Properties of OLS Estimators, Part I

Q: Suppose we are going to (1) gather data \( \{(x_i, y_i), i = 1, \ldots, n\} \)

(2) Compute \( (\hat{\beta}_0, \hat{\beta}_1) \), our estimators

How close will \( \{(\hat{\beta}_0, \hat{\beta}_1, x)\} \) be to \( E[y|x=x] \)?

Two Scenarios:

Linear model correct - 
\[
E[y|x=x] = \beta_0 + \beta_1 x
\]
for some unknown \( (\beta_0, \beta_1) \)

Linear Model Incorrect - 
\[
E[y|x=x] \neq \text{linear in } x
\]

We first consider the case where the model is correct.

---

Truth

\[
\begin{align*}
Y & \quad \downarrow \\
\beta_0 & \quad \beta_1 \\
\text{X} & \quad \text{X}
\end{align*}
\]

\[
\text{Dataset 1} \quad \text{OLS Line 1}
\]

\[
\text{Dataset 2} \quad \text{OLS Line 2}
\]

The values of \( (\hat{\beta}_0, \hat{\beta}_1) \) depend on the outcome of your experiment or study.

If your experiment has random error or noise, or your study is a random sample, then \( (\hat{\beta}_0, \hat{\beta}_1) \) are random too.

Etc
The variability of \((\hat{\beta}_0, \hat{\beta}_1)\) across possible experimental outcomes is the sampling variability of \((\hat{\beta}_0, \hat{\beta}_1)\).

The probability distribution of \((\hat{\beta}_0, \hat{\beta}_1)\) across possible outcomes is the sampling distribution of \((\hat{\beta}_0, \hat{\beta}_1)\).

What do the sampling distributions look like?

\[
\begin{align*}
\hat{\beta}_0 & \sim p(\hat{\beta}_0) \\
\hat{\beta}_1 & \sim p(\hat{\beta}_1)
\end{align*}
\]

Result #1: If \(E[Y \mid X = x] = \beta_0 + \beta_1 x\), then

\[
\begin{align*}
E[\hat{\beta}_0] &= \beta_0 \\
E[\hat{\beta}_1] &= \beta_1,
\end{align*}
\]

i.e., \((\hat{\beta}_0, \hat{\beta}_1)\) are unbiased estimators of \((\beta_0, \beta_1)\).

Proof: Recall \(\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}\)

\[
E[\hat{\beta}_1 \mid x] = \frac{\Sigma (x_i - \bar{x}) E[(y_i - \bar{y}) \mid x]}{\Sigma (x_i - \bar{x})^2}
\]
\[ E[\gamma_i | x_i] = \frac{E[\gamma_i | x_i]}{\hat{\beta}_0 + \hat{\beta}_1 x_i} \]

\[ (\hat{\beta}_0 + \hat{\beta}_1 x_i) - E[\gamma_i | x_i] \]

\[ \hat{\beta}_0 + \hat{\beta}_1 x_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \]

\[ \hat{\beta}_0 = \hat{\gamma} - \hat{\beta}_1 \bar{x} \]

\[ E[\hat{\beta}_0 | x_i] = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} - E[\hat{\beta}_1 | x_i] \]

\[ \beta_0 + \beta_1 \bar{x} = \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = \beta_0 \]

So: If \( E[\gamma_i | x_i] = \beta_0 + \beta_1 x_i \) for some \( \beta_0, \beta_1 \),

then \( E[\hat{\beta}_0, \hat{\beta}_1] = (\beta_0, \beta_1) \).

This says the sampling distribution of \((\hat{\beta}_0, \hat{\beta}_1)\) are centered around the correct values but it doesn't say how \underline{concentrated} these distributions are.

\[ \text{Concentration of } (\hat{\beta}_0, \hat{\beta}_1) \text{ around } (\beta_0, \beta_1) \text{ can be measured with variance and covariance.} \]
Result #2

1. \( E[y_i | x_i] = \beta_0 + \beta_1 x_i \) (linear model)
2. \( \text{Var}(y_i | x_i) = \sigma^2 \) (constant variance)
3. \( \text{Cov}(y_i, y_j | x_i, x_j) = 0 \) (observations are uncorrelated)

then \( E[(\hat{\beta}_0, \hat{\beta}_1) | x] = (\beta_0, \beta_1) \)

and

\[
\begin{align*}
\text{Var}(\hat{\beta}_1 | x) &= \sigma^2 \left( \frac{1}{s_{xx}} \right) \\
\text{Var}(\hat{\beta}_0 | x) &= \sigma^2 \left( \frac{1}{s_{xx}} - \frac{x^2}{s_{xxx}} \right) \\
\text{Cov}(\hat{\beta}_0, \hat{\beta}_1 | x) &= -\sigma^2 \left( \frac{x}{s_{xxx}} \right)
\end{align*}
\]

will prove when we get to multiple linear regression.

Why is this negative? What does it mean?

Interpretation: If \( \hat{\beta}_1 \) is higher than \( \beta_1 \), we expect \( \hat{\beta}_0 \) to be lower than \( \beta_0 \).

Picture:

\[
\begin{align*}
\text{MATH:} \quad & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1 | x) = \text{Cov}(\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1 | x) \\
& = \text{Cov}(\bar{y}, \hat{\beta}_1 | x) - \text{Cov}(\hat{\beta}_1, \bar{x}, \hat{\beta}_1 | x) \\
& = 0 - \bar{x} \text{ Cov}(\hat{\beta}_1, \hat{\beta}_1 | x) \\
& = -\bar{x} \text{ Var}(\hat{\beta}_1 | x) \\
& = -\bar{x} \frac{\sigma^2}{s_{xxx}}
\end{align*}
\]