Financial Econometrics
Shangai Jiao Tong University
Summer School – 2025-2026
Problem Set - Lecture 3
With Solutions

Fatih Kansoy

Reference: Stock, J.H. and Watson, M.W., Introduction to Econometrics (4th Edition, Global Edition)

Chapter 5: Regression Hypothesis Tests and Confidence Intervals

Question 1. Suppose that a researcher, using wage data on 200 randomly selected male workers and 240 female workers, estimates the OLS regression:

$$\widehat{Wage} = 10.73 + 1.78 \times Male, \quad R^2 = 0.09, \quad SER = 3.8$$

$$\text{(Standard errors:)} \quad (0.16) \quad (0.29)$$

where $Wage$ is measured in dollars per hour and $Male$ is a binary variable that is equal to 1 if the person is a male and 0 if the person is female. Define the wage gender gap as the difference in mean earnings between men and women.

(a) What is the estimated gender gap?

(b) Is the estimated gender gap significantly different from 0? (Compute the p-value for testing the null hypothesis that there is no gender gap.)

(c) Construct a 95% confidence interval for the gender gap.

(d) In the sample, what is the mean wage of women? Of men?

(e) Another researcher uses these same data but regresses $Wages$ on $Female$, a variable that is equal to 1 if the person is female and 0 if the person is male. What are the regression estimates from this regression?

Solution:

(a) **Estimated gender gap:**

The coefficient on $Male$ represents the difference in mean wages between men and women:

$$\text{Gender gap} = \hat{\beta}_1 = \boxed{\$1.78/\text{hour}}$$

Interpretation: On average, men earn \$1.78 more per hour than women.

(b) **Significance test:**

$H_0 : \beta_1 = 0$ (no gender gap) vs. $H_1 : \beta_1 \neq 0$ (gender gap exists)

Test statistic:

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{1.78}{0.29} = 6.14$$

p-value (two-sided):

$$p\text{-value} = 2 \times \Pr(Z > 6.14) \approx \boxed{0.0000}$$

Conclusion: The gender gap is highly statistically significant. We reject H_0 at any conventional significance level (1%, 5%, 10%).

(c) **95% confidence interval:**

$$\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1) = 1.78 \pm 1.96 \times 0.29 = 1.78 \pm 0.57$$

$$\boxed{[1.21, 2.35]}$$

Interpretation: We are 95% confident that the true gender wage gap is between \$1.21 and \$2.35 per hour.

(d) **Mean wages:**

Women ($Male = 0$):

$$\bar{W}_{women} = \hat{\beta}_0 = \boxed{\$10.73/\text{hour}}$$

Men ($Male = 1$):

$$\bar{W}_{men} = \hat{\beta}_0 + \hat{\beta}_1 = 10.73 + 1.78 = \boxed{\$12.51/\text{hour}}$$

(e) **Regression with $Female$ variable:**

The relationship between the two specifications:

- Original: $Wage = \beta_0 + \beta_1 \cdot Male + u$
- Alternative: $Wage = \gamma_0 + \gamma_1 \cdot Female + u$

Note that $Female = 1 - Male$. The correspondence is:

$$\gamma_0 = \text{Mean wage for men} = \beta_0 + \beta_1 = 10.73 + 1.78 = 12.51$$
$$\gamma_1 = \text{Mean wage for women} - \text{Mean wage for men} = -\beta_1 = -1.78$$

The new regression is:

$$\boxed{\widehat{Wage} = 12.51 - 1.78 \times Female, \quad R^2 = 0.09, \quad SER = 3.8}$$

Note: R^2 and SER are unchanged because this is an equivalent representation of the same model. The standard errors also remain the same: $SE(\hat{\gamma}_0) = SE(\hat{\beta}_0 + \hat{\beta}_1)$ and $SE(\hat{\gamma}_1) = SE(\hat{\beta}_1) = 0.29$.

Question 2. In the 1980s, Tennessee conducted an experiment in which kindergarten students were randomly assigned to "regular" and "small" classes and given standardized tests at the end of the year. (Regular classes contained approximately 24 students, and small classes contained approximately 15 students.) Suppose, in the population, the standardized tests have a mean score of 925 points and a standard deviation of 75 points. Let $SmallClass$ denote a binary variable equal to 1 if the student is assigned to a small class and equal to 0 otherwise. A regression of $TestScore$ on $SmallClass$ yields:

$$\widehat{TestScore} = 918.0 + 13.9 \times SmallClass, \quad R^2 = 0.01, \quad SER = 74.6$$
$$\text{(Standard errors:)} \quad (1.6) \quad (2.5)$$

(a) Do small classes improve test scores? By how much? Is the effect large? Explain.

(b) Is the estimated effect of class size on test scores statistically significant? Carry out a test at the 5% level.

(c) Construct a 99% confidence interval for the effect of $SmallClass$ on $TestScore$.

(d) Does least squares assumption 1 plausibly hold for this regression? Explain.

Solution:

(a) **Effect of small classes:**

The coefficient on *SmallClass* indicates that students in small classes score, on average:

$$\hat{\beta}_1 = \boxed{\text{13.9 points higher}}$$

Is this effect large?

To assess practical significance, compare to the standard deviation:

$$\text{Effect size} = \frac{13.9}{75} \approx 0.19 \text{ standard deviations}$$

By Cohen's conventions:

- 0.2 SD = small effect
- 0.5 SD = medium effect
- 0.8 SD = large effect

Interpretation: The effect is small to moderate—about 1/5 of a standard deviation. This is educationally meaningful: moving from the 50th to roughly the 58th percentile.

(b) **Statistical significance at 5% level:**

$H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$

Test statistic:

$$t = \frac{13.9 - 0}{2.5} = 5.56$$

Critical value (5%, two-sided): $|t| > 1.96$

Since $5.56 > 1.96$: **Reject** H_0

p-value: $2 \times \Pr(Z > 5.56) \approx 0.00$

Conclusion: The effect is **highly statistically significant** at the 5% (and 1%) level.

(c) **99% confidence interval:**

$$\hat{\beta}_1 \pm 2.576 \times SE(\hat{\beta}_1) = 13.9 \pm 2.576 \times 2.5 = 13.9 \pm 6.44$$

$$\boxed{[7.46, 20.34]}$$

Interpretation: We are 99% confident that small classes improve test scores by between 7.5 and 20.3 points.

(d) **Least squares assumption 1:** $E(u_i|X_i) = 0$

Yes, this assumption plausibly holds.

Reasoning:

- Students were **randomly assigned** to small or regular classes.
- Random assignment ensures that $SmallClass$ is **independent** of all student characteristics (observed and unobserved).
- Since u_i captures unobserved factors affecting test scores, $SmallClass \perp u_i$.
- Therefore, $E(u_i|SmallClass_i) = E(u_i) = 0$.

This is the key advantage of randomized experiments: they ensure **internal validity** by eliminating confounding.

Question 3. Suppose (Y_i, X_i) satisfy the least squares assumptions in Key Concept 4.3. A random sample of size $n = 250$ is drawn and yields:

$$\hat{Y} = 5.4 + 3.2X, \quad R^2 = 0.26, \quad SER = 6.2$$

$$\text{(Standard errors:)} \quad (3.1) \quad (1.5)$$

(a) Test $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ at the 5% level.

(b) Construct a 95% confidence interval for β_1.

(c) Suppose you learned that Y_i and X_i were independent. Would you be surprised? Explain.

(d) Suppose Y_i and X_i are independent and many samples of size $n = 250$ are drawn, regressions estimated, and (a) and (b) answered. In what fraction of the samples would H_0 from (a) be rejected? In what fraction of samples would the value $\beta_1 = 0$ be included in the confidence interval from (b)?

Solution:

(a) **Hypothesis test:**

$H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$

Test statistic:

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{3.2}{1.5} = 2.13$$

Critical value (5%, two-sided, large n): $|t| > 1.96$

Since $|2.13| > 1.96$: **Reject H_0 at the 5% level**

p-value: $2 \times \Pr(Z > 2.13) = 2 \times 0.0166 = 0.033$

(b) **95% confidence interval:**

$$\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1) = 3.2 \pm 1.96 \times 1.5 = 3.2 \pm 2.94$$

$$\boxed{[0.26, 6.14]}$$

(c) **Surprise about independence:**

Yes, I would be surprised to learn that Y and X are independent.

Reasoning:

- If Y and X were truly independent, then $\beta_1 = 0$ in the population.
- But we just **rejected** $H_0 : \beta_1 = 0$ at the 5% significance level.
- The p-value was 0.033, meaning there's only a 3.3% chance of observing $\hat{\beta}_1 = 3.2$ or more extreme if $\beta_1 = 0$.

While not impossible (we could have a Type I error), the evidence suggests Y and X are **not** independent.

(d) **Long-run properties if Y and X are truly independent:**

If $\beta_1 = 0$ (independence), then across many repeated samples:

- **Fraction rejecting H_0:**
  When H_0 is true, the probability of Type I error equals the significance level:

  $$\Pr(\text{Reject } H_0 | H_0 \text{ true}) = \alpha = \boxed{5\%}$$

  So 5% of samples would (incorrectly) reject the null.
- **Fraction with $\beta_1 = 0$ in the 95% CI:**
  By definition of a 95% confidence interval:

  $$\Pr(\beta_1 = 0 \in 95\% \text{ CI}) = \boxed{95\%}$$

  So 95% of samples would have the true value $\beta_1 = 0$ within the confidence interval.

Note: These two answers are complementary: $(100\% - 5\%) = 95\%$. Failing to reject H_0 at level α is equivalent to the hypothesized value being inside the $(1 - \alpha)$ confidence interval.