POS 3713 Fall 1999
Homework #4 Solutions

1. After Bob determines that eating chocolate donuts makes him run faster, he tells his friend Sherry to use the same training strategy to prepare for the 5-K race. Sherry is skeptical of Bob's results, so she decides to gather her own data. Below is a list of the number of chocolate donuts she ate before 7 practices and the corresponding recorded time. Use this information to answer the following questions below.

# of Chocolate Donuts Consumed	Recorded Running Time
3	26.30
5	27.27
7	28.11
10	29.36
13	33.02
17	34.50
20	35.10

a. What is the independent variable in this example? What is the dependent variable?
Independent: # of Chocolate Donuts Consumed; Dependent: Running Time

b. Calculate the correlation, or Pearson's *r*, between the number of donuts consumed and the recorded running time.

$$\mathbf{r = \frac{\Sigma(X - Xbar)(Y - Ybar)}{\sqrt{\Sigma(X - Xbar)^2 \cdot \Sigma(Y - Ybar)^2}} = \frac{134.16}{\sqrt{(237.43)(78.6)}} = \frac{134.16}{136.6} = \mathit{0.982}}$$

$$\mathbf{r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} = \frac{7(2423.38) - (75)(213.66)}{\sqrt{[7(1041) - (75)^2][7(6600.1) - (213.66)^2]}}}$$

$$\mathbf{= \frac{16963.66 - 16024.5}{\sqrt{(7287 - 5625)(46200.7 - 45650.6)}} = \frac{939.16}{956.2} = \mathit{0.982}}$$

c. Interpret your results. Do Sherry's results contradict Bob's original conclusion that the more donuts he eats, the faster he runs?
There is a strong, positive correlation between the number of donuts Sherry eats and her recorded running time, which means that the more donuts she eats, the higher her running time (or the slower she runs). Thus her results do contradict Bob's findings.

d. Calculate the regression line (i.e., find *a* and *b* in the equation Y = *a* + *b*X, using the formulas for *a* and *b* discussed in class). How do you interpret *b*? Is this interpretation consistent with what you found in c?

$$\mathbf{b = \frac{\Sigma(X - Xbar)(Y - Ybar)}{\Sigma(X - Xbar)^2} = \frac{134.16}{237.43} = 0.565}$$

a = Ybar - b(Xbar) = 30.52 - (0.565)(10.7) = 24.47

Regression Line: Y = 24.47 + 0.565X

Interpretation: For each additional donut Sherry eats, her running time increases by 0.565 minutes. Yes, this is consistent with what we found in c (a strong, positive correlation).

2. Consider the following regression output regarding crime rate in the 50 U.S. states. The dependent variable is the number of reported crimes/100,000 for each state.

Dependent Variable: CRIME RATE

Independent Variables	Mean	Standard Deviation	Description
URBAN	61.368	22.845	% of population urban
UNEM	7.312	1.905	% labor force unemployed
COPS	2.512	0.488	# of police per 1000
FMLINC	19550.3	2552.06	median state family income
BROKEN	15.684	2.898	% kids in 1-parent homes
POORKID	15.446	4.598	% kids below poverty

N = 50

Independent Variable	Estimated Coefficient (b)	Standard error (s.e.)
POORKID	-91.5	50.2
UNEM	6.12	71.2
URBAN	19.9	8.7
COPS	1504.5	415.1
BROKEN	77.4	79.9
FMLINC	-.10	.09

a) Which independent variables possess a positive relationship with crime rate? Which independent variables possess a negative relationship with crime rate? Does the direction of these relationships make sense, i.e., are the coefficients in the direction you would expect based on what you know about crime?

Positive Relationship: UNEM, URBAN, COPS, BROKEN
Negative Relationship: POORKID, FMLINC

I would expect COPS to be negative rather than positive (more cops on the street should decrease crime) and I would expect POORKID to be positive, i.e., as the % of children in poverty increases, crime should increase.

b) Which variables are statistically significant at the 0.05 level? Use a one-tailed test.

Df = N - K - 1 = 50 - 6 - 1 = 43; Critical t = +/- 1.684
Compare calculated t's to critical t; reject if calculated t exceeds critical t
H_o: β=0, H_1: β > or < 0 (depending on sign of estimate)

Variable	**Calculated t=b/s.e**	**Decision**
POORKID	**-91.5/50.2 = -1.82**	**Reject H_o**
UNEM	**6.12/71.2 = 0.086**	**Accept H_o**
URBAN	**19.9/8.7 = 2.29**	**Reject H_o**
COPS	**1504.5/415.1 = 3.62**	**Reject H_o**
BROKEN	**77.4/79.9 = 0.97**	**Accept H_o**
FMLINC	**-.10/.09 = -1.11**	**Accept H_o**

Thus POORKID, URBAN, and COPS are statistically significant at the .05 level.

3. David Romer, a professor in economics, was interested in testing the relationship between students' attendance in class and their final grades. He wanted to assess the extent to which class attendance improved the final course grade. In the fall 1990 semester, he took attendance at 6 meetings of his large intermediate macroeconomics course at University of California, Berkeley. The overall absenteeism was 25%. He measured student performance as the overall score on the three exams in the course, converted to the 4 point grading scale (3.84 and higher, A; 3.5-3.83, A-; 3.16-3.49, B+; 2.83-3.15, B; 2.50-2.82, B-; 2.17-2.49, C+; 1.84-2.16, C; 1.5-1.83, C-; 1.17-1.49, D; 1.16 and below, F). Attendance is measured as the proportion of classes a student attended. He reports the following regression results (Y = a + bX) in his article.

```
      a      b
Y = 1.25 + 2.19X        where Y = student's grade, X = proportion of classes attended
s.e.  (.27)   (.35)
```

N = 195

a. Does attendance significantly *increase* a student's final course grade at the α = .05 level? State the null and alternative hypotheses.

   H_o: $\beta = 0$
   H_1: $\beta > 0$

   t = b/s.e. = 2.19/.35 = 6.26
   Critical t = 1.645 (df = N - K - 1 = 195 - 1 - 1 = 193)
   We can reject H_o because 6.26 > 1.645 and conclude that attendance does significantly increase a student's final course grade.

b. What would you predict a student's grade would be if they attended 1/4 of the lectures (Attendance = .25)? What would you predict a student's grade would be if they attended all of the lectures (Attendance = 1)? As a student, do you perceive this difference to be significant? (Hint: use the range of grades to answer this question; A; 3.5-3.83, etc.)

   For X = .25
   Y = 1.25 + 2.19(.25) = 1.798 Grade: C

For X = 1
Y = 1.25 + 2.19(1) = 3.44 Grade: A-

Yes, there is a significant difference because the student's grade increases by almost 2 full grades if he/she attends all of the lectures.