Econometrics

Homework 4

Problem 1. Suppose that $X_1, ..., X_n$ is an i.i.d. random sample from the uniform distribution on $[0, \theta]$, where $\theta > 0$ is unknown. Denote $X_{(n)} = \max\{X_1, ..., X_n\}$. Let $Q = X_{(n)}/\theta$ and $c_n = \alpha^{1/n}$. Show that $P(Q \leq c_n) = \alpha$. Note that $P(Q \leq 1) = 1$ and therefore, $P(c_n \leq Q \leq 1) = 1 - \alpha$. Use this result to show that $\left[X_{(n)}, X_{(n)}/c_n\right]$ is a valid $1 - \alpha$ confidence interval for θ:

$$P\left(\theta \in \left[X_{(n)}, X_{(n)}/c_n\right]\right) = 1 - \alpha.$$

Solution. First, note that $Q = X_{(n)}/\theta = \max\{X_1/\theta, ..., X_n/\theta\}$ and X_i/θ is distributed as a uniform distribution on $[0, 1]$. Then,

$$\begin{aligned} P(Q \leq c_n) &= P\left(\max\{X_1/\theta, ..., X_n/\theta\} \leq \alpha^{1/n}\right) \\ &= P\left(X_1/\theta \leq \alpha^{1/n}, X_2/\theta \leq \alpha^{1/n}, ..., X_n/\theta \leq \alpha^{1/n}\right) \\ &= P\left(X_1/\theta \leq \alpha^{1/n}\right) \times \cdots \times P\left(X_n/\theta \leq \alpha^{1/n}\right) \\ &= \left(\alpha^{1/n}\right)^n \\ &= \alpha. \end{aligned}$$

Hence,

$$\begin{aligned} P\left(\theta \in \left[X_{(n)}, X_{(n)}/c_n\right]\right) &= P\left(X_{(n)} \leq \theta \leq X_{(n)}/c_n\right) \\ &= P(c_n \leq Q \leq 1) \\ &= 1 - \alpha. \end{aligned}$$

Problem 2. Find A-C in the Stata output below if the number of observations is 23. Justify your answer. According to the results, is x significant? The significance level is 5%. Explain.

y	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
x	.0910863	A	1.32	0.191	B	C
_cons	.3599422	.0241652	14.90	0.000	.3096879	.4101965

Solution.

$$A: \text{ standard error} = \frac{\widehat{\beta}_1}{t \text{ statistic}} = \frac{0.0910863}{1.32} \approx 0.069.$$

$$\begin{aligned} &B: \text{ lower bound of the confidence interval} \\ &= \widehat{\beta}_1 - t_{21,97.5\%} \times \text{ standard error} \\ &= 0.091 - 2.08 \times 0.069 \\ &\approx -0.053. \end{aligned}$$

$$
\begin{aligned}
C &= \widehat{\beta}_1 + t_{21,97.5\%} \times \text{ standard error} \\
&= 0.091 + 2.08 \times 0.069 \\
&= 0.235.
\end{aligned}
$$

$$\text{p-value} = 0.191 > 0.05 \Longrightarrow X \text{ is not significant.}$$

Problem 3. The variable *rdintens* is expenditures on research and development (R&D) as a percentage of sales. Sales are measured in millions of dollars. The variable *profmarg* is profits as a percentage of sales.
Using the data for 32 firms in the chemical industry, the following equation is estimated:

$$
\begin{aligned}
\widehat{rdintens} &= 0.472 + 0.321 \log(sales) + 0.050 profmarg \\
& (1.369) \quad (0.216) \qquad\qquad\quad (0.046) \\
& n = 32,\ R^2 = 0.099.
\end{aligned}
$$

1. Interpret the coefficient on $\log(sales)$. In particular, if *sales* increases by 10%, what is the estimated percentage point change in *rdintens*? Is this an economically large effect?

2. Test the hypothesis that R&D intensity does not change with *sales* against the alternative that it does increase with sales. Do the test at the 5% and 10% levels.

3. Interpret the coefficient on *profmarg*. Is it economically large?

4. Does *profmarg* have a statistically significant effect on *rdintens*?

Solution.

1. Holding *profmarg* fixed, $\Delta\widehat{rdintens} = 0.321 \times \log(\text{sales}) = (0.321/100) \times [100 \times \Delta\log(\text{sales})] \approx 0.00321(\%\Delta\text{sales})$. Therefore, if $\%\Delta\text{sales} = 10$, $\Delta\widehat{rdintens} \approx 0.032$, or only about $3/100$ of a percentage point. For such a large percentage increase in sales, this seems like a practically small effect.

2. $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 > 0$, where β_1 is the population slope on $\log(sales)$. The t statistic is $0.321/0.216 \approx 1.486$. The 5% critical value for a one-tailed test, with degree of freedom $= 32 - 3 = 29$, is 1.699; so we cannot reject H_0 at the 5% level. But the 10% critical value is 1.311; since the t statistic is above this value, we reject H_0 in favor of H_1 at the 10% level.

3. This is asking for your subjective opinion. No definite answer.

4. Not really. Its t statistic is only 1.087, which is well below even the 10% critical value for a one-tailed test.

Problem 4. Suppose we observe a random sample $\{(Y_i, D_i)\}_{i=1}^n$, where Y_i is the dependent variable and D_i is a binary independent variable: for all $i = 1, 2, ..., n$, $D_i = 1$ or $D_i = 0$. Suppose we regress Y_i on D_i with an intercept. Show: the LS estimate of the slope is equal to the difference between the sample averages of the dependent variable of the two groups, observations with $D_i = 1$ and observations with $D_i = 0$. Hint: The sample average of Y of observations with $D_i = 1$ can be written as $\frac{\sum_{i=1}^n D_i Y_i}{\sum_{i=1}^n D_i}$. What is the sample average of Y of observations with $D_i = 0$? Also note: $D_i = D_i^2$.

Solution. Denote $\overline{D} = n^{-1} \sum_{i=1}^{n} D_i$. The LS estimate is

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} \left(D_i - \overline{D}\right) Y_i}{\sum_{i=1}^{n} \left(D_i - \overline{D}\right)^2} = \frac{\sum_{i=1} \left(D_i - \overline{D}\right) Y_i}{\sum_{i=1}^{n} D_i^2 - n\overline{D}^2} = \frac{\sum_{i=1} D_i Y_i - n\overline{DY}}{n\overline{D} - n\overline{D}^2}.$$

The sample average of Y of observations with $D_i = 0$ is

$$\frac{\sum_{i=1}^{n} (1 - D_i) Y_i}{\sum_{i=1}^{n} (1 - D_i)}.$$

Then,

$$\begin{aligned}
\frac{\sum_{i=1}^{n} D_i Y_i}{\sum_{i=1}^{n} D_i} - \frac{\sum_{i=1}^{n} (1 - D_i) Y_i}{\sum_{i=1}^{n} (1 - D_i)} &= \frac{\sum_{i=1}^{n} D_i Y_i}{n\overline{D}} - \frac{\sum_{i=1}^{n} (1 - D_i) Y_i}{n - n\overline{D}} \\
&= \frac{\left(n - n\overline{D}\right) \sum_{i=1}^{n} D_i Y_i - \left(n\overline{D}\right) \sum_{i=1}^{n} (1 - D_i) Y_i}{n\overline{D}\left(n - n\overline{D}\right)} \\
&= \frac{\sum_{i=1}^{n} D_i Y_i - \overline{D} \sum_{i=1}^{n} D_i Y_i - n\overline{DY} + \overline{D} \sum_{i=1}^{n} D_i Y_i}{n\overline{D} - n\overline{D}^2} \\
&= \widehat{\beta}.
\end{aligned}$$

Problem 5. (25 Points) We are interested in explaining a worker's wage in terms of the number of years of education (educ) and years of experience (exper) using the following model:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u.$$

The estimated parameters by OLS for a sample of $n = 935$ observations are displayed in the table.

Several extensions of this model were considered to address the effects of being married (with the binary variable married) and/or being black (with binary black) or possible nonlinearity on the effect of years of experience.

(a) Test whether the wage regressions for married workers and unmarried workers are the same. Hint: Perform a F test. Model 2 is the unrestricted and Model 1 is the restricted.

(b) Based on a statistical test, do the effects of education and experience depend on the marriage status? Hint: Perform a F test. Model 2 is the unrestricted and Model 3 is the restricted.

(c) What we conclude about the possible nonlinearity of the relationship of $log(wage)$ with respect to the years of experience? Can you conclude that years of experience has no significant effect on $log(wage)$ in Model 5? Make two statistical tests to answer these questions. Hint: Look at Model 5 and Model 6. Use a t test to answer the first question and use a F test to answer the second question.

Variables	Model 1	Model 2	Model 3	Model 5	Model 6
educ	0.07778	0.05316	0.07815	0.071984	0.05571
	(0.00669)	(0.02085)	(0.00653)	(0.00677)	(0.00600)
exper	0.01977	0.00038	0.01829	0.01678	
	(0.00330)	(0.01066)	(0.00330)	(0.01389)	
educ×*married*		0.02813			
		(0.02194)			
exper×*married*		0.01952			
		(0.01120)			
married		-0.38069	0.20926	0.18873	0.21311
		(0.36818)	(0.04272)	(0.04763)	(0.04709)
black				-0.24128	-0.22500
				(0.08417)	(0.08212)
married×*black*				0.03543	0.01071
				(0.09404)	(0.09224)
$exper^2$				0.0000486	
				(0.00058)	
constant	5.50271	5.85694	5.32796	5.46653	5.86609
	(0.11427)	(0.34889)	(0.11574)	(0.12914)	(0.09445)
observations	935	935	935	935	935
R^2	0.13086	0.15705	0.15420	0.18132	0.15417

Solution. (a) Model 2:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ \times married + \beta_4 exper \times married + \beta_5 married + u.$$

The hypotheses are

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0$$

against

$$H_1 : \beta_3 \neq 0 \text{ or } \beta_4 \neq 0 \text{ or } \beta_5 \neq 0.$$

The test statistic is given by

$$F = \frac{R_{ur}^2 - R_r^2}{1 - R_{ur}^2} \times \frac{n-k-1}{q} = \frac{0.15705 - 0.13086}{1 - 0.15705} \times \frac{935 - 5 - 1}{3} \approx 9.6212.$$

The critical value is $F_{3,929,0.95} = 2.60$. So H_0 is rejected and regressions are different.
(b) We test

$$H_0 : \beta_3 = \beta_4 = 0$$

against

$$H_1 : \beta_3 \neq 0 \text{ or } \beta_4 \neq 0.$$

The test statistic is given by

$$F = \frac{R_{ur}^2 - R_r^2}{1 - R_{ur}^2} \times \frac{n-k-1}{q} = \frac{0.15705 - 0.15420}{1 - 0.15705} \times \frac{935 - 5 - 1}{2} \approx 1.5705.$$

The critical value is $F_{2,929,0.95} = 3.00$. We do not reject H_0. We do not find evidence supporting that the effects of education and experience depend on marriage status.
(c) Model 5:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper \\ + \beta_3 married + \beta_4 black + \beta_5 married \times black + \beta_6 exper^2 + u.$$

The first question: consider testing $H_0 : \beta_6 = 0$ against $H_1 : \beta_6 \neq 0$. The t statistic is given by $t = 0.0000486/0.00058 \approx 0.084 < 1.96$. We do not reject H_0. The second question: consider testing $H_0 : \beta_2 = \beta_6 = 0$ against $H_1 : \beta_2 \neq 0$ or $\beta_6 \neq 0$. The test statistic is given by

$$F = \frac{R_{ur}^2 - R_r^2}{1 - R_{ur}^2} \times \frac{n-k-1}{q} = \frac{0.18132 - 0.15417}{1 - 0.18132} \times \frac{935 - 6 - 1}{2} \approx 15.388.$$

The critical value is given by $F_{2,928,0.95} = 3.00$. We reject H_0. We find evidence that experience has significant effect on wage.