Simple Linear Regression

Kosuke Imai
Harvard University
Spring 2021
Linear regression is commonly used in applied research

We will explore how to use linear regression for causal effect estimation

To build intuition, we focus on the application of simple linear regression
Regression \( \sim \) conditional expectation function of \( Y \) given \( X \)

\[
\mathbb{E}(Y \mid X) = f(X) = \beta^\top X
\]

When can we interpret coefficients as causal effects?

Causal model as structural equation model

\[
Y_i(t) = \alpha + \beta t + \epsilon_i \quad \text{for } t = 0, 1, \text{ where } \mathbb{E}(\epsilon_i) = 0
\]

1. No interference between units
2. \( \mathbb{E}(Y_i(0)) = \alpha \)
3. \( Y_i(1) - Y_i(0) = \beta \) for all \( i \) \( \iff \) constant additive unit causal effect

Heterogenous treatment effect model:

\[
Y_i(t) = \alpha + \beta_i t + \epsilon_i = \alpha + \beta t + (\beta_i - \beta)t + \epsilon_i
\]

\( = \epsilon_i(t) \)

where \( \mathbb{E}(\epsilon_i) = 0 \) and \( \beta = \mathbb{E}(\beta_i) = \mathbb{E}(Y_i(1) - Y_i(0)) \)

- \( \mathbb{E}(\epsilon_i(t)) = 0 \) for \( t = 0, 1 \)
- \( \alpha = \mathbb{E}(Y_i(0)) \)
Identification Assumption

- **Strict exogeneity** assumption:

  \[ \mathbb{E}(\epsilon_i \mid T) = \mathbb{E}(\epsilon_i) = 0 \quad \text{where} \quad T = (T_1, T_2, \ldots, T_n)^\top \]

- Under this assumption, least squares estimate \( \hat{\beta} \) is unbiased for \( \beta \)

- Randomization of treatment: \( \{ Y_i(1), Y_i(0) \}_{i=1}^n \perp \perp T \)

  \[ \mathbb{E}(Y_i(t)) = \mathbb{E}(Y_i(t) \mid T_i = t) = \mathbb{E}(Y_i \mid T_i = t) = \alpha + \beta t \]

  Recall that \( \epsilon_i(t) = Y_i(t) - \alpha - \beta t \). Then, \( \mathbb{E}(\epsilon_i(t) \mid T) = \mathbb{E}(\epsilon_i(t)) = 0 \)

- Random sampling of units: \( \{ Y_i(1), Y_i(0) \} \sim_{\text{i.i.d.}} P \)

  \[ \{\epsilon_i(1), \epsilon_i(0)\} \sim_{\text{i.i.d.}} P^* \]

  independence across observations
Unbiasedness of Least Squares Estimator

- Consider a RCT with a binary treatment

- Least squares estimators:
  \[ \hat{\alpha} = \frac{1}{n_0} \sum_{i=1}^{n} (1 - T_i) Y_i, \]
  \[ \hat{\beta} = \frac{1}{n_1} \sum_{i=1}^{n} T_i Y_i - \frac{1}{n_0} \sum_{i=1}^{n} (1 - T_i) Y_i. \]

- Thus, \( \hat{\alpha} \) and \( \hat{\beta} \) are unbiased for \( \mathbb{E}(Y(0)) \) and \( \mathbb{E}(Y_i(1) - Y_i(0)) \)

- A similar conclusion holds if \( T \) is a categorical treatment
- When \( T \) is continuous, the model assumes \( Y_i(t) \) is linear in \( t \)

- What about standard errors?
Homoskedasticity

- Homoskedastic error:

\[ \mathbb{V}(\epsilon \mid T) = \sigma^2 I_n \]

1. Equal variance of potential outcomes

\[ \mathbb{V}(\epsilon_i(0)) = \mathbb{V}(\epsilon_i(1)) \]

2. Random sampling of units implies \( \epsilon_i \perp \perp \epsilon_j \) and \( \mathbb{V}(\epsilon_i) = \mathbb{V}(\epsilon_j) \)

- Under the homoskedasticity assumption,
  - model-based variance is,

\[ \mathbb{V}(\hat{\beta} \mid T) = \frac{\sigma^2}{\sum_{i=1}^{n} (T_i - \bar{T})^2} \]

- standard model-based variance estimator is,

\[ \mathbb{V}(\hat{\beta} \mid T) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} (T_i - \bar{T})^2} \quad \text{where} \quad \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{\epsilon}_i^2 \]
Violation of the Homoskedasticity Assumption

- Homoskedasticity assumption may be unrealistic: $\sigma_1^2 \neq \sigma_0^2$
- Constant additive treatment effect $\Rightarrow$ homoskedasticity
- Bias in the estimated variance when the assumption does not hold

$$
\text{Bias} = \mathbb{E} \left( \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(T_i - \bar{T})^2} \right) - \left( \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0} \right)
$$

under complete randomization

true variance

$$
= \frac{(n_1 - n_0)(n - 1)}{n_1 n_0 (n - 2)} \left( \sigma_1^2 - \sigma_0^2 \right)
$$

- Bias is zero when
  1. homoskedasticity assumption holds: $\sigma_1^2 = \sigma_0^2$
  2. design is balanced: $n_1 = n_0$
- Bias can be negative or positive
- Bias is typically small but does not go away asymptotically
Heteroskedasticity-Robust Variance Estimator

- Eicker-Huber-White (EHW) robust variance estimator:

$$
\hat{V}_{EHW}(\hat{\alpha}, \hat{\beta} | X) = \left( X^\top X \right)^{-1} \left( X^\top \text{diag}(\hat{\epsilon}_i^2) X \right) \left( X^\top X \right)^{-1}
$$

where the **sandwich** estimator is defined as:

$$
\text{bread} \left( n \sum_{i=1}^{n} x_ix_i^\top \right)^{-1} \text{meat} \left( n \sum_{i=1}^{n} \hat{\epsilon}_i^2 x_ix_i^\top \right) \text{bread} \left( n \sum_{i=1}^{n} x_ix_i^\top \right)^{-1}
$$

- The EHW robust variance estimator is asymptotically unbiased:

$$
\hat{V}_{EHW}(\hat{\beta} | T) = \bar{\sigma}_1^2 + \bar{\sigma}_0^2 \quad \text{where} \quad \bar{\sigma}_t^2 = \frac{1}{n_t} \sum_{i=1}^{n} 1\{T_i = t\}(Y_i - \bar{Y}_t)^2
$$

$$
\text{Bias} = \mathbb{E}(\hat{V}_{EHW}(\hat{\beta} | T)) - \left( \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0} \right) = -\left( \frac{\sigma_1^2}{n_1^2} + \frac{\sigma_0^2}{n_0^2} \right)
$$
Improved Robust Variance Estimators

- **HC2 heteroskedasticity-robust variance estimator:**

\[
\hat{\text{V}}_{\text{HC2}}((\hat{\alpha}, \hat{\beta}) \mid X) = (X^\top X)^{-1} \left\{ X^\top \text{diag} \left( \frac{\hat{\epsilon}_i^2}{1 - p_{ii}} \right) X \right\} (X^\top X)^{-1}
\]

where \( p_{ii} \) is the leverage

\[
p_{ii} = X_i^\top (X^\top X)^{-1} X_i = \|P_X v(i)\|^2 = \begin{cases} \frac{1}{n_1} & \text{if } T_i = 1 \\ \frac{1}{n_0} & \text{if } T_i = 0 \end{cases}
\]

- \( P_X = X(X^\top X)^{-1} X^\top \): projection matrix
- \( v(i) \): a vector whose \( i \)th element is 1 and other elements are zero

- **HC2 variance estimator is identical to the Neyman’s variance estimator and is unbiased** (Samii and Aronow. 2012. *Stat Probab Lett*)

\[
\hat{\text{V}}_{\text{HC2}}(\hat{\beta} \mid T) = \frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0}
\]
Summary

- Simple linear regression = Difference-in-means estimator

- Homoskedasticity implies the equal variance of potential outcomes

- Heteroskedasticity-robust variance estimator relaxes this assumption