Properties of OLS Estimators, Part II

Objective: Infer properties of a process from data from an experiment (population)

Post-experimental: Data are fixed, known
\( \hat{\beta}_0, \hat{\beta}_1 \) are fixed, known

Pre-experimental: Data are unknown, random
\( \beta_0, \beta_1 \) are unknown, random

Pre-experimental question: Is the probability dist. of \((\hat{\beta}_0, \hat{\beta}_1)\) centered around a meaningful representation of the process or population?

Probability Model of an experiment

Experiment: An experimental outcome \( y_i \) is generated for each experimental condition \( x_i \), \( i = 1, \ldots, n \).

Linear Model:
\[
E[y_i | x_i] = \beta_0 + \beta_1 x_i \quad \text{for some } \beta_0, \beta_1,
\]
equivalently,
\[
\{ \ y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \text{if } E[\epsilon_i] = 0 \} \quad (A1)
\]

Return to the question: Are the values of \((\hat{\beta}_0, \hat{\beta}_1)\) we will get from the data likely to be representative of the population or process?

Result 1: If \((A1)\) holds, then
\[
E[\hat{\beta}_0] = \beta_0 \quad \text{(\(\hat{\beta}_0, \hat{\beta}_1\) are unbiased for \((\beta_0, \beta_1)\)).}
\]

Proof:
\[
E[\hat{\beta}_1 | x_i] = E[\frac{S_{xy}}{S_{xx}} | x_i] = \frac{1}{S_{xx}} E[\sum (x_i - \bar{x})(y_i - \bar{y}) | x_i]
\]
\[
= \frac{1}{S_{xx}} \sum (x_i - \bar{x}) E(y_i - \bar{y} | x_i) \quad (x_i)
\]
\[
= \frac{1}{S_{xx}} \sum (x_i - \bar{x}) \beta_1 (x_i - \bar{x}) = \beta_1 \frac{S_{xx}}{S_{xx}} = S_{xx}.
\]
\[
\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}
\]

\[
E[\hat{\beta}_0 | x] = E[\bar{Y} | x] - E[\hat{\beta}_1 | x] \bar{X}
\]

\[
= \beta_0 + \beta_1 \bar{X} - \beta_1 \bar{X} = \beta_0 \quad (\text{for all } x),
\]

\[\text{Result 1: does not require independence, normality, constant variance.}\]

\[\text{but it does not tell us much.}\]

\[\text{Result 1 does not say how concentrated the dist. of } \hat{\beta}_1 \text{ is around } \beta_1.\]

\[\text{Concentration of } (\hat{\beta}_0, \hat{\beta}_1) \text{ around } (\beta_0, \beta_1) \text{ can be measured with variance and covariance. But to say something about variance, we need to make additional assumptions:}\]

\[(A1) \quad Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \quad \quad E[\varepsilon_i | x] = 0
\]

\[(A2) \quad \text{Var}(\varepsilon_i | x) = \sigma^2 \quad \text{for all } i \quad \text{(constant variance)}
\]

\[\text{Cov}(\varepsilon_i, \varepsilon_j | x) = 0 \quad \text{for all } i, j \quad \text{(uncorrelated errors)}\]

\[\text{Result 2: If } (A1) + (A2) \text{ hold, then}\]

\[E[\hat{\beta}_0, \hat{\beta}_1] = (\beta_0, \beta_1) \quad \text{and} \quad \text{Var}(\hat{\beta}_1 | x) = \frac{\sigma^2}{\sum x^2}
\]

\[\text{Var}(\hat{\beta}_0 | x) = \sigma^2 \left( \frac{1}{n} + \frac{1}{\sum x^2} \right)
\]

\[\text{Cov}(\hat{\beta}_0, \hat{\beta}_1 | x) = -\sigma^2 \left( \frac{\bar{x}}{\sum x^2} \right)
\]
Interpretation

\[
\text{Var} (\hat{\beta}_1 | x) = \frac{\sigma^2}{s_{xx}} = \frac{\sigma^2}{n \cdot s_x^2} = \frac{\sigma^2}{n \cdot S_x^2}
\]

- Depends on \( x \)
- Small if \( n \) is big
  \( \sigma^2 \) is small
  \( S_x^2 \) is big

Picture:

\[
\text{Var} (\hat{\beta}_0 | x) = \frac{\sigma^2}{n} \left( 1 + \frac{x^2}{s_{xx}} \right) = \frac{\sigma^2}{n} \left( 1 + \frac{\overline{x}^2}{s_x^2} \right) = \frac{\sigma^2}{n} \frac{1}{s_x^2} \left( \sum x_i^2 \right)
\]

Same interpretation as for \( \hat{\beta}_1 \), except \( \text{Var} (\hat{\beta}_0) \) is larger

\( x_i \)'s are not centered around \( \overline{xx} \)
\[ \text{Cov} \left( \hat{\beta}_0, \hat{\beta}_1, x \right) = -\sigma^2 \left( \frac{x}{Sxx} \right) \]  
Why negative?

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

**Picture:**

\[ \text{ols}: \hat{\beta}_1 \text{ too high} \]

\[ \text{truth} \]

\[ \hat{\beta}_0 \text{ to Cov.} \]

**Math:**

\[ \text{Cov} \left( \hat{\beta}_0, \hat{\beta}_1, x \right) = \text{Cov} \left( \hat{\gamma} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1, x \right) \]

\[ = \text{Cov} \left( \hat{\gamma}, \hat{\beta}_1, x \right) - \text{Cov} \left( \hat{\beta}_1, \bar{x}, \hat{\beta}_1, x \right) \]

\[ = 0 - \bar{x} \text{ Cov} \left( \hat{\beta}_1, \hat{\beta}_1, x \right) \]

\[ = -\bar{x} \text{ Var} \left( \hat{\beta}_1, x \right) \]

\[ = -\bar{x} \frac{\sigma^2}{Sxx} \]

**Exercise:** Suppose you want to design an experiment to estimate \( \beta_0, \beta_1 \). How should you choose your \( x_i \)’s?

**Estimating the variance:**

**Q:** Is \( \hat{\beta}_1 \) a precise estimate of \( \beta_1 \)?

**A:** \[ \text{V} \left( \hat{\beta}_1, x \right) = \frac{\sigma^2}{\bar{x}} \]

**Q:** \( S_x^2 \) is known, \( \bar{x} \) is known, but \( \sigma^2 \) isn’t.

\[ S_x^2 \text{ is known, } \bar{x} \text{ is known, but } \sigma^2 \text{ isn't.} \]

\[ S_x^2 \text{ is known, } \bar{x} \text{ is known, but } \sigma^2 \text{ isn't.} \]

**A:** Need to know \( \sigma^2 \)?
Under (A1) \& (A2),

\[ \gamma_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \text{Var}(\epsilon_i | x_i) = \sigma^2, \quad \text{Cov}(\epsilon_i, \epsilon_j | x_i) = 0 \]

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

Sample var \( \epsilon_i \) \( \approx \sigma^2 \)

But don't have \( \epsilon_i \), as we don't know \( \beta_0, \beta_1 \).

Idea: \( \hat{\beta}_0 = \beta_0, \hat{\beta}_1 = \beta_1 \), so \( \epsilon_i \approx \gamma_i - [\hat{\beta}_0 + \hat{\beta}_1 x_i] = \hat{\epsilon}_i = \text{residual} \)

\[ \text{Var}(\epsilon_i, \ldots, \epsilon_n) \approx \text{Var}(\epsilon_1, \ldots, \epsilon_n) = \sigma^2 \]

More precisely:

Let \( \text{RSS} = \sum (\gamma_i - [\hat{\beta}_0 + \hat{\beta}_1 x_i])^2 = \sum \hat{\epsilon}_i^2 \)

then \( \mathbb{E}[\text{RSS}'] = \sigma^2 \cdot (n-2) \)

\( \mathbb{E}[\text{RSS}/(n-2)] = \sigma^2 \)

so \( \hat{\sigma}^2 = \frac{\text{RSS}}{n-2} \) is an unbiased estimate of \( \sigma^2 \).

Q: Is \( \hat{\beta}_1 \) a precise \( \text{est.} \) of \( \beta_1 \)?

A: \( \text{V}(\hat{\beta}_1 | x) = \frac{\sigma^2}{S_x^2} \quad \text{SD}(\hat{\beta}_1 | x) = \frac{\sigma}{\sqrt{S_x}} \frac{1}{S_x} \)

\[ \approx \frac{\sigma^2}{S_x^2} \quad \approx \frac{\sigma}{\sqrt{S_x}} \frac{1}{S_x} = \text{SE}(\hat{\beta}_1 | x) \]

The estimated standard deviation is the standard error.

Similarly, \( \text{SE}(\hat{\beta}_0 | x) = \sqrt{\hat{\sigma}^2 \left(1 + \frac{x^2}{S_x} \right)} = \hat{\sigma} \sqrt{1 + \frac{x^2}{S_x}} \)

\( \hat{\sigma}^2, \text{SE}(\hat{\beta}_0 | x), \text{SE}(\hat{\beta}_1 | x) \) will help us with \{ confidence intervals, hypothesis tests, prediction intervals \}. 
Normal Linear Model

Confidence intervals require we know more about the distribution of $(\hat{\beta}_0, \hat{\beta}_1)$ beyond expectations, variances.

Result 3: If $\varepsilon_1, \ldots, \varepsilon_n$ are normally distributed, then

\[
\hat{\beta}_0, \hat{\beta}_1 \text{ also.}
\]

Result 3b: Even if $\varepsilon_1, \ldots, \varepsilon_n$ are not normally distributed, $\hat{\beta}_0, \hat{\beta}_1$ are approximately normally distributed for large $n$.

More specifically:

\[
\begin{align*}
(A1) & \quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad E[\varepsilon_i] = 0 \\
(A2) & \quad \text{Var}(\varepsilon_i | x_i) = \sigma^2, \quad \text{Cov}(\varepsilon_i, \varepsilon_j | x_i) = 0 \quad \forall i, j \\
(A3) & \quad \varepsilon_i \sim \text{norm} \quad \forall i \\
(A23) & \quad \varepsilon_1, \ldots, \varepsilon_n \sim \text{ind} \quad N(0, \sigma^2)
\end{align*}
\]

Result 3: If $A1 + A2 + A3$ hold, then given in Result 2,

\[
(\hat{\beta}_0, \hat{\beta}_1) \sim N \left( \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \right)
\]

This says:

\[
\begin{align*}
E[\hat{\beta}_0 | x] &= \beta_0 \\
\text{Cov}(\hat{\beta}_0, \hat{\beta}_1 | x) &= \text{Cov} \quad \checkmark \text{ already showed this}
\end{align*}
\]

\[
\begin{align*}
(\hat{\beta}_0, \hat{\beta}_1) \text{ are jointly normally distributed.} & \checkmark \text{ need to show this}
\end{align*}
\]
Recall: (1) \( z \sim N(0,1) \rightarrow a \cdot z \sim N(a, b^2) \)

(2) \( w_1, \ldots, w_n \) indep. normal, \( w_i \sim N(\mu_i, \sigma_i^2) \)

then \( a \cdot w : E_i \cdot w_i \sim N(E_i \cdot \mu_i, E_i \cdot \sigma_i^2) \)

\[
\hat{e}_i = \frac{s_{xy}}{s_{xx}} = \frac{1}{s_{xx}} (x_i - \bar{x}) (y_i - \bar{y})
\]

\[
(x_i - \bar{x}) = (I - 11^T/8) x = Cx
\]

\( (x_i - \bar{x}) (y_i - \bar{y}) = (x_i - \bar{x})^T C x \)

\( \Rightarrow \hat{e}_i = \frac{1}{s_{xx}} (x_i - \bar{x})^T C x \)

Now by (1), \( y_i = (\beta_0 + \beta_1 x_i) + e_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \)

(2) \( y_1, \ldots, y_n \) indep. normal, so \( a \cdot x \) is normal

(already computed the mean and variance)