Multiple Choice Questions. Notes:

(1) Please choose only one of the answer choices (a)-(e).

(2) Write down your answers on the scantron answer sheet.

(3) Each question is worth 5 points.

1. Suppose that X is a random variable that has the characteristic function $\exp(-\frac{1}{2}\lambda^2)$, where λ is a real number. Denote $Z_1 := (1+X)^4$ and $Z_2 := \sin(X) + \frac{X}{1+X^2}$. Which of the following is right?

   a. $E(Z_1) = 8$ and $E(Z_2) = -2$
   b. $E(Z_1) = 10$ and $E(Z_2) = 0$
   c. $E(Z_1) = 12$ and $E(Z_2) = 2$
   d. $E(Z_1) = 14$ and $E(Z_2) = 4$
   e. None of the above choices (a)-(d).

2. Let X be a random variable that has the probability density function:

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

   where $x \in \mathbb{R}$ and $\varsigma > 0$ is the parameter. In addition, we define $Z_1 := \frac{d}{d\varsigma}\ln f(X,\varsigma)$ and $Z_2 := Z_1^2 + \frac{d^2}{d\varsigma^2}\ln f(X,\varsigma)$. Which of the following is right?

   a. $E(Z_1) = \varsigma\pi$ and $E(Z_2) = \frac{\pi^2\varsigma^2}{3}$
   b. $E(Z_1) = 0$ and $E(Z_2) = 0$
   c. $E(Z_1) = \varsigma\pi$ and $E(Z_2) = \frac{2\pi^2\varsigma^2}{3}$
   d. $E(Z_1) = \frac{2}{3}\varsigma\pi$ and $E(Z_2) = \frac{\pi^4\varsigma^2}{5}$
   e. None of the above choices (a)-(d).

3. Let $\{Z_i\}_{i=0}^{n}$ be an IID sequence of $N(0,1)$-distributed random variables. Consider the following two processes:

$$X_i = X_{i-1} + Z_i,$$

   with $X_0 := Z_0$, and

$$Y_i = Z_{i-1} + Z_i.$$

   Define $\bar{Y} := n^{-1}\sum_{i=1}^{n} Y_i$. Please calculate $var(X_n)$ and $var(\bar{Y})$.

   a. $var(X_n) = n$ and $var(\bar{Y}) = \frac{4n-8}{n^2}$
   b. $var(X_n) = n+1$ and $var(\bar{Y}) = \frac{4n-6}{n^2}$
   c. $var(X_n) = n+1$ and $var(\bar{Y}) = \frac{4n-2}{n^2}$

見背面

d. $var(X_n) = n$ and $var(\bar{Y}) = \frac{4n+2}{n^2}$
e. None of the above choices (a)-(d).

4. Let $\{X_i\}_{i=1}^n$ be an IID sequence of $N(0,1)$ random variables, and $\{Z_i\}_{i=1}^n$ be another IID sequence of $N(0,1)$ random variables. In addition, $\{X_i\}_{i=1}^n$ is independent of $\{Z_i\}_{i=1}^n$, and we define $Y_i := 2X_i^3 + 3Z_i^4$. Consider a simple linear regression:

$$Y_i = \beta_0 + \beta_1 X_i + e_i,$$

where β_0 and β_1 are regression coefficients, and e_i is a zero-mean error. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the ordinary least squares estimators for β_0 and β_1, respectively. Suppose that $\hat{\beta}_0$ and $\hat{\beta}_1$ are, respectively, consistent for β_0^* and β_1^*, as $n \to \infty$. Please calculate β_0^* and β_1^*.

a. $\beta_0^* = 9$ and $\beta_1^* = 6$
b. $\beta_0^* = 7$ and $\beta_1^* = 4$
c. $\beta_0^* = 5$ and $\beta_1^* = 3$
d. $\beta_0^* = 1$ and $\beta_1^* = 1$
e. None of the above choices (a)-(d).

5. Let $\{Y_i\}_{i=1}^n$ be an IID sequence of Student's $t(2)$-distributed random variables. Denote $\mu := E(Y)$, $\bar{Y} := n^{-1}\sum_{i=1}^n Y_i$, $\hat{\sigma}^2 := n^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$ and

$$Z_n := \frac{\sqrt{n}(\bar{Y} - \mu)}{\hat{\sigma}}.$$

Note that "$\xrightarrow{d}$" means that "convergence in distribution." Which of the following is right?

a. $Z_n \xrightarrow{d} N(0,4)$ and $Z_n^2 \xrightarrow{d} \chi^2(4)$, as $n \to \infty$
b. $Z_n \xrightarrow{d} N(0,3)$ and $Z_n^2 \xrightarrow{d} \chi^2(3)$, as $n \to \infty$
c. $Z_n \xrightarrow{d} N(0,2)$ and $Z_n^2 \xrightarrow{d} \chi^2(2)$, as $n \to \infty$
d. $Z_n \xrightarrow{d} N(0,1)$ and $Z_n^2 \xrightarrow{d} \chi^2(1)$, as $n \to \infty$
e. None of the above choices (a)-(d).

6. Let $\{X_i\}_{i=1}^n$ be an IID sequence of random variables with the zero mean and the variance $\sigma^2 < \infty$. Define $\bar{X} := n^{-1}\sum_{i=1}^n X_i$. Let $C_n(\cdot)$ be the characteristic function of the statistic $n^{1/2}\bar{X}/\sigma$, and $C(\cdot)$ be the characteristic function of the random variable X_i. We also let λ be an arbitrary real number. Which of the following is right?

接次頁

a. $C_n(\lambda) = C(\frac{\lambda}{\sigma\sqrt{n}})^n$

b. $C_n(\lambda) = \frac{1}{\sigma\sqrt{n}} C(\frac{\lambda}{\sigma\sqrt{n}})^n$

c. $C_n(\lambda) = \frac{1}{\sigma^2\sqrt{n}} C(\frac{\lambda}{\sigma\sqrt{n}})^{n-2}$

d. $nC_n'(\lambda) = \frac{1}{\sigma^2\sqrt{n}} C(\frac{\lambda}{\sigma\sqrt{n}})^{n-1}$

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

7. Let Y_i be a random variable with two possible outcomes: 0 and 1, and X_i be a $N(0,1)$-distributed random variable. Suppose that $\{(Y_i, X_i)\}_{i=1}^n$ is an IID sequence, and $P(Y_i = 1|X_i) = p(X_i) > 0$. Consider a simple linear regression:

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

where e_i is a zero-mean error. Let $\hat{\beta}$ be the ordinary least squares estimator for β, and define $\hat{e}_i := Y_i - \hat{\beta} X_i$ and $\hat{\sigma}^2 := n^{-1} \sum_{i=1}^n \hat{e}_i^2$. Suppose that $\hat{\beta}$ and $\hat{\sigma}^2$ are, respectively, consistent for β^* and σ^{*2}, as $n \to \infty$. Please calculate β^* and σ^{*2}.

a. $\beta^* = E(p^2(X_i))$ and $\sigma^{*2} = E(p(X_i)(1 - p(X_i)))$

b. $\beta^* = E(X_i p(X_i))$ and $\sigma^{*2} = E(p(X_i)(1 - p(X_i)))$

c. $\beta^* = E(X_i p(X_i))$ and $\sigma^{*2} = \beta^{*2} + E(p(X_i))$

d. $\beta^* = E(p^2(X_i))$ and $\sigma^{*2} = \beta^{*2} + E(p(X_i)(1 - p(X_i)))$

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

8. Let $\{(Y_i, X_{1i}, X_{2i})\}_{i=1}^n$ be an IID sequence of trivariate normal random variables. The mean vector and the covariance matrix of $(Y_i, X_{1i}, X_{2i})'$ are unknown. Consider the following linear regression:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i,$$

where e_i is a zero-mean error. Let P be the p-value of the t statistic for the following hypotheses:

$$H_0 : \beta_2 = 0;$$
$$H_1 : \beta_2 > 0.$$

Please calculate the second moment and the variance of P under H_0.

a. $E(P^2) = 1/5$ and $var(P) = 1/16$

b. $E(P^2) = 1/4$ and $var(P) = 1/14$

c. $E(P^2) = 1/3$ and $var(P) = 1/12$

d. $E(P^2) = 1/2$ and $var(P) = 1/10$

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

9. Let $\{X_i\}_{i=1}^n$ be an IID sequence of random variables with $\mu = E(X_i)$ and $\sigma^2 = var(X_i)$, and denote $\bar{X} := n^{-1}\sum_{i=1}^n X_i$. Suppose that we have the following inequality:

$$P(|\bar{X} - \mu| \geq \delta) \leq \frac{var(\bar{X})}{\delta^2},$$

for an arbitrary constant $\delta > 0$. Which of the following is right as $n \to \infty$?

a. $var(\bar{X}) = \frac{\sigma^2}{n-1}$, and we may use this inequality to show the normality of $\bar{X}$

b. $var(\bar{X}) = \frac{\sigma^2}{n}$, and we may use this inequality to show the efficiency of $\bar{X}$

c. $var(\bar{X}) = \frac{\sigma^2}{n}$, and we may use this inequality to show the normality of $\bar{X}$

d. $var(\bar{X}) = \frac{\sigma^2}{n}$, and we may use this inequality to show the consistency of $\bar{X}$

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

10. Let $\{Z_i\}_{i=1}^n$ be an IID sequence of logistic random variables with $\mu = E(Z_i)$ and $\sigma^2 = var(Z_i)$. Denote $\bar{Z} := n^{-1}\sum_{i=1}^n Z_i$, $\hat{v}_i = Z_i - \bar{Z}$ and $v_i := Z_i - \mu$. Which of the following is right?

a. $n^{-1/2}\sum_{i=1}^n \hat{v}_i - n^{-1/2}\sum_{i=1}^n v_i$ degenerates to zero, as $n \to \infty$

b. $n^{-1/2}\sum_{i=1}^n \hat{v}_i^2 - n^{-1/2}\sum_{i=1}^n v_i^2$ degenerates to zero, as $n \to \infty$

c. $\left(n^{-1}\sum_{i=1}^n v_i^2\right) n^{1/2}(\bar{Z} - \mu)$ degenerates to zero, as $n \to \infty$

d. $n(\bar{Z} - \mu)^2$ degenerates to zero, as $n \to \infty$

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

11. Suppose that A and B are two events. $P(A)$ is the probability of the happening of event A. If $P(A|B) < P(A)$, which of the following is right.

a. $P(A) < P(B)$;

b. $P(B) < P(A)$;

c. $P(A) < P(A \cap B)$;

d. $P(A|B) < P(A \cap B)$;

e. $P(B|A) < P(B)$.

12. The number of red chips and white chips in an urn is not known, but it is known that the proportion, p, of reds is either 1/5, 1/3, 1/2, or 3/4. A sample of size 4, drawn with replacement, yields the sequence red, white, white, white. The MLE for p is:

a. 1/5;

b. 1/4;
c. 1/3;
d. 1/2;
e. 3/4.

13. Let $\{X_i\}_{i=1}^4$ be an IID sequence of random variables drawn from a normal distribution $N(\mu, \sigma^2)$. Define $\bar{X} := \frac{1}{3}\sum_{i=1}^3 X_i$. Suppose that Z is the standard normal distribution ($Z \sim N(0,1)$). What is the probability that $X_4 \in [\bar{X} - 2\sigma, \bar{X} + 2\sigma]$?

    a. $Pr(X_4 \in [\bar{X} - 2\sigma, \bar{X} + 2\sigma]) = Pr(Z \in [-\sqrt{2}, \sqrt{2}])$;
    b. $Pr(X_4 \in [\bar{X} - 2\sigma, \bar{X} + 2\sigma]) = Pr(Z \in [-\sqrt{3}, \sqrt{3}])$;
    c. $Pr(X_4 \in [\bar{X} - 2\sigma, \bar{X} + 2\sigma]) = Pr(Z \in [-2, 2])$;
    d. $Pr(X_4 \in [\bar{X} - 2\sigma, \bar{X} + 2\sigma]) = Pr(Z \in [-1.96, 1.96])$;
    e. None of the above choices (a)-(d).

14. A random variable X is normally distributed, $X \sim N(\mu, \sigma^2)$. A sample of size $n = 4$ is drawn and yields: $\sum_{i=1}^4 X_i = 40$, and $\sum_{i=1}^4 (X_i - \bar{X})^2 = 48$, where $\bar{X} = \frac{1}{4}\sum_{i=1}^4 X_i$. We would like to test the following hypothesis: $H_0 : \mu = 5$ and $H_1 : \mu \neq 5$. We are given that $t_{3,0.99} = 4.541$, $t_{3,0.975} = 3.182$, $t_{3,0.95} = 2.353$, and $t_{3,0.90} = 1.638$, where $t_{d,\alpha}$ is the t-statistic with degree of freedom d and cumulative probability of α. Which of the following is true?

    a. We cannot reject the null at 90% confidence level;
    b. We can reject the null at 90% confidence level but not at 95% confidence level;
    c. We can reject the null at 95% confidence level but not at 97.5% confidence level;
    d. We can reject the null at 97.5% confidence level but not at 99% confidence level;
    e. We can reject the null at 99% confidence level.

15. Consider a normal distribution of the form $N(\mu, 4)$. You want to test the hypothesis that $H_0 : \mu = 5$ against the alternative $H_1 : \mu > 5$. A random sample $X_1, X_2, ..., X_n$ was obtained.

    Suppose you would like to test the above one-sided hypothesis using a 5% level of significance. How would you test it?

    a. Set up a Z test where $z = \frac{\bar{X}-5}{\sqrt{4n}}$ and reject the null if $z > 1.96$;
    b. Set up a Z test where $z = \frac{\bar{X}-5}{\sqrt{4n}}$ and reject the null if $z > 1.645$;

c. Set up a Z test where $z = \frac{\bar{X}-5}{\sqrt{4/n}}$ and reject the null if $z > 1.96$;

d. Set up a Z test where $z = \frac{\bar{X}-5}{\sqrt{4/n}}$ and reject the null if $z > 1.645$;

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

16. Korosensei is a junior high school teacher. A few of his students performed weakly during the midterm exam, and Korosensei decided to provide additional but mandatory tutor sessions to those who scored below 60 in the midterm. After the final exam, Korosensei collects the scores and calculate d_i as the score difference between the final exam and the midterm exam for each student i. Let's express $\overline{d_{[x:y]}}$ as the average d_i among students who scored between x and y in the midterm exam. Let's also make an assumption that students have similar intrinsic ability (not statistically different) if their midterm exam score difference is less or equal to 10 points. Before conducting any hypothesis testing, among the following statistics, which one would be more informative about the effect of Korosensei's additional tutor session?

a. $\overline{d_{[0:100]}}$;
b. $\overline{d_{[0:60)}}$;
c. $\overline{d_{[60:100]}} - \overline{d_{[0:60)}}$;
d. $\overline{d_{[95:100]}} - \overline{d_{[55:60)}}$;
e. $\overline{d_{[60:65]}} - \overline{d_{[55:60)}}$.

17. Bob is testing the following empirical model using a sample of 1,000 observations:
$Y_i = \alpha + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \epsilon_i$.
He found that the F-statistic of the joint test for the entire model is 500. He now considers expressing the model's strength using (unadjusted) R-square. Please calculate the R-square value for him.

a. 0.334;
b. 0.501;
c. 0.601;
d. 0.668;
e. None of the above choices (a)-(d).

18. In the two-variable model:
$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i, \qquad i = 1, 2, 3, ..., 11$
Suppose that $X_1'X_1 = 2$, $X_2'X_2 = 2$, $X_1'X_2 = 1$, $X_1'Y = 1$, $X_2'Y = 1$, and $Y'Y = 4/3$, where X_1, X_2, and Y are the column vectors with typical elements X_{1i}, X_{2i}, and Y_i respectively. Furthermore, X_1', X_2', and Y' are the transpose of X_1, X_2, and Y

respectively. Assume $\epsilon_i \sim IID\ N(0, \sigma_\epsilon^2)$.
Now suppose you would like to make out-of-sample predictions about the dependent variable for one hypothetical observation (Y_j, X_{1j}, X_{2j}) for some $j > 11$. We can observe that $X_{1j} = 5$ and $X_{2j} = -2$. Please estimate the expected value and variance of Y_j using the formula: $\widehat{Y}_j = \widehat{\beta}_1 X_{1j} + \widehat{\beta}_2 X_{2j}$.
Note: please estimate the variance-covariance matrix of $\hat{\beta}$ as $s^2(X'X)^{-1}$ where s^2 is the residual sum of squares divided by the degree of freedom and $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$.

a. $E(\widehat{Y}_j)$ =2, $Var(\widehat{Y}_j)$ =3.42;
b. $E(\widehat{Y}_j)$ =1, $Var(\widehat{Y}_j)$ =2.00;
c. $E(\widehat{Y}_j)$ =2, $Var(\widehat{Y}_j)$ =2.54;
d. $E(\widehat{Y}_j)$ =1, $Var(\widehat{Y}_j)$ =1.92;
e. None of the above choices (a)-(d).

19. Suppose that a data generating process is as the following:
$Y_i = a + bX_i + \epsilon_i$
However, we have some measurement error in X_i. The collectable $\tilde{X}_i = X_i + u_i$. We know that $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ and $u_i \sim N(0, \sigma_u^2)$. u_i is also uncorrelated to X_i, Y_i, and ϵ_i. If we run the regression of Y on $\tilde{X}_i$ along with an intercept, what is the relationship between the estimated $\widehat{\beta}$ and the true b?

a. $plim(\widehat{\beta}) = b$;
b. $plim(\widehat{\beta}) = \frac{var(X_i)}{var(X_i)+\sigma_\epsilon^2} b$;
c. $plim(\widehat{\beta}) = \frac{var(X_i)}{var(X_i)+\sigma_u^2} b$;
d. $plim(\widehat{\beta}) = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2+\sigma_u^2} b$;
e. None of the above choices (a)-(d).

20. Consider the following linear model:
$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_i, \quad E(u_i|X_i, Z_i) = 0, u \sim N(0, \sigma_u^2)$.
Assume that you want to conduct an alternative regression model where the dependent variable is the difference between Y_i and Z_i:
$(Y_i - Z_i) = \alpha^* + \beta_1^* X_i + \beta_2^* Z_i + u_i^*$.
What is the relationship between the OLS estimate $\widehat{\beta_1^*}$ and β_1?

a. $E(\widehat{\beta_1^*}) = \beta_1$;
b. $E(\widehat{\beta_1^*}) < \beta_1$;
c. $E(\widehat{\beta_1^*}) > \beta_1$;
d. $E(|\widehat{\beta_1^*}|) < |\beta_1|$;
e. $E(\widehat{\beta_1^*})$ could be larger or smaller than β_1.