國立臺北大學115學年度碩士班一般入學考試試題

系（所）組別：經濟學系
科　　　目：統計學

☑可　☐不可使用計算機

I. 選擇題（每題5分，共50分）

1. A person buys a new toothbrush every quarter. Only two colors are available: blue and pink. If the person uses a blue toothbrush in a given quarter, the probabilities of buying a blue or pink toothbrush in the next quarter are 0.3 and 0.7, respectively. If the person uses a pink toothbrush in a given quarter, the probabilities of buying a blue or pink toothbrush in the next quarter are 0.5 and 0.5, respectively. Suppose the person buys a blue toothbrush in quarter 1. Given that the person buys a pink toothbrush in quarter 4, the probability that the toothbrush colors in quarters 2 and 3 are the same lies in which of the following ranges?
   (A) Between 0.35 and 0.39
   (B) Between 0.40 and 0.44
   (C) Between 0.45 and 0.49
   (D) Between 0.50 and 0.54
2. Suppose a random variable has a right-skewed (positively skewed) distribution. Which of the following is correct?
   (A) Mean < Median < Mode
   (B) Mean < Mode < Median
   (C) Median < Mode < Mean
   (D) Mode < Median < Mean
3. Let X and Y be random variables with the following joint probability density function. $f(X, Y) = e^{-X}$, $0 < Y < X$
   First derive E(Y | X), and then evaluate it at X = 2. What is the resulting expected value?
   (A) $\frac{1}{2}$　(B) 1　(C) 2　(D) None of the above
4. A random variable X has the following probability density function. $f(X) = \frac{e^{-X}}{(1+e^{-X})^2}$, $-\infty < X < \infty$
   Let $Y = (1 + e^{-X})^{-1}$. Which of the following distributions does the probability density function of Y, f(Y), belong to?
   (A) Uniform distribution
   (B) Normal distribution
   (C) Exponential distribution
   (D) Poisson distribution
5. Suppose $x_1, x_2, \ldots, x_n$ is a random sample from a normal distribution with mean μ and variance σ^2, where both μ and σ^2 are unknown. Assume $x_1 = 2$, $x_2 = 4$, $x_3 = 4$, $x_4 = 6$, $x_5 = 8$, $x_6 = 8$, and $x_7 = 10$. What is the 90% confidence interval for σ^2? Hint: $\chi^2_{0.05,6} = 1.635$, $\chi^2_{0.05,7} = 2.167$, $\chi^2_{0.95,6} = 12.592$, and $\chi^2_{0.95,7} = 14.067$.
   (A) (3.81, 29.36)
   (B) (4.45, 34.25)
   (C) (3.41, 22.15)
   (D) (3.98, 25.84)
6. Consider the hypothesis test for a normal distribution with mean μ at significance level α: H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$. Which of the following is correct?
   (A) Pr (critical region | H_0) = 1 - α
   (B) Pr (critical region | H_1) = Pr (type II error)
   (C) Pr (reject H_0 | H_0) = p-value
   (D) Pr (reject H_0 | H_1) = power
7. Suppose $x_1, x_2, \ldots, x_n$ is a random sample of size n from the following double exponential distribution. $f(x \mid \beta) = \frac{1}{2\beta} e^{-\frac{|x|}{\beta}}$
   Assume that $x_1 = -3$, $x_2 = 2$, $x_3 = 4$, $x_4 = -1$, and $x_5 = 1$. First derive the maximum likelihood estimator of β, and then evaluate it at the given sample values. What is the resulting estimate?
   (A) 0.6　(B) 1.67　(C) 2.2　(D) 0.45

8. Suppose $x_1, x_2, \ldots, x_n$ is a random sample of size n from a Poisson distribution with mean λ, and assume n = 2k for some integer k. Consider the following estimator. $\hat{\lambda} = \frac{1}{2k}\sum_{i=1}^{k}(x_{2i} - x_{2i-1})^2$ Which of the following is correct?
   (A) $\hat{\lambda}$ is a linear and an unbiased estimator.
   (B) $\hat{\lambda}$ is a linear but biased estimator.
   (C) $\hat{\lambda}$ is not a linear but an unbiased estimator.
   (D) $\hat{\lambda}$ is neither a linear nor an unbiased estimator.

9. Suppose $x_1, x_2, \ldots, x_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2. What is the Cramér-Rao lower bound for σ^2?
   (A) $\frac{\sigma^2}{n}$ (B) $\frac{\sigma^2}{n-1}$ (C) $\frac{2\sigma^4}{n}$ (D) $\frac{2\sigma^4}{n-1}$

10. Let $\{X_n\}$ be a sequence of random variables indexed by the sample size n. Suppose that X_n takes two values, 1 and n, with probabilities 1- $\frac{1}{n}$ and $\frac{1}{n}$, respectively. Which of the following is correct?
   (A) X_n converges in probability.
   (B) X_n converges in mean square.
   (C) X_n converges in both probability and mean square.
   (D) X_n converges in neither probability nor mean square.

II. 選擇題 (每題 5 分，共 50 分)

Instructions. For each question, choose exactly one option (A)-(D). Unless otherwise stated, assume an i.i.d. sample $i = 1, \ldots, n$ and that all stated moments exist. Notation: $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$, $Cov(a, b) = E[(a - Ea)(b - Eb)]$, and $Var(a) = Cov(a, a)$.

1. Consider the linear model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$, where $E(u_i \mid x_{1i}, x_{2i}) = 0$ and $Var(u_i \mid x_{1i}, x_{2i}) = \sigma^2$ (homoskedasticity). Let R_1^2 be the R^2 from the auxiliary regression of x_{1i} on $(1, x_{2i})$. Suppose $\hat{\sigma}^2 = 4.00$, $\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2 = 200$, and $R_1^2 = 0.75$. Using the standard homoskedastic OLS variance formula, what is the implied standard error of $\hat{\beta}_1$?
   (A) 0.071
   (B) 0.100
   (C) 0.283
   (D) 0.400

2. Consider $y_i = \beta_0 + \beta_1 D_i + \beta_2 x_i + \beta_3 (D_i x_i) + u_i$, where $D_i \in \{0,1\}$ and $x_i \in \mathbb{R}$, with $E(u_i \mid D_i, x_i) = 0$. Define the conditional mean $\mu(D, x) = E(y_i \mid D_i = D, x_i = x)$. Which statement is correct?
   (A) $\beta_1 = \mu(1, x) - \mu(0, x)$ for all x
   (B) The marginal effect of x when $D = 1$ is β_2
   (C) $\beta_3 = \left.\frac{\partial\mu(1,x)}{\partial x}\right|_{x=0} - \left.\frac{\partial\mu(0,x)}{\partial x}\right|_{x=0}$
   (D) $\beta_2 + \beta_3$ equals the intercept difference between $D = 1$ and $D = 0$

3. The true model is $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + u_i$, $E(u_i \mid x_i, z_i) = 0$,but you estimate the short regression $y_i = \alpha_0 + \alpha_1 x_i + e_i$ (omitting z_i). Assume the standard OVB formula:
$$plim(\hat{\alpha}_1) = \beta_1 + \beta_2 \frac{Cov(x_i, z_i)}{Var(x_i)}.$$
   Suppose $\beta_1 = 1.0$, $\beta_2 = -2.0$, $Var(x_i) = 4.0$, and $Cov(x_i, z_i) = 1.0$. Which is closest to $plim(\hat{\alpha}_1)$?
   (A) 0.0
   (B) 0.5
   (C) 1.0
   (D) -0.5

4. Consider the simple regression $y_i = \beta_0 + \beta_1 x_i + u_i$. Which condition, if violated *alone*, implies OLS is generally *not BLUE* but can remain unbiased?

(A) $E(u_i \mid x_i) = 0$

(B) $Var(u_i \mid x_i) = \sigma^2$ (constant variance)

(C) $Var(x_i) > 0$

(D) Linear-in-parameters conditional mean: $E(y_i \mid x_i) = \beta_0 + \beta_1 x_i$

5. In the multiple regression $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$, you want to test $H_0: \beta_1 - \beta_2 = 0$ (a single linear restriction).

Suppose you have $\hat{\beta}_1 = 0.80$, $\hat{\beta}_2 = 0.50$, $\widehat{Var}(\hat{\beta}_1) = 0.040$, $\widehat{Var}(\hat{\beta}_2) = 0.010$, $\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) = 0.005$.

Compute the t-statistic $t = \frac{(\hat{\beta}_1 - \hat{\beta}_2) - 0}{\sqrt{\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2)}}$, $\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2) = \widehat{Var}(\hat{\beta}_1) + \widehat{Var}(\hat{\beta}_2) - 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)$.

Which is closest?

(A) 1.50

(B) 2.12

(C) 3.00

(D) 4.24

6. Consider the structural model $y_i = \beta_0 + \beta_1 x_i + u_i$, where $Cov(x_i, u_i) \neq 0$. You have one instrument z_i satisfying $Cov(z_i, x_i) \neq 0$ (relevance), $Cov(z_i, u_i) = 0$ (exogeneity).

Assume all variables are mean-zero for simplicity. The just-identified IV estimand is

$$\beta_1^{IV} = \frac{Cov(z_i, y_i)}{Cov(z_i, x_i)}.$$

Suppose $Cov(z_i, y_i) = -0.18$ and $Cov(z_i, x_i) = 0.12$. What is β_1^{IV}?

(A) -0.67

(B) -0.5

(C) 0.5

(D) -1.50

7. In a just-identified IV setup with one endogenous regressor x_i and one instrument z_i, the first stage (with an intercept) is $x_i = \pi_0 + \pi_1 z_i + v_i$. You estimate $\hat{\pi}_1$ with t-statistic $t_{\pi_1} = 1.9$ (two-sided). A common rule-of-thumb uses the first-stage F statistic for $H_0: \pi_1 = 0$, which in this single-instrument case satisfies $F = t_{\pi_1}^2$. Which statement is most accurate?

(A) F ≈ 1.9, so the instrument is strong

(B) F ≈ 3.6, suggesting potential weak-instrument concerns and distorted 2SLS inference

(C) F ≈ 19, so 2SLS is more efficient than OLS

(D) F ≈ 0.5, so 2SLS is unbiased in finite samples

8. Let a categorical variable $G_i \in \{1,2,3,4\}$ and define dummies $D_{ki} = \mathbf{1}\{G_i = k\}$, $k = 1,2,3,4$. Consider the regression with an intercept: $y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + u_i$. Which statement is correct?

(A) The model is identified because the dummies are mutually exclusive

(B) The model is not identified because $\sum_{k=1}^{4} D_{ki} = 1$ for all i; drop one dummy or drop the intercept

(C) The model is identified if n is large enough

(D) Identification fails only if one group has zero observations

9. Suppose you estimate a log-level model $\ln(y_i) = \beta_0 + \beta_1 x_i + u_i, \quad y_i > 0,$
with $E(u_i \mid x_i) = 0$. Using the standard approximation for small changes, which interpretation of β_1 is correct?
(A) A one-unit increase in x changes y by exactly β_1 units
(B) A one-percent increase in x changes y by approximately β_1 percent
(C) A one-unit increase in x changes $E(y \mid x)$ by approximately $100\beta_1$ percent
(D) β_1 is the price elasticity of demand

10. Consider the wage equation $\ln(wage_i) = \beta_0 + \beta_1 educ_i + u_i$, where u_i contains unobserved ability $ability_i$ and local labor-market quality L_i. Assume $Cov(educ_i, ability_i) > 0$ and L_i directly affects wages. A proposed instrument is distance to nearest college, z_i. Suppose z_i is strongly correlated with L_i (e.g. urban areas have smaller distance and higher wages even holding education fixed), and L_i is omitted from the regression. Which statement is most correct?
(A) The exclusion restriction may fail because z_i can affect $\ln(wage_i)$ through L_i, so $Cov(z_i, u_i)$ may be nonzero
(B) If $Cov(z_i, educ_i) \neq 0$, IV is always consistent regardless of L_i
(C) Adding an intercept is sufficient to control for L_i
(D) If OLS is biased, IV must be unbiased in finite samples