PLSC 503: Solutions to Problem Set 1

Thad Dunning

Department of Political Science

Yale University

Spring 2010

1. **Question**: Show that if $z_i = a + bx_i$ for all i, then $\bar{z} = a + b\bar{x}$. How does this relate to the fact that the regression line passes through the point of averages?

   Solution: The first part follows directly from the hint:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} z_i &= \frac{1}{n}\sum_{i=1}^{n}(a + bx_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} a + \frac{1}{n}\sum_{i=1}^{n} bx_i \\ &= a + b\frac{1}{n}\sum_{i=1}^{n} x_i \\ &= a + b\bar{x}. \end{aligned} \quad (1)$$

   Notice that $y_i = a + bx_i + e_i$ for all i, where a and b are the fitted intercept and slope of the regression line and e_i is the residual for unit i. Thus, generalizing (1), we have that

$$\bar{y} = a + b\bar{x} + \bar{e} \implies \bar{y} = a + b\bar{x}, \quad (2)$$

   since the average of the residuals is zero. Moreover, $\hat{y}_i = a + bx_i$, so $\bar{\hat{y}}_i = a + b\bar{x}$ and thus $\bar{\hat{y}}_i = \bar{y}$. Therefore, the regression line passes through the point of averages.

2. **Question**: Show that adding a constant does not change the variance: if $z_i = x_i + d$ for all i, then

$$\text{var}(z) = \text{var}(x). \quad (3)$$

   Solution: First, since $\bar{z} = \bar{x} + d$ (see above) and $z_i = x_i + d$,

$$z_i - \bar{z} = (x_i + d) - (\bar{x} + d) = x_i - \bar{x}. \quad (4)$$

Thus, squaring both sides, summing over all i, and dividing through by n, we have

$$\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x)^2, \tag{5}$$

that is, $\text{var}(z) = \text{var}(x)$.

3. **Question**

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

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

   Solution:

   (a) First, note that since $\bar{z} = c\bar{x} + d$, we have

$$\begin{aligned} z_i - \bar{z} &= (cx_i + d) - (c\bar{x} + d) \\ &= c(x_i - \bar{x}). \end{aligned} \tag{6}$$

Thus,

$$\begin{aligned} \text{Var}(z) &= \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}[c(x_i - \bar{x})]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}[c^2x_i^2 - 2c^2x_i\bar{x} + c^2\bar{x}^2] \\ &= \frac{1}{n}\sum_{i=1}^{n}[c^2x_i^2 - c^2\bar{x}^2] \\ &= c^2\frac{1}{n}\sum_{i=1}^{n}[x_i^2 - \bar{x}^2] \\ &= c^2\text{Var}(x), \end{aligned} \tag{7}$$

which completes the solution. In moving from the third to the fourth line of (7), we use the fact that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} 2c^2 x_i \bar{x} &= 2c^2 \bar{x}\frac{1}{n}\sum_{i=1}^{n} x_i \\ &= 2c^2(\bar{x})^2. \end{aligned} \tag{8}$$

In turn, the first line of (8) uses the fact that $2c^2\bar{x}$ does not depend on the index of summation (so we can take this term outside the summation sign), and the second line uses the definition of the average, $\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$.

(b) Since $\hat{y}_i = a + bx_i$ for all i, the result above implies that $\text{var}(\hat{y}_i) = b^2\text{var}(x_i)$.

4. **Question**: Let $z_i = au_i + bv_i + c$ for all subjects i. Show that

$$\text{Var}(z) = a^2\text{Var}(u) + 2ab\text{Cov}(u, v) + b^2\text{Var}(v). \tag{9}$$

Solution: First, by the solution to Question 2, z_i has the same variance as $z_i' = au_i + bv_i$, because the addition of the constant c does not change the variance. Thus, we must simply show that

$$\text{Var}(z') = a^2\text{Var}(u) + 2ab\text{Cov}(u, v) + b^2\text{Var}(v). \tag{10}$$

Note also, by a generalization of the solution to Question 1, that $\bar{z}' = a\bar{u} + b\bar{v}$.

Now, we can write the variance as

$$\text{Var}(z') = (\frac{1}{n}\sum_{i=1}^{n}(z_i')^2) - (\bar{z})^2, \tag{11}$$

that is, the variance is the average of the squares minus the square of the average. (Refer to the lecture notes or to Freedman and Lane (1981: p. 28). Then, we simply substitute for z_i'

and $\bar{z}$ in equation (11) and multiply and rearrange terms to complete the solution:

$$\begin{aligned}
\mathrm{Var}(z') &= (\frac{1}{n}\sum_{i=1}^{n}(au_i + bv_i)^2) - (a\bar{u} + b\bar{v})^2 \\
&= (\frac{1}{n}\sum_{i=1}^{n}[a^2u_i^2 + 2abu_iv_i + b^2]) - a^2\bar{u}^2 - 2ab\bar{u}\bar{v} - b^2\bar{v}^2 \\
&= a^2(\frac{1}{n}\sum_{i=1}^{n}u_i^2 - \bar{u}^2) + 2ab(\frac{1}{n}\sum_{i=1}^{n}u_iv_i - \bar{u}\bar{v}) + b^2(\frac{1}{n}\sum_{i=1}^{n}v_i^2 - \bar{v}^2) \\
&= a^2\mathrm{Var}(u) + 2ab\mathrm{Cov}(u,v) + b^2\mathrm{Var}(v) \qquad (12)
\end{aligned}$$

In the final line, we made use again of the alternate way of writing variances, and also the alternate way of writing covariances (see the lecture notes or Freedman and Lane p. 43, proposition 2). This completes the solution.

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

$$z_i = \frac{(x_i - \bar{x})}{\mathrm{SD}_x}, \qquad (13)$$

where SD_x is the standard deviation of x. Show that

$$\bar{z} \equiv \frac{1}{n}\sum_{i=1}^{n} z_i = 0 \qquad (14)$$

and

$$\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 = 1. \qquad (15)$$

(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.

Solution: Let z_i be x_i in standard units:

$$z_i = \frac{(x_i - \bar{x})}{\mathrm{SD}_x} \tag{16}$$

(a) First, we want to prove that the average $\bar{z} = 0$. By the definition of the average,

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i. \tag{17}$$

Then, substituting terms,

$$\begin{aligned} \bar{z} &= \frac{1}{n}\sum_{i=1}^{n}[\frac{(x_i - \bar{x})}{\mathrm{SD}_x}] \\ &= \frac{1}{\mathrm{SD}_x}\frac{1}{n}\sum_{i=1}^{n}[x_i - \bar{x}] \\ &= \frac{1}{\mathrm{SD}_x}[\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n}\bar{x}] \\ &= \frac{1}{\mathrm{SD}_x}[\bar{x} - \bar{x}] \\ &= 0. \end{aligned} \tag{18}$$

The second line of (18) uses the distributive law, while the third uses the associative law and the fact that

$$\frac{1}{n}\sum_{i=1}^{n}\bar{x} = \frac{n\bar{x}}{n} = \bar{x}. \tag{19}$$

Thus, when converted to standard units, a variable has average equal to 0.

(b) Second, by the definition of variance,

$$\begin{aligned} \text{Var}(z) &= \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(z_i)^2 \end{aligned} \tag{20}$$

since $\bar{z} = 0$. Then, using the definition of z_i, we have

$$\begin{aligned} \text{Var}(z) &= \frac{1}{n}\sum_{i=1}^{n}(\frac{x_i - \bar{x}}{\text{SD}_x})^2 \\ &= \frac{1}{(\text{SD}_x)^2}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= \frac{(\text{SD}_x)^2}{(\text{SD}_x)^2} \\ &= 1. \end{aligned} \tag{21}$$

The penultimate step of (21) uses the fact that $\text{Var}_x = (\text{SD}_x)^2$, and the right-most term in the second line is simply $(\text{SD}_x)^2$.

Then, since $\sqrt{\text{Var}(z)} = \text{SD}_z$, we have also that $\text{SD}_z = 1$.

Thus, when a variable converted to standard units, it has variance and standard deviation equal to 1.

6. **Hooke's Law**

Solution See Freedman and Lane (1981) for a sketch of the solutions.