Simple Linear Regression Beta Coefficient Variance Derivations

BIOS 6611

CU Anschutz

Week 8

Formula Refreshers

Formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$

Our formulas for our beta coefficients are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{S_{XY}}{S_{XX}}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Our formulas for the standard errors of our beta coefficients are

$$SE(\hat{\beta}_0) = \sqrt{\sigma^2_{Y|X}\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)}$$

$$SE(\hat{\beta}_1) = \sqrt{\frac{\sigma^2_{Y|X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

Helpful Properties and Formula Refresher

$$Var(Y) = E(Y^2) - [E(Y)]^2 \implies E(Y^2) = Var(Y) + [E(Y)]^2$$

$$Var(cY) = c^2 \times Var(Y)$$

$$SE(cY) = |c| \times SE(Y)$$

$$Cov(aX + b, cY + d) = ac \times Cov(X, Y)$$

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$$

$$Var(Y) = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n - 1}$$

$$Cov(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

$$r_{x,y} = \frac{Cov\,(X, Y)}{SD\,(X)\,SD\,(Y)} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}$$

Calculating the Variances

Deriving $Var(\hat{\beta}_1)$

Before deriving the variance of $\hat{\beta}_1$, let's do a little rearranging:

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} \\
&= \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right) Y_i - \bar{Y} \sum_{i=1}^{n} (X_i - \bar{X})}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} \\
&= \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right) Y_i}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} \\
&= \frac{\left(X_1 - \bar{X}\right) Y_1 + \ldots + \left(X_n - \bar{X}\right) Y_n}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}
\end{aligned}
$$

where $\bar{Y} \sum_{i=1}^{n} (X_i - \bar{X}) = 0$ because

$$
\sum_{i=1}^{n} (X_i - \bar{X}) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X} = \frac{n}{n} \times \sum_{i=1}^{n} X_i - n\bar{X} = n\bar{X} - n\bar{X} = 0
$$

Deriving $Var(\hat{\beta}_1)$

Now we are ready to derive our variance for the slope:

$$Var(\hat{\beta}_1) = Var\left[\frac{(X_1 - \bar{X})Y_1 + ... + (X_n - \bar{X})Y_n}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

Let's cut to our "whiteboard" to work through the math...

Deriving $Var(\hat{\beta}_0)$

Now we are ready to derive our variance for the intercept:

$$Var(\hat{\beta}_0) = Var(\bar{Y} - \hat{\beta}_1 \bar{X})$$

Let's cut to our "whiteboard" to work through the math...

Showing $Cov(\bar{Y}, \hat{\beta}_1 \bar{X}) = 0$

First we will note that $E(Y_i Y_j) = E(Y_i) E(Y_j)\ \forall i \neq j$ by independence and that $\sum_{i=1}^n \left(X_i - \bar{X}\right) = 0$.

Then we can note another representation of $\hat{\beta}_1$:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
&= \frac{\sum_{i=1}^n [(X_i - \bar{X}) Y_i] - \bar{Y} \sum_{i=1}^n (X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
&= \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2}
\end{aligned}$$

Okay, we are ready to show (on the whiteboard):

$$Cov(\bar{Y}, \hat{\beta}_1 \bar{X}) = Cov\left(\frac{\sum_{i=1}^n Y_i}{n},\ \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right) Y_i}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2} \times \bar{X} \right) = 0$$

The Standard Error of the Mean

Deriving $SE(\hat{\mu}_{Y|X_0})$

We can manipulate our predicted regression equation for this calculation:

$$\hat{\mu}_{Y|X_0} = \hat{\beta}_0 + \hat{\beta}_1 X_0 = \left(\bar{Y} - \hat{\beta}_1 \bar{X}\right) + \hat{\beta}_1 X_0 = \bar{Y} + \hat{\beta}_1 (X_0 - \bar{X})$$

Then we can calculate the variance as

$$\begin{aligned}
Var\left(\hat{\mu}_{Y|X_0}\right) &= Var\left(\bar{Y} + \hat{\beta}_1\left(X_0 - \bar{X}\right)\right) \\
&= Var\left(\bar{Y}\right) + \left(X_0 - \bar{X}\right)^2 Var\left(\hat{\beta}_1\right) + 2(X_0 - \bar{X})Cov(\bar{Y}, \hat{\beta}_1) \\
&= Var\left(\bar{Y}\right) + \left(X_0 - \bar{X}\right)^2 \frac{\hat{\sigma}^2_{Y|X}}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \\
&= \frac{\hat{\sigma}^2_{Y|X}}{n} + \frac{\hat{\sigma}^2_{Y|X}}{n-1}\left(\frac{\left(X_0 - \bar{X}\right)^2}{\hat{\sigma}^2_X}\right)
\end{aligned}$$

because $\hat{\sigma}^2_X = \sum_{i=1}^{n} \frac{\left(X_i - \bar{X}\right)^2}{n-1}$ and $Var\left(\hat{\beta}_1\right) = \frac{\hat{\sigma}^2_{Y|X}}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$.