pts) For this part only, assume the distribution is extremely right skewed. At least what percent of the bills will be between \$25 and \$145?
   $\bar{x} \pm 2s$
   By Chebychev's at least
   $1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} =$ 75%

3. The exam scores of 16 students are sorted from lowest to highest and shown below.

39, 57, 59, 59 | 61, 63, 66, 68 | 69, 70, 71, 74 | 79, 82, 84, 93

a. (2 pts) Calculate the median. $m = \frac{68+69}{2} = 68.5$

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

$Q_1 = \frac{59+61}{2} = 60$

$Q_3 = \frac{74+79}{2} = 76.5$

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

$\text{lower} = Q_1 - 1.5(Q_3 - Q_1) = 60 - 1.5(76.5 - 60) = 60 - 24.75 = 35.25$

$\text{upper} = Q_3 + 1.5(Q_3 - Q_1) = 76.5 + 24.75 = 101.25$

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

No outliers.

4. A survey is conducted to study computer usage among college students. The following data are collected for each student: class level (freshman, sophomore, junior, senior), academic major, whether or not the student owns a computer, the number of hours per week of computer use, and the number of emails received yesterday. A portion of the resulting data frame is shown below.

Name	Class level	Major	Own computer?	Hours of computer use per week	Number of e-mails received yesterday
Elmer Fudd	Freshman	Wildlife Biology	No	1.5	3
Bugs Bunny	Senior	Engineering	Yes	10	15
:	:	:	:	:	:
:	:	:	:	:	:
John Smith	Junior	Business	Yes	5	10

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

i. Class level: (Q) C D
ii. Major: (Q) C D
iii. Own computer? (Q) C D
iv. Hours of computer use per week Q (C) D
v. Number of e-mails received yesterday Q C (D)

b. (3 pts) You want to compare the hours of computer use per week across different majors.

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.

Calculate $\bar{x}$ and s for each group and compare

Ex.

Major	Business	Nursing	Biology
$\bar{x}$	5	3.2	4.1
s	0.8	1.0	0.5

c. (1 pts) What type of graph would help you understand the relationship between hours of computer use per week and number of e-mails received yesterday?

scatterplot

d. (1 pts) Suppose you want to study the relationship between computer ownership and class level. What type of data summary would help in your investigation?

contingency table

computer	Fr	So	Jr	Sr
Y				
N				

5. Consider the bivariate data in the table below for this problem. X = weight of car in thousands of pounds, Y= gas mileage of car (in miles per gallon).

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$
5	15	$5-3=2$	$15-21=-6$	-12
2	21	$2-3=-1$	$21-21=0$	0
4	13	$4-3=1$	$13-21=-8$	-8
1	35	$1-3=-2$	$35-21=14$	-28
Total 12	84	0		~~-32~~ -48

$\bar{x} = \frac{12}{4} = 3$ $\quad \bar{y} = \frac{84}{4} = 21$

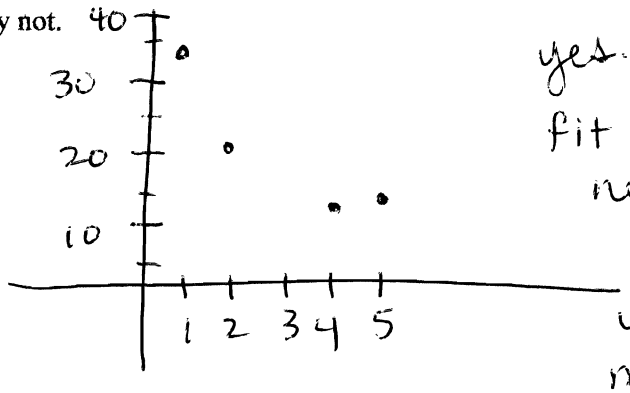
a. (7 pts) Calculate the covariance for this data by hand. Use the formula $s_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$.

$$s_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1} = \frac{-48}{4-1} = \frac{-48}{3} = -16$$

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. the scatterplot fit to a line with negative slope. as weight of car increases, gas mileage increases, consistent with negative correlation.

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 number of empty seats in our classroom, y = the number of occupied seats in our classroom

negative.

b. x = The number of points a pro basketball player makes in a season, y = The salary of a pro basketball player

positive.

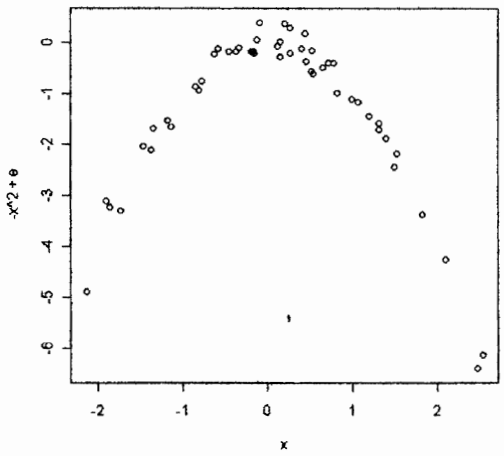
c. x = The number of pairs of socks a student owns, y = the student's score on this exam

near zero

d. x = The number of exams a student has in a week, y = the hours of leisure time a student has in the same week

negative.

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



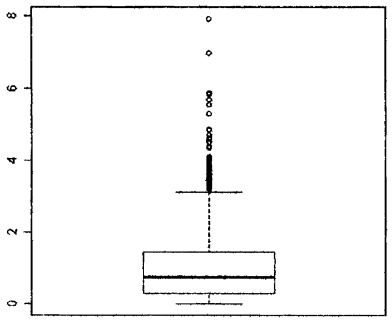
near zero

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

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. skewed right!	T	(F)	C
b.	Do not contain outliers.	T	(F)	C
c.	The mean is greater than the median. due to right skewness	(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

↓
Bell-shaped must be symmetric. This data is skewed.



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 first quartile is between \$0 and \$2.	(T)	F	C
d.	The third quartile is larger than the mean.	T	F	(C)

9. Suppose k is a fixed but unspecified number.

a. (4 pts) Show the two data sets have the same mean. (Hint: The mean will not be a number, it will be an algebraic expression involving the variable k.)

Data set 1: k-1, k, k+1

$$\bar{x}_1 = \frac{(k-1)+k+(k+1)}{3} = \frac{3k}{3} = k$$

Data set 2: k-100, k, k+100

$$\bar{x}_2 = \frac{(k-100)+k+(k+100)}{3} = \frac{3k}{3} = k$$

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

Data Set 2 is more spread out!

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

x	$x-\bar{x}$	$(x-\bar{x})^2$
$k-100$	$(k-100)-k=-100$	$(-100)^2=10{,}000$
k	$k-k=0$	0
$k+100$	$(k+100)-k=100$	$(100^2)=10{,}000$
		$20{,}000$

$$s^* = \sqrt{\frac{20{,}000}{3-1}} = \sqrt{10{,}000} = 100$$