PLSC 503: Problem Set 1

Thad Dunning

Department of Political Science

Yale University

Spring 2010

Due in class, Tuesday, January 19, 2010

The following facts may be useful in working these problems. First, for a constant b,

$$\sum_{i=1}^{n} b = nb, \tag{1}$$

$$\frac{1}{n}\sum_{i=1}^{n} b = b, \tag{2}$$

and

$$\sum_{i=1}^{n} bx_i = b\sum_{i=1}^{n} x_i. \tag{3}$$

Also, recall that

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i. \tag{4}$$

Finally, remember the definitions of variance and covariance presented in lecture.

Questions:

1. Show that if $z_i = a + bx_i$ for all i, then $\bar{z} = a + b\bar{x}$.

   (Hint: write

   $$\frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}(a + bx_i) \tag{5}$$

   and then use the rules above).

   How does this relate to the fact that the regression line passes through the point of averages?

2. Show that adding a constant does not change the variance: if $z_i = x_i + d$ for all i, then $\text{var}(z) = \text{var}(x)$. (Hint: what is $z_i - \bar{z}$, expressed in terms of x_i, $\bar{x}$, and d?).

3. (a) Let $z_i = cx_i + d$ for all i. Show that $\text{Var}(z) = c^2\text{Var}(x)$.

   (Hints: use 1 and 2 to rewrite $z_i - \bar{z}$. Then, use the definition of variance, substitute for $z_i - \bar{z}$, and multiply out terms. You will also need to use the alternate definition of variance presented in class).

(b) Now, recall that the equation for the regression line is $\hat{y}_i = a + bx_i$. So what is $\text{var}(\hat{y}_i)$?

4. Let $z_i = au_i + bv_i$ for all subjects i. Show that

$$\text{Var}(z) = a^2\text{Var}(u) + 2ab\text{Cov}(u, v) + b^2\text{Var}(v). \tag{6}$$

(Hints: You will need to use the alternate definitions of variance and covariance used in lecture, as well as your results above).

5. Let the variable z_i be x_i in standard units:

$$z_i = \frac{(x_i - \bar{x})}{\text{SD}_x}, \tag{7}$$

where SD_x is the standard deviation of x. Show that

$$\bar{z} \equiv \frac{1}{n}\sum_{i=1}^{n} z_i = 0 \tag{8}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 = 1. \tag{9}$$

(Note: "≡" means "defined as" or "identically equal to").

In other words, when a variable is converted to standard units,

(a) its average is 0;

(b) its variance and standard deviation are equal to 1.

(Hints: Use the properties of summation defined above. Also, remember that $(\text{SD}_x)^2 = \text{var}(x)$, that is, the variance is the square of the standard deviation).

6. **Hooke's Law** Hooke's law states that when a load (weight) is placed on a spring, the length is proportional to the weight. That is,

$$\text{length under load} = \text{length under no load} + \text{constant} \cdot \text{load}$$

Physicists test this prediction in the lab and obtain the following results:

Load (kg)	Length (cm)
0	287.12
1	287.18
1	287.16
3	287.25
4	287.33
4	287.35
6	287.40
12	287.75

(a) The two lengths for a load of 1 kg differ. Why might that be?

(b) Find the regression equation for predicting length from load. (Show your work, including calculations of the relevant variances and covariances!).

(c) Use the equation to predict length at the following loads: 2 kg, 3 kg, 5 kg, 105 kg.

(d) For a load of 3 kg, the answer to (c) is different from the number in the table. Under the load of 3 kg, would you use the number in the table, or the regression equation? Explain carefully. (You may want to refer to Chapter 12 of Freedman, Pisani, and Purves, 2007).

(e) Estimate the length of the spring under no load.

(f) Estimate the constant (i.e., the slope coefficient) in Hooke's law.