※ 注意：全部題目均請作答於試卷內之「非選擇題作答區」，請標明題號依序作答。

第一部分（共 50 分）

1. A cherry orchard is testing three different irrigation methods (X, Y, Z) to determine their impact on cherry yield (measured in kilograms per tree). The ANOVA summary table is as follows:

Source of Variation	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Between Groups	150.45	2	75.225	(b)	0.098
Within Groups	416.10	30	13.87		
Total	566.55	(a)			

1.1 Calculate the (a) total degrees of freedom and (b) F Statistic. (5 points)

1.2 State the null and alternative hypotheses. (5 points)

1.3 If the significance level is set at 0.05, what can you conclude about the irrigation methods? (5 points)

2. A bank is analyzing the time customers spend waiting in line at their branches. The waiting time (W) is modeled as a random variable following a normal distribution: $W \sim N(\mu, \sigma^2)$.

μ: The true mean waiting time (in minutes).

σ^2: The true variance in waiting time, known to be $\sigma^2 = 9$

The bank collects a sample of $n = 15$ customer waiting times: $w_1, w_2, w_3, \ldots, w_{15}$ and calculates the following sample statistics: sample mean $\bar{w} = 12.2$ minutes.

2.1 Calculate the Cramer-Rao Lower Bound (CRLB) for the variance of $\hat{\mu}$, given that σ^2 and n are known. (5 points)

2.2 Calculate the MLE of μ using the provided sample mean ($\bar{w}$). (5 points).

2.3 Compare the variance of the MLE for μ to the CRLB derived in 3.1. What does this tell you about the efficiency of the MLE for estimating the mean waiting time? (10 points)

3. Suppose $X_1, X_2, \ldots, X_n$ is a random sample drawn from a normal distribution

$$X_i \sim N(\mu, \sigma^2), i = 1,2 \ldots, n$$

Where μ is the population mean, and σ^2 is the population variance (both unknown). X_{n+1} and X_{n+2} are independently and identically distributed from the same distribution as the original sample. You are tasked with predicting $Y = X_{n+1} + X_{n+2}$ using the information from the random sample $\{X_1, X_2, \ldots, X_n\}$.

3.1 Derive (a) the point estimate of $Y = X_{n+1} + X_{n+2}$ based on the sample mean $\bar{X}$ and (b) the variance of Y. (5 points)

3.2 Derive the $100(1-\alpha)\%$ confidence interval for Y, assuming μ and σ^2 are estimated using $\bar{X}$ and S^2 (sample variance). (5 points)

3.3 If $n = 20$, the sample mean $\bar{X} = 10$, and the sample variance $S^2 = 4$. Calculate the 95% confidence interval for Y ($t_{20,0.05} = 1.724$; $t_{19,0.05} = 1.729$; $t_{20,0.025} = 2.085$; $t_{19,0.025} = 2.093$). (5 points)

見背面

第二部分（每題 5 分，共 50 分）：第 1 至 6 為選擇題，可複選，全對才給分；第 7 至 10 為簡答題。

1. A student in the Ag Econ Department is estimating the treatment effect of randomly assigned fake news on NTU students' support for political parties. Her advisor says that it is useful to control for experiment participants' socio-demographic variables. What is the main reason(s) for doing so?

(a) To achieve the conditional independence assumption
(b) To increase model's R-squared
(c) To reduce the standard error of the estimate of the treatment effect
(d) All of the above

2. Another student is studying wage determinants in Taiwan by regressing an individual's wage on her characteristics with OLS. A professor shrugs her shoulders says that life is about the that error term. What do you think the professor is suggesting statistically?

(a) The model's R-squared would be low
(b) There would be heteroskedasticity in the error term
(c) The error term will be correlated with both the outcome and covariate(s)
(d) All of the above

3. Applied economists commonly take the natural logarithm of the outcome variable when running regressions with OLS. What is the main reason(s) for that?

(a) The estimated coefficients can have a percentage change interpretation
(b) To mitigate the issue of heteroskedasticity
(c) To increase R-squared
(d) All of the above

4. Under what circumstance(s) can using a (naturally) logged outcome be problematic with OLS?

(a) The distribution of the outcome variable is highly skewed
(b) The outcome variable has many zeros
(c) The outcome variable has negative values
(d) All of the above

5. A student is estimating the following model to understand how rice yield in kilograms per acre (Y) in county i in year t is associated with air quality (AQ), measured by annual average $PM_{2.5}$ concentration in $\mu g/m^3$, and temperature (TP), measured by annual average temperature in degree Celsius:

$$Y_{it} = \alpha + \beta AQ_{it} + \gamma TP_{it} + \delta(AQ_{it} \times TP_{it}) + \varepsilon$$

Using county-year panel data of Taiwan between year 2000 and 2023, she estimates $\hat{\beta} = -100$, $\hat{\gamma} = -100$, and $\hat{\delta} = -1$. Based on the results, which of the following statements is sensible?

(a) When the annual average $PM_{2.5}$ concentration increases by 1 $\mu g/m^3$, rice yield per acre would on average decrease by 100 kg.
(b) When the annual average $PM_{2.5}$ concentration increases by 1 $\mu g/m^3$, rice yield per acre would decrease by 101 kg.
(c) Without knowing the average of TP, it is difficult to assess the average change in rice yield given a unit change in AQ.
(d) All of the above are nonsensible.

接次頁

6. Following question 5, some colleagues wonder if technological advances could be an omitted variable that bias the results. What are some of the remedies?

(a) Find some measures of technology and control for those measures

(b) Control for year fixed effects

(c) Use time-series regression models

(d) All of the above

7. A data analyst at Google want to estimate the association between a store i's visits (Y) and review scores (X). She is considering the following three specifications with different functional form to provide the most useful insight to the marketing team.

(1) $Y_i = \alpha + \beta X_i + \varepsilon$, where X_i is a continues variable ranging from 1 to 5.

(2) $Y_i = \alpha + \beta_1 1[X_i < 2] + \beta_2 1[X_i \geq 2 \ and \ X_i < 3] + \beta_3 1[X_i \geq 3 \ and \ X_i < 4] + \beta_4 1[X_i \geq 4 \ and \ X_i \leq 5] + \varepsilon$,
where $1[\cdot]$ is a binary indicator variable equal to 1 when condition within the bracket is satisfied and 0 otherwise.

(3) $Y_i = \alpha + \beta_1 1[X_i \geq 2 \ and \ X_i < 3] + \beta_2 1[X_i \geq 3 \ and \ X_i < 4] + \beta_3 1[X_i \geq 4 \ and \ X_i \leq 5] + \varepsilon$,
where $1[\cdot]$ is a binary indicator variable equal to 1 when condition within the bracket is satisfied and 0 otherwise.

Which model would you pick and why? Please keep your answer within 50 words (in either English or Mandarin Chinese, excluding punctuation) and include a word count at the end of your answer. Answers exceeding the word limit will not be graded.

8. Many master's theses, particularly those in business schools, rely on survey data collected through voluntary participation on social media platforms. Discuss the disadvantages of this data collection approach and when such an approach can be appropriate, especially from a sampling perspective. Please keep your answer within 100 words (in either English or Mandarin Chinese, excluding punctuation) and include a word count at the end of your answer. Answers exceeding the word limit will not be graded.

9. Provide an example of "correlation is not causation." Please keep your answer within 50 words (in either English or Mandarin Chinese, excluding punctuation) and include a word count at the end of your answer. Answers exceeding the word limit will not be graded.

10. What is the most important or useful thing you have learned in statistics or econometrics, and why? Please keep your answer within 100 words (in either English or Mandarin Chinese, excluding punctuation) and include a word count at the end of your answer. Answers exceeding the word limit will not be graded.

We would like to thank you for being interested in our program. Hope you enjoy the exam! Our program is a really fun place to be—likely even more fun than you think right now. With the ever-rising craze for sustainability, the career prospects of our graduates are brighter than ever. That said, please keep in mind that whether you join our program or choose a different path, the outcome of your graduate school applications will likely have very low predictive power of your life decades from now. We look forward to seeing you this fall!

試題隨卷繳回