Stat 160 Fall 2008
Solutions to Assignment #4

1. (a) In this problem, *year* is the explanatory variable and *population* is the response variable. Hence, if we denote *year* by x, then

$$\overline{x} = \frac{1952 + 1962 + 1972 + 1982 + 1992 + 2002}{6} = 1977$$

and

$$s_x = \sqrt{\frac{(1952 - 1977)^2 + \cdots + (2002 - 1977)^2}{5}} = \sqrt{350} = 18.70829.$$

Furthermore, if we denote *population* by y, then

$$\overline{y} = \frac{3157000 + 5368613 + 7684783 + 10492578 + 13552830 + 16769878}{6} = 9504280$$

and

$$s_y = \sqrt{\frac{(3157000 - 9504280)^2 + \cdots + (16769878 - 9504280)^2}{5}} = 5115832.$$

The correlation is given by

$$r = \frac{1}{5}\left[\frac{1952 - 1977}{18.70829} \cdot \frac{3157000 - 9504280}{5115832} + \cdots + \frac{2002 - 1977}{18.70829} \cdot \frac{16769878 - 9504280}{5115832}\right]$$
$$= 0.9970367.$$

Since the least squares regression line is given by $\hat{y} = a + bx$ where

$$b = r\frac{s_y}{s_x} = 0.9970367 \times \frac{5115832}{18.70829} = 272642.4$$

and

$$a = \overline{y} - b\overline{x} = 9504280 - 1977 \times 272642.4 = -529509745$$

we conclude that the required equation is

$$\hat{y} = -529509745 + 272642.4x.$$

1. (b) Using our regression line from **(a)**, we predict Florida's population in 2007 to be

$$\hat{y} = -529509745 + 272642.4 \times 2007 = 17683552.$$

It is interesting to note that data from Florida's Office of Economic & Demographic Research lists the 2007 population as 18717990. Our linear model is incredibly accurate!

#5.32 Recall that r^2 is the fraction of the variation in the values of the response variable (in this case, grade) that is explained by the least-squares regression on the explanatory variable (in this case, attendance). Since $r^2 = 0.16$, the two possible values for the correlation are $\sqrt{0.16} = \pm 0.40$. We know that high attendance goes with high grades so the correlation must be positive, and thus we conclude that $r = 0.40$.

#5.50 One example of how grade inflation in high school might account for this pattern is as follows. The student who in the past might have received a grade of B (and a lower SAT score) now receives an A (but has a lower SAT score than an A student in the past). While this is a bit of an oversimplification, this means that today's A students are yesterday's A and B students, today's B students are yesterday's C students, and so on. Because of the grade inflation, we are not comparing students with equal abilities in the past and today.

#8.28 This is an experiment, because the treatment is selected (randomly, we assume) by the interviewer. The explanatory variable (treatment) is the level of identification, and the response variable is whether or not the interview is completed.

#8.46 (a) The wording is clear, but will almost certainly be slanted toward a high positive response. (Would anyone hear the phrase "brain cancer" and not be inclined to agree that a warning label is a good idea?)

#8.46 (b) The question makes the case for a national health care system, and so will slant responses toward "yes."

#8.46 (c) This survey question is most likely to produce a response similar to: "Uhh...yes? I mean, no? I'm sorry, could you repeat the question?" (And, if the person is able to understand the question, it is slanted in favour of day care subsidies.)