STA 6127: Exercises for Reviewing Basic Regression

1. Anthropologists often try to reconstruct information using partial human remains at burial sites. For instance, after finding a femur (thighbone), they may want to predict how tall an individual was. An equation they use to do this is the regression line, $\hat{y} = 61.4 + 2.4x$, where $\hat{y}$ is the predicted height and x is the length of the femur, both in centimeters (S. Junger, *Vanity Fair*, October 1999).
   a) Identify the y-intercept and slope of the equation. Interpret the slope.
   b) A femur found at a particular site has length of 50 cm. What is the predicted height of the person who had that femur?

2. The OECD (Organization for Economic Cooperation and Development) consists of 20 advanced, industrialized countries. For these nations, the prediction equation relating Y = child poverty rate in 2000 to X = social expenditure as a percent of gross domestic product is $\hat{y} = 22 - 1.3x$. The y-values ranged from 2.8% (Finland) to 21.9% (U.S.). The x-values ranged from 2% (U.S.) to 16% (Denmark).
   a) Interpret the y-intercept and the slope.
   b) The correlation is -0.79. Interpret.

3. For recent UN data from 39 countries on Y = per capita carbon dioxide emissions (metric tons per capita) and X = per capita gross domestic product (GDP, in dollars), the prediction equation was $\hat{y} = 1.26 + 0.346x$.
   a) For the U.S., $x = 34.3$ and $y = 19.7$. Find the predicted CO2 value. Find the residual, and interpret.
   b) For Switzerland, $x = 28.1$ and $y = 5.7$. Find the predicted CO2 value and residual. Interpret.

4. For students who take Statistics 101 at Lake Wobegon College in Minnesota, both X = midterm exam score and Y = final exam score have mean = 75 and standard deviation = 10.
   a) The prediction equation is $\hat{y} = 30 + 0.60x$. Find the predicted final exam score for a student who has (i) midterm score = 100, (ii) midterm score = 50. Note that the predicted final exam score regresses toward the mean.
   b) Show that the correlation equals 0.60.
   c) If instead, $\hat{y} = x$, show that $r = 1.0$
   d) If instead, $\hat{y} = 75$, show that $r = 0.0$.

5. Is political ideology associated with income? When GSS data for 779 cases in 2004 were used to regress Y = political views (POLVIEWS, using scores 1-7 with 1 = extremely liberal and 7 = extremely conservative) on X = respondent's income (RINCOME, using scores 1-12 for the 12 income categories) on , we get the results shown in Table 1.
   a) Show all steps of the test of the hypothesis that political views are independent of income. Interpret.
   b) Construct a 95% confidence interval for the population slope. Interpret.
   c) What is SPSS reporting under "Beta" in this printout? How would you interpret this value?

Table 1: Political Views Regressed on Income

R Square 0.00024

Variable	B	SE B	Beta	T	Sig T
Constant	4.12668	0.18271		22.58	0.000
RINCOME	0.00739	0.01706	0.01554	0.43	0.665

6. Refer to the previous exercise. When political ideology is regressed on X = number of hours spent in the home on religious activity in the past month (RELHRS1), we obtain:

```
-------------------------------------------------------
  Variable     B       SE B     Beta        T     Sig T
  Constant   4.0115   0.0422            95.10   0.0000
  RELHRS1    0.0064   0.0020    0.087    3.20   0.0015
-------------------------------------------------------
```

a) Report and interpret the P-value for testing the hypothesis that these variables are independent.
b) Use these results to illustrate that statistical significance does not imply practical significance.

7. The headline of an article in the Gainesville Sun (October 17, 2003) stated "Height can yield a taller paycheck." It described an analysis of four large studies in the U.S. and Britain by a University of Florida professor on subjects' height and salaries. The article reported that for each gender, "an inch is worth about $789 a year in salary. So, a person who is 6 feet tall will earn about $5,523 more a year than a person who is 5 foot 5."
a) For the interpretation in quotes, identify the response variable and explanatory variable, and state the slope of the prediction equation, when height is measured in inches and income in dollars.
b) Explain how the value $5,523 relates to the slope.

8. Annual income, in dollars, is the response variable in a regression analysis. For a British version of the report on the analysis, all responses are converted to British pounds sterling (1 pound equals about 1.9 dollars, as of 2007).
a) How, if at all, does the slope of the prediction equation change?
b) How, if at all, does the correlation change?

9. For a class of 100 students, the teacher takes the 10 students who perform poorest on the midterm exam and enrolls them in a special tutoring program. The overall class mean is 70 both on the midterm and final, but the mean for the specially tutored students increases from 50 to 60. Can we conclude that the tutoring program was successful? Explain.

10. * Refer to the formula $a = \bar{y} - b\bar{x}$ for the y-intercept.
a) Show that substituting $x = \bar{x}$ into the prediction equation $\hat{y} = a + bx$ yields the predicted y-value of $\hat{y} = \bar{y}$. Show that this means that the least squares prediction equation passes through the point with coordinates $(\bar{x}, \bar{y})$, the center of gravity of the data.
b) Show that an alternative way of expressing the regression model is as $(\hat{y} - \bar{y}) = b(x - \bar{x})$.
c) Let Y = final exam score and X = midterm exam score. Suppose the correlation is 0.70 and the standard deviation is the same for each set of scores. Show that $(\hat{y} - \bar{y}) = 0.70(x - \bar{x})$; that is, the predicted difference between your final exam grade and the class mean is 70% of the difference between your midterm exam score and the class mean, so your score is predicted to regress toward the mean.