Financial Econometrics
Shangai Jiao Tong University Summer School 2025-2026
Problem Set - Lecture 5

Fatih Kansoy

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

Chapter 7: Hypothesis Tests and Confidence Intervals in Multiple Regression

Questions 7.1–7.6 refer to the following regression table using data on employees in a developing country ($n = 10,973$). The dependent variable is log average weekly earnings (AWE).

Regressor	(1)	(2)	(3)
High school graduate (X_1)	0.352	0.373	0.371
	(0.021)	(0.021)	(0.021)
Male (X_2)	0.458	0.457	0.451
	(0.021)	(0.020)	(0.020)
Age (X_3)		0.011	0.011
		(0.001)	(0.001)
North (X_4)			0.175
			(0.037)
South (X_5)			0.103
			(0.033)
East (X_6)			−0.102
			(0.043)
Intercept	12.84	12.471	12.390
	(0.018)	(0.049)	(0.057)
F-statistic (regional effects = 0)			21.87
SER	1.026	1.023	1.020
R^2	0.0710	0.0761	0.0814

Question 1. For each of the three regressions, add * (5% level) and ** (1% level) to the table to indicate the statistical significance of the coefficients.

Question 2. Using the regression results in column (1):

(a) Is the high school earnings difference estimated from this regression statistically significant at the 5% level? Construct a 95% confidence interval of the difference.

(b) Is the male–female earnings difference estimated from this regression statistically significant at the 5% level? Construct a 95% confidence interval for the difference.

Question 3. Using the regression results in column (2):

(a) Is age an important determinant of earnings? Use an appropriate statistical test and/or confidence interval to explain your answer.

(b) Suppose Alvo is a 30-year-old male college graduate, and Kal is a 40-year-old male college graduate. Construct a 95% confidence interval for the expected difference between their earnings.

Question 4. Using the regression results in column (3):

(a) Are there any important regional differences? Use an appropriate hypothesis test to explain your answer.

(b) Juan is a 32-year-old male high school graduate from the North. Mel is a 32-year-old male college graduate from the West. Ari is a 32-year-old male college graduate from the East.

(i) Construct a 95% confidence interval for the difference in expected earnings between Juan and Mel.

(ii) Explain how you would construct a 95% confidence interval for the difference in expected earnings between Juan and Ari.

Question 5. The regression shown in column (2) was estimated again, this time using data from 1993 (5000 observations selected at random and converted into 2007 units using the Consumer Price Index). The results are:

$$\log AWE = 9.32 + 0.301 \cdot \text{High school} + 0.562 \cdot \text{Male} + 0.011 \cdot \text{Age}$$

$$\text{(SE:)} \quad (0.20) \quad (0.019) \quad (0.047) \quad (0.002)$$

$$SER = 1.25, \quad R^2 = 0.08$$

Comparing this regression to the regression for 2007 shown in column (2), was there a statistically significant change in the coefficient on High school?

Question 6. In all of the regressions in the previous Exercises, the coefficient of High school is positive, large, and statistically significant. Do you believe this provides strong statistical evidence of the high returns to schooling in the labor market?

Question 7. Question 6.5 reported the following regression (where standard errors have been added):

$$\begin{aligned} \widehat{Price} = {} & 109.7 + 0.567 \cdot BDR + 26.9 \cdot Bath + 0.239 \cdot Hsize \\ & (22.1) \quad (1.23) \qquad (9.76) \qquad (0.021) \\ & + 0.005 \cdot Lsize + 0.1 \cdot Age - 56.9 \cdot Poor \\ & (0.00072) \qquad (0.23) \qquad (12.23) \end{aligned}$$

with $R^2 = 0.85$ and $SER = 45.8$.

(a) Is the coefficient on BDR statistically significantly different from zero?

(b) Typically, four-bedroom houses sell for more than three-bedroom houses. Is this consistent with your answer to (a) and with the regression more generally?

(c) A homeowner purchases 2500 square feet from an adjacent lot. Construct a 95% confidence interval for the change in the value of her house.

(d) Lot size is measured in square feet. Do you think that another scale might be more appropriate? Why or why not?

(e) The F-statistic for omitting BDR and Age from the regression is $F = 2.38$. Are the coefficients on BDR and Age statistically different from zero at the 10% level?