Introductory Econometrics
Multiple Hypothesis Testing

by Hieu Nguyen

Fall 2024

1.

File `wage.csv` contains a cross-sectional dataset on 526 working individuals for the year 1976 in the US. Using this labor market data, estimate a simple model describing the impact of years of education and work experience on hourly wage in USD per hour:

$$\text{wage} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \epsilon.$$

(a) Import data into Gretl from the .csv file. Carry out a basic inspection of data (display values, visually, descriptive statistics).

(b) Comment on the expected signs of coefficients β_1 and β_2 first and then estimate the model.

(c) Evaluate the statistical significance of β_1 and β_2 based on the Gretl output.

(d) How much of the variation in wage for these 526 individuals is explained by `educ` and `exper`? Explain.

(e) Estimate also the model without `exper`, compare R^2 and R^2_{adj}. Which is a better model? Why?

(f) Test formally the following hypotheses at the 5% significance level:

  (i) Education has a significant impact on wages.

  (ii) Workforce experience has a significantly positive impact on wages.

  (iii) The regression is overall significant.

(g) Set up a 90% confidence interval for β_2 (and a 99% confidence interval for β_1).

(h) How would the estimated coefficients, standard errors, and t-statistics have differed if we transformed the wage variable into monthly income and `exper` into decades? Explain.

2.

Answer the following questions about data on the sales prices of houses in the UK. The variables in this study are:

- $PRICE_i$: sales price for house i;
- $ASSESS_i$: assessed price of house i;
- $LOTSIZE_i$: size of lot (in square feet) for house i;
- $BDRMS_i$: number of bedrooms for house i;
- $BATH_i$: number of bathrooms for house i;
- $OCEAN_i$: a variable equal to 1 if house i is located within 10 miles of the ocean, 0 otherwise;
- $URBAN_i$: a variable equal to 1 if house i is located in an area classified as urban, 0 otherwise;
- $LAKE_i$: a variable equal to 1 if house i is located within 10 miles of a lake, 0 otherwise;

Table 1: Results of regressions

Dependent variable $\texttt{PRICE}_i$**,** $n = 238$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\texttt{ASSESS}_i$	0.90	0.90	0.91	0.90	0.89	0.90	0.90
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
$\texttt{LOTSIZE}_i$	0.0035	0.00059	0.00059	0.00057	0.00058	0.00059	0.00060
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
$\texttt{BDRMS}_i$	11.5	9.74	7.65	8.74	10.43		
	(2.32)	(3.11)	(3.29)	(3.54)	(3.77)		
$\texttt{BATH}_i$			3.57	3.78			
			(2.24)	(1.11)			
$\texttt{OCEAN}_i$	15.6	14.32	16.76	15.32	14.56		
	(11.43)	(5.21)	(4.32)	(4.98)	(7.01)		
$\texttt{URBAN}_i$		9.54	10.29	12.32			
		(8.99)	(5.43)	(5.22)			
$\texttt{LAKE}_i$				11.36	12.87	11.98	
				(4.28)	(8.32)	(6.43)	
INTERCEPT	261.9	-38.91	-40.30	-43.21	-36.54	-42.37	-38.44
	(11.98)	(6.78)	(7.32)	(6.99)	(5.87)	(7.22)	(9.43)
RSS	145.69	142.99	136.66	134.54	135.38	135.22	136.54
R^2	0.143	0.159	0.196	0.209	0.204	0.205	0.197

- $INTERCEPT$: intercept in the model.

Table 1 lists estimated coefficients with standard errors in parentheses below.

(a) Using the reported regressions, could you test whether the value of the house near water was different from the value of the house away from water at the 5% significance level, controlling for assessed value, lot size, and the number of bedrooms? If so, perform the test. If not, explain what results you would need to do the test.

(b) Could you test whether bathrooms change the house value, controlling for assessed value, lot size, and the number of bedrooms at the 5% significance level? If so, perform the test. If not, explain what results you would need to do the test.

(c) Can you test whether the assessed value and number of bedrooms are jointly significant, controlling for lot size? If yes, perform the test at the 5% significance level. If not, explain what you would need to perform this test.

(d) Could you test whether all 7 of the listed variables (excluding the intercept) are jointly significant at the 5% significance level? Be sure to state any assumptions you are making.