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. Consider the following model: $Y = \alpha + \beta X + U$, where $E(U \mid X) = 0$, $E(U^2 \mid X) = \sigma^2 > 0$ and the conditional distribution of U given X is $N(0, \sigma^2)$. (a) What is $E(Y \mid X)$? What is the conditional distribution of Y given X? (b) In this question, use the following result: if the conditional distribution of Y given X is $N\left(m(X), s(X)^2\right)$ $(s(X) > 0)$, then the conditional distribution of $\frac{Y-m(X)}{s(X)}$ given X is $N(0,1)$ and

$$P\left(\frac{Y - m(X)}{s(X)} \leq z \mid X\right) = \Phi(z),$$

where $\Phi(z)$ is the standard normal CDF. Here, the left hand side means the conditional probability of $\frac{Y-m(X)}{s(X)} \leq z$ given X. For a random variable Z, its τ-th quantile ($\tau \in (0,1)$) q_τ is defined by the equation: $P(Z \leq q_\tau) = \tau$. Similarly, a function $q_\tau(X)$ is the τ-th quantile of the conditional distribution of Y given X if

$$P(Y \leq q_\tau(X) \mid X) = \tau.$$

Find the expression of $q_\tau(X)$. Hint: Let z_τ denote the τ-th quantile of the standard normal distribution so that $\Phi(z_\tau) = \tau$. (c) Suppose that $E(U^2 \mid X) = e^{2X}$ and the conditional distribution of U given X is $N\left(0, e^{2X}\right)$. Find the expression of $q_\tau(X)$.

Solution. (a) $\mathrm{E}[Y \mid X] = \alpha + \beta X$ and $Y \mid X \sim \mathrm{N}(\alpha + \beta X, \sigma^2)$. (b)

$$
\begin{aligned}
\Pr\left(\frac{Y - (\alpha + \beta X)}{\sigma} \leq z_\tau \mid X\right) = \tau \quad &\Longrightarrow \quad \Pr(Y \leq \sigma z_\tau + \alpha + \beta X \mid X) = \tau \\
&\Longrightarrow \quad q_\tau(X) = \sigma z_\tau + \alpha + \beta X.
\end{aligned}
$$

(c) $Y \mid X \sim \mathrm{N}\left(\alpha + \beta X, e^{2X}\right)$.

$$
\begin{aligned}
\Pr\left(\frac{Y - (\alpha + \beta X)}{e^X} \leq z_\tau \mid X\right) = \tau \quad &\Longrightarrow \quad \Pr\left(Y \leq e^X z_\tau + \alpha + \beta X \mid X\right) = \tau \\
&\Longrightarrow \quad q_\tau(X) = e^X z_\tau + \alpha + \beta X.
\end{aligned}
$$

Problem 4. 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:

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

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 5. A researcher has data for 50 countries on N, the average number of newspapers purchased per adult in one year, and G, GDP per capita, measured in US \$, and fits the following regression (RSS = residual sum of squares)

$$\hat{N} = 25.0 + 0.020G,\ R^2 = 0.06,\ RSS = 4000.0.$$

The researcher believes that GDP in each country has been underestimated by 50% and that N should have been regressed on G^*, where $G^* = 2G$. Explain, how the following components of the output would have differed: (a) the coefficient of GDP; (b) R^2.

Solution. The coefficient of GDP: $G_i^* - \overline{G}^* = 2\left(G_i - \overline{G}\right)$ for all i. Hence the new slope coefficient is

$$\begin{aligned} b_2^* &= \frac{\sum\left(G_i^* - \overline{G}^*\right)\left(N_i - \overline{N}\right)}{\sum\left(G_i^* - \overline{G}^*\right)^2} \\ &= \frac{\sum 2\left(G_i - \overline{G}\right)\left(N_i - \overline{N}\right)}{\sum 4\left(G_i - \overline{G}\right)^2} \\ &= \frac{b_2}{2} \\ &= 0.01, \end{aligned}$$

where $b_2 = 0.02$ is the slope coefficient in the original regression.
The intercept: The new intercept is

$$b_1^* = \overline{N} - b_2^* \overline{G}^* = \overline{N} - \frac{b_2}{2} 2\overline{G} = \overline{N} - b_2 \overline{G} = b_1 = 25,$$

the original intercept.
RSS: The residual in observation i in the new regression, e_i^*, is given by

$$e_i^* = N_i - b_1^* - b_2^* G_i^* = N_i - b_1 - \frac{b_2}{2} 2G_i = e_i$$

the residual in the original regression. Hence RSS is unchanged. Then R^2 is unchanged since RSS and $\sum \left(N_i - \overline{N}\right)^2$ are unchanged.

Problem 6. 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 7. Consider the following model:

$$Y_i = \beta + U_i,$$

where U_i are iid $N(0, 1)$ random variables, $i = 1, \ldots, n$.

1. Find the OLS estimator of β and its mean, variance, and distribution.
2. Suppose that a data set of 100 observation produced OLS estimate $\widehat{\beta} = 0.167$.
   (a) Construct 90% and 95% symmetric two-sided confidence intervals for β.
   (b) Construct a 95% one-sided confidence interval of the form $[A, +\infty)$ for β. In other words, find a random variable A such that $\Pr(\beta \in [A, +\infty)) = 1 - \alpha$, where $\alpha \in (0, 0.5)$ is a known constant chosen by the econometrician.
   (c) Construct a 95% one-sided confidence interval of the form $(-\infty, A]$ for β.

Solution. The model is $Y_i = \beta + U_i$, with $\{U_i\}_{i=1}^n$ i.i.d random variables and $U_i \sim N(0, 1)$, $i = 1, ..., n$. OLS estimator for β is $\widehat{\beta} = \frac{1}{n}\sum_{i=1}^n Y_i = \overline{Y}$. Notice the following

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^n Y_i = \frac{1}{n}\sum_{i=1}^n (\beta + U_i) = \beta + \frac{1}{n}\sum_{i=1}^n U_i.$$

Hence,

$$\begin{aligned} \mathrm{E}\left[\widehat{\beta}\right] &= \beta + \frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[U_i\right] = \beta \\ \mathrm{Var}(\widehat{\beta}) &= \mathrm{Var}(\frac{1}{n}\sum_{i=1}^n U_i) \\ &= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(U_i) \quad \text{since } U_i\text{'s are i.i.d.} \\ &= \frac{n}{n^2} = \frac{1}{n}. \end{aligned}$$

Since $\widehat{\beta}$ is just a linear combination of iid normal random variables, $\widehat{\beta} \sim N(\beta, \frac{1}{n})$. $\widehat{\beta} = 0.167$. Confidence interval for significance level α is

$$\widehat{\beta} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \le \beta \le \widehat{\beta} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$$

Plugging in the values for $\widehat{\beta} = 0.167$, $\sqrt{\frac{\sigma^2}{n}} = 0.1$, $z_{1-\frac{\alpha}{2}} = 1.645$ when $\alpha = 0.1$, $z_{1-\frac{\alpha}{2}} = 1.96$ when $\alpha = 0.05$. We obtain $CI_{90\%} = [0.0025, 0.3315]$ and $CI_{95\%} = [-0.029, 0.363]$.
One sided confidence interval for significance level $\alpha = 0.05$ of the form $[a, +\infty)$ is

$$\beta \ge \widehat{\beta} - z_{1-\alpha}\sqrt{\frac{\sigma^2}{n}}.$$

Plugging in the values for $\widehat{\beta} = 0.167$, $\sqrt{\frac{\sigma^2}{n}} = 0.1$, $z_{1-\alpha} = 1.645$. We obtain the one-sided confidence interval $CI_{95\%} = [0.0025, \infty)$.
One sided confidence interval for significance level $\alpha = 0.05$ of the form $(-\infty, a]$ is

$$\beta \le \hat{\beta} + z_{1-\alpha}\sqrt{\frac{\sigma^2}{n}}$$

Plugging in the values for $\widehat{\beta} = 0.167$, $\sqrt{\frac{\sigma^2}{n}} = 0.1$, $z_{1-\alpha} = 1.645$. We obtain the one-sided confidence interval $CI_{95\%} = (-\infty, 0.3315]$.