請用 2B 鉛筆作答於答案卡，並先詳閱答案卡上之「畫記說明」。

Part I. 統計與計量 (50%, 選擇題，答案可能不只一個選項，每題 5 分)

1. Given the function:
$$g(x) := \exp(-\lambda x),$$
for $x \in \mathbb{R}_+$ and for some $\lambda > 2$, we define
$$\nu_1 := \int_0^\infty (x + x^2) g(x) \mathrm{d}x$$
and
$$\nu_2 := \int_0^\infty \exp(2x) g(x) \mathrm{d}x.$$
Which of the following is right?
   a. $\nu_1 = \frac{1}{\lambda} + \frac{3}{\lambda^2}$ and $\nu_2 = \frac{\lambda}{\lambda - 1}$.
   b. $\nu_1 = \frac{1}{\lambda} + \frac{3}{\lambda^2}$ and $\nu_2 = \frac{1}{\lambda - 1}$.
   c. $\nu_1 = \frac{1}{\lambda} + \frac{2}{\lambda^3}$ and $\nu_2 = \frac{1}{\lambda - 2}$.
   d. $\nu_1 = \frac{1}{\lambda} + \frac{2}{\lambda^3}$ and $\nu_2 = \frac{\lambda}{\lambda - 2}$.
   e. None of the above choices (a)-(d).

2. Let $\{X_i\}_{i=1}^n$ be a sequence of independently and $N(\mu, \sigma^2)$-distributed random variables. Suppose that we utilize Chebyshev's inequality to establish the following result:
$$P(A) \le \alpha,$$
where
$$A := \left\{ \left| \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \sigma^2 \right| \ge \frac{1}{2} \right\}.$$
Which of the following is right?
   a. $\alpha = \frac{8\sigma^4}{n}$.
   b. $\alpha = \frac{4\sigma^4}{n}$.
   c. $\alpha = 0.04$ when $\sigma = 1$ and $n = 100$.
   d. $\alpha \to 0$, as $n \to \infty$.
   e. None of the above choices (a)-(d).

3. Suppose that Y is a continuous random variable with $E[Y] = \mu$ and $var[Y] = \sigma^2$, and X is a discrete random variable with K possible outcomes: $x_1, x_2, \ldots, x_K$. Denote $p_k := P(X = x_k)$, $\mu_k := E[Y|X = x_k]$ and $\sigma_k^2 := var[Y|X = x_k]$, for $k = 1, 2, \ldots, K$. Which of the following is right?
   a. $\mu = \sum_{k=1}^K p_k \mu_k$ and $\sigma^2 = \sum_{k=1}^K p_k(\sigma_k^2 + \mu_k^2) - \mu^2$.
   b. $\mu = \sum_{k=1}^K p_k \mu_k$ and $\sigma^2 = \sum_{k=1}^K p_k \sigma_k^2$.

見背面

c. $\mu = \sum_{k=1}^{K} p_k \mu_k$ and $\sigma^2 = \sum_{k=1}^{K} p_k \sigma_k^2 - \mu^2$.

d. $\mu \neq \sum_{k=1}^{K} p_k \mu_k$ and $\sigma^2 \neq \sum_{k=1}^{K} p_k \sigma_k^2$.

e. None of the above choices (a)-(d).

4. Assume that $Y = X + X^2 + Z + Z^2$, where X and Z are two uncorrelated standard normal random variables. Which of the following is right?

a. $E[Y^2] = 8$ and $E[Y^2 X] = 10$.

b. $E[Y^2] = 10$ and $E[Y^2 X] = 8$.

c. $E[Y^2] = 12$ and $E[Y^2 X] = 6$.

d. $E[Y^2] = 14$ and $E[Y^2 X] = 4$.

e. None of the above choices (a)-(d).

5. Suppose that $\{Y_i\}_{i=1}^{n}$ is a sequence of independently and $N(\mu, \sigma^2)$-distributed random variables. Let $f(\cdot|\mu, \sigma^2)$ be the probability density function of $N(\mu, \sigma^2)$. Also, let $\hat{\mu}$ and $\hat{\sigma}^2$ be the maximum likelihood estimators for μ and σ, respectively, that are solved by maximizing the log-likelihood function:

$$L_n(\mu, \sigma^2) := \frac{1}{n} \sum_{i=1}^{n} \log f(Y_i|\mu, \sigma^2).$$

Which of the following is right?

a. $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\mu})^2$.

b. $E\left[\frac{\partial}{\partial \mu} \log f(Y_i|\mu, \sigma^2)\right] \neq E\left[\frac{\partial}{\partial \sigma^2} \log f(Y_i|\mu, \sigma^2)\right]$.

c. $E\left[\frac{\partial}{\partial \mu} \log f(Y_i|\mu, \sigma^2) \frac{\partial}{\partial \sigma^2} \log f(Y_i|\mu, \sigma^2)\right] = 0$.

d. $var\left[\frac{\partial}{\partial \sigma^2} \log f(Y_i|\mu, \sigma^2)\right] = \frac{1}{2}$ if $\sigma^2 = 1$.

e. None of the above choices (a)-(d).

6. Let $\{X_i\}_{i=1}^{n}$ be a sequence of independently and identically distributed random variables with $\mu := E[X_i]$ and a finite variance $\sigma^2 := var[X_i]$. Assume that $n > 2$. Denote $\bar{X} := \frac{1}{n-2} \sum_{i=1}^{n} X_i$ and $\hat{\sigma}^2 := \frac{1}{n-2} \sum_{i=1}^{n} (X_i - \bar{X})^2$. Which of the following is right?

a. $n^{1/2} \bar{X} / \hat{\sigma}$ has the limiting distribution $N(0, 1)$, as $n \to \infty$, in general.

b. $n^{1/2} \bar{X} / \hat{\sigma}$ has the limiting distribution $N(0, 1)$, as $n \to \infty$, when $\mu = \delta / \sqrt{n}$ for some fixed $\delta \neq 0$.

c. $n^{1/2} \bar{X} / \hat{\sigma}$ has the limiting distribution $N(\delta/\sigma, 1)$, as $n \to \infty$, when $\mu = \delta / \sqrt{n}$ for some fixed $\delta \neq 0$.

接次頁

d. $n\bar{X}^2/\hat{\sigma}^2$ has the limiting distribution $\chi^2(1)$, as $n \to \infty$, when $\mu = \delta/\sqrt{n}$ for some fixed $\delta \neq 0$.

e. None of the above choices (a)-(d).

7. Let X be an $n \times 3$ matrix of Gaussian random variables with $n > 3$. Assume that $X'X$ is positive definite. Define the matrix:

$$V := X(X'X)^{-1}X'.$$

Which of the following is right?

a. V is positive definite.

b. $V \neq VV$.

c. $\text{trace}(VVV) = 3$.

d. $\text{rank}(VV) = n - 3$.

e. None of the above choices (a)-(d).

8. Let $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ be a sequence of independently and identically distributed random vectors with finite variances. Consider the following two simple regressions:

$$M_1 \ : \ Y_i = \beta X_i + e_i$$

and

$$M_2 \ : \ Y_i = \gamma Z_i + u_i.$$

Let $\hat{\beta}$ and $\hat{\gamma}$ be the LS estimators for β and γ, respectively. Suppose that $\hat{\beta}$ and $\hat{\gamma}$ are, respectively, consistent for β_o and γ_o under the condition: $E[e_i|X_i, Z_i] = 0$. Which of the following is right?

a. $\gamma_o = 0$ if $E[Z_iX_i] = 0$.

b. $\gamma_o = \beta_o$ if $X_i = Z_i + V_i$, where V_i is uncorrelated to Z_i, and $E[V_i] = 0$.

c. $\gamma_o = 0.5\beta_o$ if $E[Z_iX_i] = 0.5E[Z_i^2]$.

d. $\gamma_o = \frac{1}{3}\beta_o$ if X_i and Z_i are two independent $\chi^2(1)$-distributed random variables.

e. All of the above choices (a)-(d).

9. Assume that $\{(X_i, Z_i, e_i)\}_{i=1}^n$ is a sequence of independently and identically distributed random vector. In particular, the random vector (X_i, Z_i, e_i) has the distribution $N(0, \Sigma)$, where Σ is a 3×3 matrix:

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho_1 & \rho_2 \\ \rho_1 & \sigma_z^2 & \rho_3 \\ \rho_2 & \rho_3 & \sigma_e^2 \end{bmatrix}.$$

見背面

國立臺灣大學 112 學年度碩士班招生考試試題

Suppose that Y_i is generated by the following regression:

$$Y_i = \beta X_i + e_i,$$

for some $\beta \in \mathbb{R}$ and $i = 1, 2, \ldots, n$. Define two estimators: $\hat{\beta}_1 := \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$ and $\hat{\beta}_2 := \frac{1}{n}\sum_{i=1}^{n} Z_i Y_i$. Which of the following is right?

a. $\hat{\beta}_1$ is consistent for β if $\sigma_x^2 = 1$ and $\rho_2 = 0$.

b. $\hat{\beta}_1$ is consistent for β if $\sigma_x^2 = 1$ and $\rho_2 \neq 0$.

c. $\hat{\beta}_2$ is consistent for β if $\rho_1 = 1$ and $\rho_3 = 0$.

d. $var[\sqrt{n}(\hat{\beta}_1 - \hat{\beta}_2)] \to \infty$, as $n \to \infty$, if $\sigma_x^2 = 1$, $\rho_1 = 1$ and $\rho_2 = \rho_3 = 0$.

e. None of the above choices (a)-(d).

10. Let $\{(Y_i, X_i, Z_i)\}_{i=1}^{n}$ be a sequence of independently and identically distributed random vectors, where X_i has the logistic probability density function:

$$f(x) = \frac{\exp(-x/\kappa)}{\kappa(1+\exp(-x/\kappa))^2},$$

for $x \in \mathbb{R}$ and for some $\kappa > 0$, Z_i is a $\chi^2(q)$-distributed random variable which is independent of X_i, and $Y_i = X_i Z_i$ for all i's. Consider a simple regression:

$$Y_i = \beta X_i + e_i,$$

where β is the regression coefficient, and e_i is the regression error. Let $\hat{\beta}$ be the least squares estimator for β. Which of the following is right?

a. $\hat{\beta}$ converges in probability to κ/q, as $n \to \infty$.

b. $\hat{\beta}$ converges in probability to q/κ, as $n \to \infty$.

c. $\hat{\beta}$ converges in probability to κ, as $n \to \infty$.

d. $\hat{\beta}$ converges in probability to q, as $n \to \infty$.

e. None of the above choices (a)-(d).

接次頁

Part II. 統計個案分析 (50%, 選擇題，答案可能不只一個選項，每題 5 分)

Ash Ketchum, the owner of Pikachu, surveyed attack powers of all 23 Pokémon monsters after he won the 2022 Pokémon World Championships. He studied all monsters for five years and regress attack powers of Pokémon on their health points (HPs). A simple linear regression analysis was specified as

$y_{it}=a_0+ a_1 x_{it} + e_{it},$

where i and t indicate subscripts for monster and year, respectively. Y is the attack power, and x is the HP. Regression was performed by ordinary least squares (OLS) method with results below.

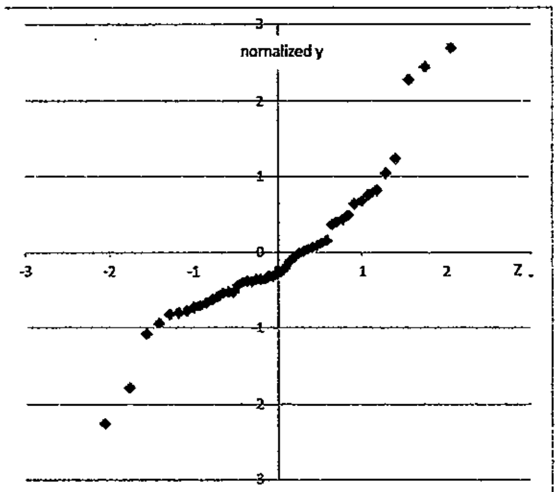
Summary	
R-sq	0.090
Adj. R-sq	☐
SE	56.800
N	115.000

ANOVA	DF	SS	MS	F	P-value
Regression	1.000	36129.233	36129.233	11.199	0.001
Residual	113.000	364560.781	3226.202		
Total	114.000	400690.014			

Regression estimates	Coefficient	SE	t-value	P-value
Intercept	91.344	17.096	5.343	0.000
HP	1.732	0.518	3.346	0.001

11. The data form of Ash's study could be known as
    a. Cross-sectional
    b. Time-series
    c. Panel
    d. Count
    e. Categorical

12. Ash wants to know the distribution of the y variable, i.e., the attack power, by QQ-plot. To plot the QQ-plot, he obtained a figure in the right. Which one(s) of following statement should be true?
    a. Data are not randomly sampled
    b. Data are left skewed
    c. Data are right skewed
    d. Data are fat-tailed
    e. Data are generally a Poisson distribution



13. Based on information from ANOVA, the unconditional variance of y, which is equal to $\Sigma(y_{it}-\bar{y})^2/(n-1)$, is about
    a. 56.800
    b. 61.523
    c. 3226.201
    d. 3514.825
    e. 4058.211

14. Ash does not interpret the volatility of y by standard error but by variance. Which of following one(s)

is right statement(s) of the standard error (SE) under assumptions of *i.i.d.*?

a. SE is a non-linear function of y
b. SE is a biased estimator
c. SE is an inconsistent estimator
d. SE has the same unit of y
e. SE is more volatile than the variance

15. In the ANOVA table, MS indicates

a. Mean of Squares
b. Sum of Squares divided by corresponding degree of freedom
c. Mode of Symbol
d. Max Security
e. Maximum of Squares

16. The adj. R-sq is not presented in the tables. The adj. R-sq is equal to.

a. 0.082
b. 0.090
c. 0.910
d. 0.918
e. 0.995

17. What is SS of the independent variable x?

a. 106.112
b. 3226.202
c. 10240.904
d. 400690.014
e. 36129.232

18. Given the data form (preferring to the answer in question 1 of this case study),which of following statement(s) is(are) true?

a. The cross-sectional variations in the data generally violate the normality assumption in a large sample
b. The time series variations in the data generally violate the independence assumption
c. We can further specify a moving-average regression model for the residual e to fit the identical assumption.
d. The cross-sectional variations in the data generally can be modeled by an autoregressive regression model for the residual e.
e. Regression model using panel form of data cannot be estimated by maximum likelihood method.

19. Ash considered an alternative regression model that include a new variable and year fixed effect:
$y_{it}=a_0+ a_1 x_{it} +a_2 w_t +b_1 v2018_t +b_2 v2019_t +b_3 v2020_t +b_4 v2021_t +b_5 v2022_t + e_{it}$.

接次頁

v_t indicates year fixed effect. New variable, w_t, is the time trend, which is equal to 1, 2, 3, 4 and 5 for year 2018, 2019, 2020, 2021 and 2022, respectively. *v2018, v2019,...* and *v2022* are dummies for year fixed effects. Which one(s) of following statements is(are) true?

a. Coefficient a_2 is generally negative.
b. Average of b_1 to b_5 is zero.
c. Results can be estimated by OLS
d. A linear combination of w_t, *v2018, v2019,...* and *v2022* could be a constant.
e. There is a perfect collinearity issue.

20. Ash also plans to estimate following regression model:

$x_{it}=d_0+ d_1\, y_{it} + \varepsilon_{it},$

where i and t indicate subscripts for monster and year, respectively. Y is the attack power, and x is the HP. Regression was performed by ordinary least squares (OLS) method. Which one(s) of following statements is(are) true?

a. $d_1 < 0.01$
b. $d_1 = 0.052$
c. $d_1 = 0.577$
d. $d_1 > 1$
e. R-sq of regressing x on y is 0.073

試題隨卷繳回