Lecture 7: Confidence intervals

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

April 5, 2026

Point vs interval estimators

- Recall our model:
  1. $Y_i = \beta_0 + \beta_1 X_i + U_i, \qquad i = 1, \dots, n.$
  2. $\mathrm{E}[U_i \mid \mathbf{X}] = 0$ for all i's.
  3. $\mathrm{E}[U_i^2 \mid \mathbf{X}] = \sigma^2$ for all i's.
  4. $\mathrm{E}[U_i U_j \mid \mathbf{X}] = 0$ for all $i \neq j$.
- So far we have established that conditionally on $\mathbf{X}$:
  - $\mathrm{E}\left[\hat{\beta}_1 \mid \mathbf{X}\right] = \beta_1$ (unbiasedness),
  - $\mathrm{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \dfrac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$.
- If the U_i's are continuously distributed, then with probability **one**, $\hat{\beta}_1 \neq \beta_1$:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (X_i - \bar{X}) U_i}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

- An **interval estimator** is a random interval $[LB, UB]$ that contains the true parameter value with a pre-specified probability.
- To construct an interval estimator for β_1, we need to know the **distribution** of $\hat{\beta}_1$.
- This requires an additional assumption about the distribution of U_i's. Let's first review the normal distribution.

Normal distribution

- A normal rv is a continuous rv that can take on any value. The PDF of a normal rv X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{ where}$$

$$\mu = \mathrm{E}[X] \text{ and } \sigma^2 = \mathrm{Var}(X).$$

  We usually write $X \sim N(\mu, \sigma^2)$.
- If $X \sim N(\mu, \sigma^2)$, then $a + bX \sim N(a + b\mu, b^2\sigma^2)$.

Standard normal distribution

- A standard normal rv has $\mu = 0$ and $\sigma^2 = 1$. Its PDF is $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$.
- Symmetric around zero (mean): if $Z \sim N(0,1)$,

$$P(Z > 0) = P(Z < 0) = 0.5,$$
$$P(Z > z) = P(Z < -z) \text{ for any } z.$$

- Thin tails: $P(-1.96 \leq Z \leq 1.96) = 0.95$.
- If $X \sim N(\mu, \sigma^2)$, then $(X - \mu)/\sigma \sim N(0, 1)$.

Bivariate normal distribution

- X and Y have a bivariate normal distribution if their joint PDF is given by:

$$f(x,y) = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_X^2\sigma_Y^2}} \exp\left[-\frac{Q}{2(1-\rho^2)}\right],$$

where

$$Q = \frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y},$$

$\mu_X = \mathrm{E}[X]$, $\mu_Y = \mathrm{E}[Y]$, $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$, and $\rho = \mathrm{Corr}(X, Y)$.

Properties of bivariate normal

- If X and Y have a bivariate normal distribution, then $a + bX + cY \sim N(\mu^*, (\sigma^*)^2)$, where

$$\mu^* = a + b\mu_X + c\mu_Y,$$
$$(\sigma^*)^2 = b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\rho\sigma_X\sigma_Y.$$

- $\mathrm{Cov}(X, Y) = 0 \Longrightarrow X$ and Y are independent.
- This can be generalized to more than 2 variables (multivariate normal).

Normality of the OLS estimator

- **Assumption 5:** U's are jointly normally distributed conditional on $\mathbf{X}$.
- Then $Y_i = \beta_0 + \beta_1 X_i + U_i$ are also jointly normally distributed conditional on $\mathbf{X}$.
- Since $\hat{\beta}_1 = \sum_{i=1}^n w_i Y_i$, where $w_i = \frac{X_i - \bar{X}}{\sum_{l=1}^n (X_l - \bar{X})^2}$ depend only on $\mathbf{X}$, $\hat{\beta}_1$ is also normally distributed conditional on $\mathbf{X}$.
- Conditionally on $\mathbf{X}$:

$$\hat{\beta}_1 \mid \mathbf{X} \sim N\left(\beta_1, \mathrm{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)\right),$$
$$\mathrm{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}.$$

Interval estimation problem

- We want to construct an **interval estimator** for β_1:
  - The interval estimator is called a **confidence interval** (CI).
  - A CI contains the **true** value β_1 **with some pre-specified probability** $1 - \alpha$, where α is a small probability of error.
  - For example, if $\alpha = 0.05$, then the random CI will contain β_1 with probability 0.95.
- $1 - \alpha$ is called the **coverage probability**.
- Confidence interval: $CI_{1-\alpha} = [LB_{1-\alpha}, UB_{1-\alpha}]$. The lower bound (LB) and upper bound (UB) should depend on the coverage probability $1 - \alpha$.

- The formal definition of CI: It is a **random interval** $CI_{1-\alpha}$ such that conditionally on $\mathbf{X}$,

$$P\left(\beta_1 \in CI_{1-\alpha} \mid \mathbf{X}\right) = 1 - \alpha.$$

Note that the random element is $CI_{1-\alpha}$.

- Sometimes, a CI is defined as $P\left(\beta_1 \in CI_{1-\alpha}\right) \geq 1 - \alpha$.

Symmetric CIs

- One approach to constructing CIs is to consider a **symmetric** interval around the estimator $\hat{\beta}_1$:

$$CI_{1-\alpha} = \left[\hat{\beta}_1 - c_{1-\alpha}, \hat{\beta}_1 + c_{1-\alpha}\right]$$

- The problem is choosing $c_{1-\alpha}$ such that $P\left(\beta_1 \in CI_{1-\alpha} \mid \mathbf{X}\right) = 1 - \alpha$.
- In choosing $c_{1-\alpha}$, we will be relying on the fact that, given our assumptions and conditionally on $\mathbf{X}$:

$$\hat{\beta}_1 \mid \mathbf{X} \sim N\left(\beta_1, \operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)\right),$$
$$\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}.$$

- Note that conditionally on $\mathbf{X}$:

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}} \sim N(0,1).$$

Standard normal quantiles

- Let $Z \sim N(0,1)$. The τ-th **quantile** (percentile) of the standard normal distribution is z_τ such that

$$P\left(Z \leq z_\tau\right) = \tau.$$

- **Median**: $\tau = 0.5$ and $z_{0.5} = 0$. ($P(Z \leq 0) = 0.5$).
- If $\tau = 0.975$ then $z_{0.975} = 1.96$. Due to symmetry, if $\tau = 0.025$ then $z_{0.025} = -1.96$.

σ^2 is known (infeasible CIs)

- **Suppose** (for a moment) that σ^2 is known, and we can compute exactly the variance of $\hat{\beta}_1$:

$$\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}.$$

- Consider the following CI:

$$CI_{1-\alpha} = \left[\hat{\beta}_1 - z_{1-\alpha/2}\sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)},\right.$$
$$\left.\hat{\beta}_1 + z_{1-\alpha/2}\sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}\right].$$

- For example, if $1 - \alpha = 0.95 \iff \alpha = 0.05 \iff z_{1-\alpha/2} = z_{0.975} = 1.96$, and $CI_{0.95}$ is

$$\hat{\beta}_1 \pm 1.96\sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}.$$

Infeasible CI validity (σ^2 known)

- **Goal:** show that $P\left(\beta_1 \in CI_{1-\alpha} \mid \mathbf{X}\right) = 1-\alpha$.
- **Notation:** $\sigma_{\hat{\beta}_1} = \sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}$.
- **Key fact:** $Z = \dfrac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$ conditionally on $\mathbf{X}$.

$$
\begin{aligned}
& P\left(\beta_1 \in CI_{1-\alpha} \mid \mathbf{X}\right) \\
& = P\left(\hat{\beta}_1 - z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \mid \mathbf{X}\right) \\
& = P\left(-z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \leq \beta_1 - \hat{\beta}_1 \leq z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \mid \mathbf{X}\right) \\
& = P\left(-z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \leq \hat{\beta}_1 - \beta_1 \leq z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \mid \mathbf{X}\right) \\
& = P\left(-z_{1-\alpha/2} \leq \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \leq z_{1-\alpha/2} \mid \mathbf{X}\right) \\
& = P\left(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2} \mid \mathbf{X}\right) \\
& = P\left(Z \leq z_{1-\alpha/2} \mid \mathbf{X}\right) - P\left(Z \leq -z_{1-\alpha/2} \mid \mathbf{X}\right) \\
& = (1-\alpha/2) - \alpha/2 \\
& = 1-\alpha.
\end{aligned}
$$

Feasible CIs (σ^2 unknown)

- Since σ^2 is unknown, we must estimate it from the data:

$$
\begin{aligned}
s^2 &= \frac{1}{n-2}\sum_{i=1}^{n} \hat{U}_i^2 \\
&= \frac{1}{n-2}\sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2 .
\end{aligned}
$$

- The **standard error** of $\hat{\beta}_1$ is defined as

$$
\begin{aligned}
\operatorname{se}\left(\hat{\beta}_1\right) &= \sqrt{\widehat{\operatorname{Var}}\left(\hat{\beta}_1\right)} \\
&= \sqrt{\frac{s^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}}.
\end{aligned}
$$

- Replacing σ by its estimate does not give a normal distribution anymore:

$$
\frac{\hat{\beta}_1 - \beta_1}{\operatorname{se}\left(\hat{\beta}_1\right)} \mid \mathbf{X} \sim t_{n-2}.
$$

  Here t_{n-2} denotes the t-distribution with $n-2$ degrees of freedom.
- The degrees of freedom depend on
  - the sample size (n),
  - and the number of parameters one has to estimate to compute s^2 (two in this case, β_0 and β_1).
- Let $t_{df,\tau}$ be the τ-th quantile of the t-distribution with the number of degrees of freedom df: If $T \sim t_{df}$ then

$$
P\left(T \leq t_{df,\tau}\right) = \tau.
$$

- Similarly to the normal distribution, the t-distribution is centered at zero and is symmetric around zero: $t_{n-2,1-\alpha/2} = -t_{n-2,\alpha/2}$.
- We can now construct a feasible $CI_{1-\alpha}$ as

$$\hat{\beta}_1 \pm t_{n-2,1-\alpha/2} \times \text{se}\left(\hat{\beta}_1\right).$$

Example

- Data: `rental` from the `wooldridge` R package. 64 US cities in 1990.
  - `rent`: average monthly rent ($)
  - `avginc`: per capita income ($)
- Model: $\text{Rent}_i = \beta_0 + \beta_1 \text{AvgInc}_i + U_i$.
- R implementation:

```
# Load data and run OLS regression
library(wooldridge)
data("rental")
rental90 <- subset(rental, y90 == 1)
reg <- lm(rent ~ avginc, data = rental90)
summary(reg)
```

```

Call:
lm(formula = rent ~ avginc, data = rental90)

Residuals:
   Min     1Q Median     3Q    Max 
-94.67 -47.27 -13.68  25.65 228.46 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.488e+02  3.210e+01   4.635 1.89e-05 ***
avginc      1.158e-02  1.308e-03   8.851 1.34e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 66.56 on 62 degrees of freedom
Multiple R-squared:  0.5582,	Adjusted R-squared:  0.5511 
F-statistic: 78.34 on 1 and 62 DF,  p-value: 1.341e-12
```

- 95% CI for the slope coefficient:

```
confint(reg, "avginc", level = 0.95)
```

```
             2.5 %     97.5 %
avginc 0.008964625 0.01419539
```

- 90% CI for the slope coefficient:

```
confint(reg, "avginc", level = 0.90)
```

```
               5 %       95 %
avginc 0.009395296 0.01376472
```

The effect of estimating σ^2

- The t-distribution has heavier tails than the normal.
- $t_{df,1-\alpha/2} > z_{1-\alpha/2}$, but as df increases $t_{df,1-\alpha/2} \to z_{1-\alpha/2}$.

- When the sample size n is large, $t_{n-2,1-\alpha/2}$ can be replaced with $z_{1-\alpha/2}$.
- In R, use `qt()` for t-quantiles and `qnorm()` for z-quantiles:

```
# z critical value for 95% CI
qnorm(0.975)
```

```
[1] 1.959964
```

```
# t critical values for 95% CI with different df
qt(0.975, df = 30)
```

```
[1] 2.042272
```

```
qt(0.975, df = 100)
```

```
[1] 1.983972
```

```
qt(0.975, df = 1000)
```

```
[1] 1.962339
```

```
qt(0.975, df = 10000)
```

```
[1] 1.960201
```

Interpretation of confidence intervals

- The confidence interval $CI_{1-\alpha}$ is a function of the **sample** $\{(Y_i, X_i) : i = 1, \ldots, n\}$, and therefore is **random**. This allows us to talk about the probability of $CI_{1-\alpha}$ containing the true value of β_1.
- Once the confidence interval is computed given the data, we have its **one realization**. The realization of $CI_{1-\alpha}$ (the computed confidence interval) is not random, and it does not make sense anymore to talk about the probability that it includes the true β_1.
- **Once the confidence interval is computed, it either contains the true value β_1 or it does not.**