Econ 103: Introduction to Econometrics
Week 5: Practice Questions

Francisco Díaz-Valdés*

Fall 2025

Question 1

A study looked at whether there is a difference in hourly wages (measured in $) between alumni of USC and UCLA, other things being equal. The model is given by

$$\text{Wage } = b_1 + b_2 * (\text{USC})$$

The variable USC is called an indicator variable, and it only has the value 0 or 1. The variable $USC = 1$ if the person is a USC alumni and $USC = 0$ if they are a $UCLA$ alumni.

Some of the regression output is given below:

- $\hat{b}_1 = 70.10, \text{se}(\hat{b}_1) = 2.10$
- $\hat{b}_2 = -20.51, \text{se}(\hat{b}_2) = 3.00$

(a) Interpret the coefficients of the model.

(b) Construct a 99% confidence interval (CI) for the slope of the model. Assume that t-critical=2.576 for a 99% CI.

(c) Test the hypothesis that the average hourly wage for UCLA alumni is $65.00 at the 5% confidence level. Assume that t-critical = 1.96 for a 95% CI.

(a) Note that

$\mathbb{E}[\hat{w}_i | USC_i = 1] = \hat{b}_1 + \hat{b}_2 \rightarrow$ avg. estimated salary for a student who went to USC

$\mathbb{E}[\hat{w}_i | USC_i = 0] = \hat{b}_1 \rightarrow$ avg. estimated salary for a student who went to UCLA

Assumption: sample contains only USC and UCLA alumni

$$\Rightarrow E[\hat{w}_i \mid usc_i = 1] - E[\hat{w}_i \mid usc_i = 0] = \hat{b}_2$$

$\Rightarrow \hat{b}_2$ could be interpret as the "premium lost" for going to USC instead of UCLA

(b) 99% CI for b_2

we Know that $t = \dfrac{\hat{b}_2 - b_2}{se(\hat{b}_2)}$

and $Pr(-t.stat \leq t \leq t.stat) = 0.99$

$$\Rightarrow Pr\left(-t.stat \leq \frac{\hat{b}_2 - b_2}{se(\hat{b}_2)} \leq t.stat\right) = 0.99$$

$$\Rightarrow Pr\left(\underbrace{\hat{b}_2 - t.stat \cdot se(\hat{b}_2)}_{\text{CI lower bound}} \leq b_2 \leq \underbrace{\hat{b}_2 + t.stat \cdot se(\hat{b}_2)}_{\text{CI upper bound}}\right) = 0.99$$

$$CI = [\hat{b}_2 - t.stat \cdot se(\hat{b}_2), \hat{b}_2 + t.stat \cdot se(\hat{b}_2)]$$

Use the calculator and the given values to obtain the numbers

(c) $H_0: b_1 = 65$

$t.crit = 1.96$ for 95%

t - statistic under H_0

$$t = \frac{\hat{b}_1 - b_1}{se(\hat{b}_1)} = \frac{70.1 - 65}{2.1}$$

$$= \frac{5.1}{2.1} > 1.96$$

reject $H_0 : b_1 = 65$
at 95% CI

Question 2

Assume you estimate the simple regression equation below. The data are from a charity where the variable GIFT represents the gift amount (in \$) and MAILYR represents the number of mailings per year.

$$\text{GIFT} = b_1 + (2.65) * \text{MAILYR}$$

You are told the following:

- $se(\hat{b}_1) = 0.74$ with t.stat $= 2.72$
- $se(\hat{b}_2) =?$ with t.stat $= 7.72$

(a) What is the estimated equation intercept?

(b) What is the standard error of the estimated slope?

(c) Interpret the slope coefficient. If each mailing costs \$1.00, is the charity expected to make a net gain on each mailing? Explain.

(d) Construct a 95% confidence interval (CI) estimate of the slope of this relationship.

(a) $\hat{b}_1 = ?$

if $H_0 : b_1 = 0$, then

$$t = \frac{\hat{b}_1}{se(\hat{b}_1)} \Rightarrow \hat{b}_1 = t \cdot se(\hat{b}_1)$$

$$\hat{b}_1 = 2.72 \cdot 0.74$$

$$\hat{b}_1 = 2.0128$$

(b) $se(\hat{b}_2) = ?$

if $H_0 : b_2 = 0$

$$\Rightarrow t = \frac{\hat{b}_2}{se(\hat{b}_2)} \Rightarrow se(\hat{b}_2) = \frac{\hat{b}_2}{t}$$

$$\Rightarrow se(\hat{b}_2) = \frac{2.65}{7.72}$$

$$\Rightarrow se(\hat{b}_2) = 0.343 \,//$$

(c) Interpret b_2

$$\frac{\partial GIFT}{\partial MAILYR} = b_2$$

$$\Rightarrow \Delta GIFT = b_2 \cdot \Delta MAILYR$$

$\Rightarrow$ a extra unit of mail increases, on average, the gift variable by \$2.65

If each mailing cost \$1

$\Rightarrow$ net gain per mail $= \$2.65 - \$1$

$= \$1.65$ //

(d) CI at 95% for b_2

given that we do not know the degrees of freedom, I will use the t-critical provided in the previous question (95%)

$t.stat = 1.96$

$$Pr(-t.stat \leq t \leq t.stat) = 0.95$$

$$t = \frac{\hat{b}_2 - b_2}{se(\hat{b}_2)} = \frac{2.65 - b_2}{0.343}$$

$$\Rightarrow Pr\left(2.65 - 1.96 \cdot 0.343 \leq b_2 \leq 2.65 + 1.96 \cdot 0.343\right) = 0.95$$

$$\Rightarrow Pr\left(1.977 \leq b_2 \leq 3.322\right) = 0.95$$

$CI = [1.977, 3.322]$ //

Question 3

We are given the following model of home prices as a function of the square foot size of the house.

$$\text{price} = b_1 + b_2\text{sqrt}$$

where sqrt is measured in 100 sqft and price in \$1.000s. After running a regression, we see that $\hat{b}_2 = 13.40$ and $\text{se}(\hat{b}_2) = 0.449$.

Find the p-value associated with whether increasing the size of the house by 100sqft is associated with an increase in the price by \$13.000 versus the alternative hypothesis that it increases by more than \$13.000

units! sqrt is in 100 sqft

price is in \$1.000s

$\hat{b}_2 = 13.40$, $se(\hat{b}_2) = 0.449$

Find the p-value of $H_0 : b_2 = 13$

(assume $\alpha = 5\%$) $H_A : b_2 > 13$

$$t_2 = \frac{13.40 - 13}{0.449}$$

$\Rightarrow t_2 = 0.89$

p-value for this alternative is

$$\text{p-value} = Pr(0.89 < t) = 1 - Pr(t < 0.89)$$

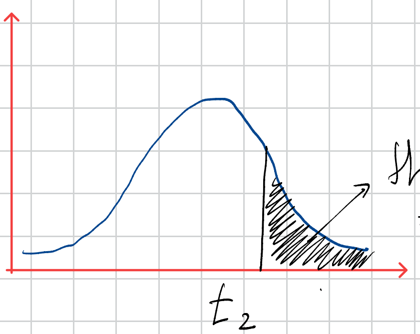
$$= 1 - Pr(t < 0.89)$$

I approximate this probability using a standard normal since we do not know the degrees of freedom

$$= 1 - 0.8133$$

$$= 0.1867 > 0.05$$

do not reject H_0 //



this area = p-value (for this type of H_A)

H_A : alternative hypothesis

Question 4

Let Y denote food expenditure, and let X denote household income. Our model is given by

$$Y = b_1 + b_2 * X$$

(a) After running a regression we are told that $\hat{b}_1 = 10$ and $\hat{b}_2 = 69.81$. Calculate the expected food expenditures for a household with $X = 10$.

(b) Building off part (a), suppose that $\mathbb{V}(\hat{b}_1) = 0.25$, $\mathbb{V}(\hat{b}_2) = 12.50$, and $\mathbb{C}(\hat{b}_1, \hat{b}_2) = 5.98$. Furthermore, suppose we are willing to assume that the residuals are homoskedastic, but we do not know the value of σ. Construct a 95% confidence interval of the expected value from part (a) using the critical value from the standard normal. Is your answer exact?

(c) Next, we want to test whether the expected food expenditures for a household with $X = 10$ equals 710 against the alternative that it does not equal 710. Calculate the $p-value$. Do you reject the null if $\alpha = 0.1$? What about $\alpha = 0.05$?

(a) $\hat{b}_1 = 10, \quad \hat{b}_2 = 69.81$

We are looking for the average estimated expenditure (Y) conditional on household income equal to 10 $(x = 10)$

$$E[\hat{Y} \mid X = 10] = \hat{b}_1 + \hat{b}_2 \cdot 10$$
$$= 10 + 69.81 \cdot 10$$
$$= 10 + 698.1$$
$$= 708.1$$

(b) $V(\hat{b}_1) = 0.25$

$V(\hat{b}_2) = 12.5$

$Cor(\hat{b}_1, \hat{b}_2) = 5.98$

Denote $\hat{\lambda} = E[\hat{Y} \mid X = 10] = \hat{b}_1 + 10 \cdot \hat{b}_2$

We want to construct a CI for λ

using t-stat = 1.96

$$\Rightarrow \quad t = \frac{\hat{\lambda} - \lambda}{se(\hat{\lambda})} \quad = ?$$

$$V(\hat{\lambda}) = V(\hat{b}_1 + 10\hat{b}_2)$$

$$\Rightarrow V(\hat{\lambda}) = V(\hat{b}_1) + 100\, V(\hat{b}_2) + 2 \cdot 10 \cdot cov(\hat{b}_1, \hat{b}_2)$$

$$\Rightarrow V(\hat{\lambda}) = 0.25 + 100 \cdot 12.5 + 20 \cdot 5.98$$

$$\Rightarrow V(\hat{\lambda}) = 1{,}369.85$$

$$\Rightarrow \quad se(\hat{\lambda}) = \sqrt{V(\hat{\lambda})} = 37.011$$

then

$$Pr(\hat{\lambda} - 1.96 \cdot se(\hat{\lambda}) \leq \lambda \leq \hat{\lambda} + 1.96 \cdot se(\hat{\lambda})) = 0.95$$

$$\hat{\lambda} = 708.1$$

$$\hat{\lambda} - 1.96 \cdot se(\hat{\lambda}) = 635.55$$

$$\hat{\lambda} + 1.96 \cdot se(\hat{\lambda}) = 780.64$$

(c) $H_0 : \lambda = 710$, $H_A : \lambda \neq 710$

$$t_\lambda = \frac{\hat{\lambda} - \lambda}{se(\hat{\lambda})}$$

$$= \frac{708.64 - 710}{37.011}$$

$$= -0.036$$

p-value for this type of H_A

is $\text{P-value} = 2 \cdot \Pr(t_\lambda < |t|)$

since t_λ is so small, p-value > 10%

for sure $\Rightarrow$ do not reject H_0

$\Rightarrow$ if we work with,

$\alpha = 5\%$ is the same

result since

p-value > 10% > 5%

do not reject the null

How do I know that t_λ is too small, because at 5%, we are looking for $|t_\lambda| > 1.96$ approx to reject the null and if $\alpha = 10\%$, we are looking for $|t_\lambda| > 1.645$ to reject the null //

Note that as α gets smaller, it is harder to reject the null, because a bigger (in absolute value) t-statistic is needed //

Question 5

Suppose we conduct an experiment where some individuals are treated with a medicine and others are not. Consider the model:

$$Y_i = \beta_1 + \beta_2 D_i + \varepsilon_i$$

Where D_i is equal to 1 if individual i was treated with the medicine and equal to 0 otherwise.

Y_i is the overall health of the individual 6 months after the experiment. Explain in 1 or 2 sentences how to interpret the parameter β_2. In other words, what does β_2 measure?

$$D_i = \begin{cases} 1 & \text{if } i \text{ was treated} \\ 0 & \text{otherwise} \end{cases}$$

$$E[Y_i \mid D_i = 1] = \beta_1 + \beta_2$$

$$E[Y_i \mid D_i = 0] = \beta_1$$

$$\Rightarrow \underbrace{E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0]}_{} = \beta_2$$

$= \beta_2$: average health of treated individuals

Question 6

Let WAGE denote wages and EDUC years of education, and consider the model.

$$\log(\text{WAGE}) = \beta_1 + \beta_2 * \text{EDUC} + \varepsilon$$

We estimate the regression in R and find the following output.

Table 1: Regression Coefficients

	Estimate	Std. Error	t value	$\Pr(>\lvert t\rvert)$
(Intercept)	0.4260866	0.0430064	9.908	$< 2 \times 10^{-16}$ ***
educ	0.1135814	0.0029442	38.577	$< 2 \times 10^{-16}$ ***

Note: Signif. codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Assume that the degrees of freedom are 4,730. Mark each of the following statements either TRUE or FALSE

(a) These estimates suggest that, on average, one additional year of education is associated with an increase in wages of approximately 11.3%.

(b) The confidence interval for the estimate of β_2 will include the value 0.1135814, regardless of the confidence level.

(c) There are 4,733 observations in this sample.

(d) The p-value associated with $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$, is very close to 5%.

$W_i = WAGE_i$

$X_i = EDUC_i$

$\ln(W_i) = \beta_1 + \beta_2 X_i + \varepsilon_i$

(a) $\dfrac{\partial \ln(W_i)}{\partial X_i} = \beta_2$

Note that $\dfrac{\partial \ln(W_i)}{\partial X_i} \approx \dfrac{\Delta\% W_i}{\Delta X_i}$

$\Rightarrow \Delta\% W_i \approx \hat{\beta}_2 \Delta X_i$, $\hat{\beta}_2 = 0.1135814$

if $\Delta X_i = 1 \Rightarrow$ an extra year of educ is associated, on average with an increase of 11,3 % in wages

(a) true

(b) CI at $(1-\alpha)\%$

$\Rightarrow \Pr\left(\hat{\beta}_2 - |t.\text{stat}_{df,\frac{\alpha}{2}}|\, se(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + |t.\text{stat}_{df,\frac{\alpha}{2}}| \cdot se(\hat{\beta}_2)\right) = (1-\alpha)\%$

$\Rightarrow \hat{\beta}_2 \in$ Confidence interval for all α, I used that

$|t.\text{stat}\ df, \tfrac{\alpha}{2}| = |t.\text{stat}\ df, 1-\tfrac{\alpha}{2}|$

since t is symmetrical if $df > 3$

(b) true

(c) df $= N - 2$

df $= 4,730$

$\Rightarrow N = 4,732$ (c) False //

(d) pvalue / false the p-value ≈ 0 not 5%

(d) FALSE

Question 7

We are given 500 observations of single family homes sold in Los Angeles during 2018-2020. The data includes PRICE (in thousands of dollars) and number of windows. The regression model is

$$\text{PRICE} = \beta_1 + \beta_2 * (\text{WINDOWS}^2) + \varepsilon$$

From the data we obtain the estimates

- $\hat{\beta}_1 = 93.56$
- $\hat{\beta}_2 = 0.186$

Compute the elasticity of PRICE with respect to WINDOWS for a home with 20 windows.

Denote $Y = \text{PRICE}$

$X = \text{WINDOWS}$

$$\Rightarrow Y_i = \beta_1 + \beta_2 X_i^2 + \varepsilon_i$$

Elasticity: $\varepsilon_{Y,X}(Y,X) = \frac{dY}{dX} \cdot \frac{X}{Y}$

$$\frac{dY}{dX} = 2\hat{\beta}_2 X \Rightarrow \varepsilon_{Y,X}(Y,X) = \frac{2\hat{\beta}_2 X \cdot X}{Y|X}$$

$$= 2\hat{\beta}_2 \frac{X^2}{Y|X}$$

$$(Y|X=20) = 93.56 + 0.186(20)^2 = 167.96$$

$$= \frac{2 \cdot 0.186 \cdot (20)^2}{167.96}$$

$$= 0.89 //$$

Question 8

Below we summarize the output from a regression of monthly sales (SALES are measured in \$1,000s) on the price of their popular burger (PRICE is measured in dollars).

Table 2: OLS regression of *sales* on *price*

Source	SS	df	MS
Model	1219.09103	1	1219.09103
Residual	1896.39084	73	25.9779567
Total	3115.48187	74	42.1011063

Number of obs	75
$F(1, 73)$	46.93
Prob > F	0.0000
R-squared	0.3913
Adj R-squared	0.3830
Root MSE	5.0969

	Coef.	Std. Err.	t	$P>\|t\|$	[95% Conf. Interval]	
price	-7.829074	1.142865	-6.85	0.000	-10.1068	-5.551348
_cons	121.9002	6.526291	18.68	0.000	108.8933	134.9071

Which of the following statements provide the best interpretation of the slope?

(a) We expect monthly revenue to increase by \$7,829 for a decrease in price of \$1. ✓

~~(b)~~ An increase in price of \$1 will lead to a fall in monthly revenue of \$7,829. → Expected (or average) → not a totally certain amount //

~~(c)~~ An increase in price of \$1 will lead to a increase in monthly revenue of \$7,829. → decrease in revenue

(d) None of the above because the confidence interval endpoints are both negative.

Y is measured in \$1.000

X is measured in \$1

if X decrease by \$1

$\Rightarrow Y$ will decrease by 7.829 on average (or expected)

in dollars will be a

decrease of $7.829 \cdot \$1.000$

$= \$7829$

$\Rightarrow$ (a) is the correct alternative //