# OLS Regression Uncertainty

## UC Berkeley, MIDS w203

Statistics for Data Science

February 11, 2022

# Large-Sample Regression

We assume that the best linear predictor of Y given $\mathbf{X}$ is $\mathbf{X}\boldsymbol{\beta}$.

$$\mathrm{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon.$$

We know that

$$\mathbb{E}\left[\varepsilon\right] = 0, \text{ and } \mathbb{E}\left[\mathbf{X}^{\mathrm{T}}\varepsilon\right] = \mathbf{0}.$$

# OLS Coefficients

We also assume that the dataset $\{(\mathrm{Y_i}, \mathbf{X}_\mathrm{i})\}$ is taken i.i.d. from the joint distribution of $(\mathrm{Y}, \mathbf{X})$. For each i, we can write

$$\mathrm{Y_i} = \mathbf{X_i}\boldsymbol{\beta} + \varepsilon_\mathrm{i}.$$

In matrix notation, we can write

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

And, the solution (when exists) can be written as

$$\widehat{\boldsymbol{\beta}} = \left[\mathbb{X}^\mathrm{T}\mathbb{X}\right]^{-1}\mathbb{X}^\mathrm{T}\mathbf{Y}.$$

## An Alternative Form of Coefficients

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \left[\mathbb{X}^{\mathrm{T}}\mathbb{X}\right]^{-1}\mathbb{X}^{\mathrm{T}}\mathbf{Y}\\ &= \left[\mathbb{X}^{\mathrm{T}}\mathbb{X}\right]^{-1}\mathbb{X}^{\mathrm{T}}(\mathbb{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})\\ &= \left[\mathbb{X}^{\mathrm{T}}\mathbb{X}\right]^{-1}(\mathbb{X}^{\mathrm{T}}\mathbb{X})\boldsymbol{\beta}+\left[\mathbb{X}^{\mathrm{T}}\mathbb{X}\right]^{-1}\mathbb{X}^{\mathrm{T}}\boldsymbol{\varepsilon}\\ &= \boldsymbol{\beta}+\left[\mathbb{X}^{\mathrm{T}}\mathbb{X}\right]^{-1}\mathbb{X}^{\mathrm{T}}\boldsymbol{\varepsilon}\end{aligned}$$

So,

$$\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta} = \left[\mathbb{X}^{\mathrm{T}}\mathbb{X}\right]^{-1}\mathbb{X}^{\mathrm{T}}\boldsymbol{\varepsilon} \tag{1}$$

# Asymptotic Variance

We multiply Eq. (1) by $\sqrt{n}$ to get

$$\begin{aligned}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) &= \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{X}_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{X}_i^T\varepsilon_i\right)\\ &= \widehat{\mathbb{Q}}_{\mathbf{XX}}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{X}_i^T\varepsilon_i\right)\end{aligned}$$

From the consistency of OLS estimators, we already have

$$\widehat{\mathbb{Q}}_{\mathbf{XX}} \xrightarrow[p]{} \mathbb{Q}_{\mathbf{XX}}$$

# Asymptotic Variance

$\{\mathbf{X}_i\varepsilon_i\}_i$ are i.i.d. from $\mathbf{X}\varepsilon$.

$$\mathbb{E}\left[\mathbf{X}_i^T\varepsilon_i\right] = \mathbb{E}\left[\mathbf{X}^T\varepsilon\right] = \mathbf{0}.$$

Now,

$$\mathbb{V}[\mathbf{X}_i^T\varepsilon_i] = \mathbb{E}\left[\mathbf{X}_i^T\varepsilon_i\left(\mathbf{X}_i^T\varepsilon_i\right)^T\right] = \mathbb{E}\left[\mathbf{X}^T\mathbf{X}\varepsilon^2\right] \stackrel{\text{def}}{=} \mathbb{A}.$$

By the (multivariate) Central Limit Theorem, as $n \to \infty$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{X}_i^T\varepsilon_i \xrightarrow[d]{} \mathcal{N}(\mathbf{0}, \mathbb{A}).$$

There is a small technicality here, we must have $\mathbb{A} < \infty$. This can be imposed by a stronger regularity condition on the moments, e.g., $\mathbb{E}\left[Y^4\right], \mathbb{E}\left[||\mathbf{X}||^4\right] < \infty$.

# The Final Form

Putting everything together, we conclude

$$\begin{aligned}
\sqrt{\mathrm{n}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &\xrightarrow[\mathrm{d}]{} \mathbb{Q}_{\mathbf{XX}}^{-1}\mathcal{N}(\mathbf{0}, \mathbb{A}) \\
&= \mathcal{N}\left(\mathbf{0}, \left[\mathbb{Q}_{\mathbf{XX}}^{-1}\right]^{\mathrm{T}} \mathbb{A}\mathbb{Q}_{\mathbf{XX}}^{-1}\right) \\
&= \mathcal{N}\left(\mathbf{0}, \mathbb{Q}_{\mathbf{XX}}^{-1}\mathbb{A}\mathbb{Q}_{\mathbf{XX}}^{-1}\right)
\end{aligned}$$

## Asymptotic Variance Theorem

1. The observations $\{(\mathrm{Y}_i, \mathbf{X}_i)\}_{i=1}^{n}$ are i.i.d from the joint distribution of $(\mathrm{Y}, \mathbf{X})$
2. $\mathbb{E}\left[\mathrm{Y}^4\right] < \infty$
3. $\mathbb{E}\left[||\mathbf{X}||^4\right] < \infty$
4. $\mathbb{Q}_{\mathbf{XX}} = \mathbb{E}\left[\mathbf{XX}'\right]$ is positive-definite.

Then, as $n \to \infty$

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow[d]{} \mathcal{N}\left(\mathbf{0}, \mathbb{V}_{\boldsymbol{\beta}}\right),$$

where

$$\mathbb{V}_{\boldsymbol{\beta}} \stackrel{\text{def}}{=} \mathbb{Q}_{\mathbf{XX}}^{-1} \mathbb{A} \mathbb{Q}_{\mathbf{XX}}^{-1}$$

and $\mathbb{Q}_{\mathbf{XX}} = \mathbb{E}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]$, $\mathbb{A} = \mathbb{E}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\varepsilon^2\right]$.

# Estimating the Asymptotic Variance Matrix

The asymptotic variance of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is

$$\mathbb{V}_{\boldsymbol{\beta}} = \mathbb{Q}_{\mathbf{XX}}^{-1} \mathbb{A} \mathbb{Q}_{\mathbf{XX}}^{-1}.$$

- $\widehat{\mathbb{Q}}_{\mathbf{XX}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{X}_i$ is a natural estimator for $\mathbb{Q}_{\mathbf{XX}}$.
- For $\mathbb{A}$, we use the moment estimator

$$\widehat{\mathbb{A}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{X}_i e_i^2,$$

where $e_i = (Y_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}})$ is the i-th residual. As it turns out, $\widehat{\mathbb{A}}$ is a consistent estimator for $\mathbb{A}$.

# Robust Error Variance Estimators

As a result, we get the following plug-in estimator for $\mathbb{V}_{\boldsymbol{\beta}}$:

$$\widehat{\mathbb{V}}_{\boldsymbol{\beta}} = \widehat{\mathbb{Q}}_{\mathbf{XX}}^{-1} \widehat{\mathbb{A}} \widehat{\mathbb{Q}}_{\mathbf{XX}}^{-1}$$

$$\begin{aligned}
\widehat{\mathbb{V}}\left[\widehat{\boldsymbol{\beta}}\right] &= \frac{1}{n} \widehat{\mathbb{Q}}_{\mathbf{XX}}^{-1} \widehat{\mathbb{A}} \widehat{\mathbb{Q}}_{\mathbf{XX}}^{-1} \\
&= \frac{1}{n} \left( \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{X}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} e_i^2 \mathbf{X}_i^T \mathbf{X}_i \right) \left( \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{X}_i \right)^{-1} \\
&= \left( \mathbb{X}^T \mathbb{X} \right)^{-1} \mathbb{X}^T \mathbb{D} \mathbb{X} \left( \mathbb{X}^T \mathbb{X} \right)^{-1}
\end{aligned}$$

where $\mathbb{D}$ is an $n \times n$ diagonal matrix with diagonal entries $e_1^2, e_2^2, \ldots, e_n^2$.