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:

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

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

Problem 3. Consider the following model: $Y = \alpha + \beta X + U$, where $E(U|X) = 0$, $E(U^2|X) = \sigma^2 > 0$ and the conditional distribution of U given X is $N(0, \sigma^2)$. (a) What is $E(Y|X)$? 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

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

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

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

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 $E(U^2|X) = e^{2X}$ and the conditional distribution of U given X is $N\left(0, e^{2X}\right)$. Find the expression of $q_\tau(X)$.

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:

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

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 $rdintens$?

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.

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: $D_i = D_i^2$.

Problem 7. Consider the following model:

$$Y_i = \beta + U_i,$$

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

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.