※ 115 年度台大財金 ※

1. Ash Ketchum collected 100 samples of Pikachus and measured their weights. Few outliers are found, and which of following way(s) can be potential measures to reduce effect of outliers?

(A) Winsorize outliers.

(B) Remove outliers.

(C) Take natural logarithm of outliers.

(D) Take exponential of outliers.

2. Ash Ketchum studies the power of four different PocketMons, and four independent groups are observed, each with the same sample size. The sample means and pooled mean square error from a one-way ANOVA are given below:

Group	Name of PockMon	Sample Mean
1	Purin	70
2	Pikuchu	75
3	Zenigame	78
4	Fushigidane	82

Each group has sample size $n = 10$. The pooled mean square error is $MSE = 36$. The researcher is interested in testing whether the average power of Pikuchu, Zenigame, and Fushigidane differs from the Purin. Ash Ketchum decides to use contrast test for this hypothesis testing, where $H_0 : L = 0$; with the contrast $L = \mu_1 - (\mu_2 + \mu_3 + \mu_4)/3$. Which of following one(s) is correct?

(A) The contrast coefficients must sum to zero.

(B) The test statistic follows a t-distribution with 39 degrees of freedom.

(C) The t-value is about -3.80.

(D) A significant contrast implies that the overall ANOVA F-test is significant.

3. Let $X \sim \text{Bernoulli}(p)$, where $0 < p < 1$. Consider the event $A = \{(X - p)^2 > 4p(1 - p)\}$, Which of the following value(s) can the probability $P(A)$ take for some value of $p \in (0, 1)$?

(A) 0

(B) p

(C) $1 - p$

(D) 0.999

4. Let X be a continuous random variable with probability density function: $f_X(x) = 2x, 0 \le x \le 1$. Define a new random variable $Y = X^2$. Which of the following statement(s) are correct?

(A) Y takes values in (0,1)

(B) The cumulative distribution of Y is $F_Y(y) = y, \forall 0 \le y \le 1$.

(C) The probability density function of Y is $f_Y(y) = 1, \forall 0 \le y \le 1$

(D) $E(Y) = 1/3$.

5. Which of following estimator(s) could be biased to population mean given i.i.d. random variables $\{X_1, \ldots, X_n\}$ with finite population mean μ?

(A) X_1

(B) $\check{X} = (X_1 X_3 X_5)^{\frac{1}{3}}$

(C) $\bar{X} = \sum_1^n X_i / n$

(D) $\tilde{X} = 2 \sum_1^n i X_i / n(n + 1)$

6. Let $\{X_1, \ldots, X_n\}$ be i.i.d. random variables such that $X_i \sim N(\mu, \sigma^2)$, where σ^2 is known. Consider the hypothesis test $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$. Denote $z_{1-\alpha}$ is the $(1-\alpha)$ quantile of the standard normal distribution. Which of the following is the power function of this test?

   (A) $\pi(\mu) = 1 - \Phi(z_{1-\alpha} - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma})$

   (B) $\pi(\mu) = \Phi(z_{1-\alpha} - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma})$

   (C) $\pi(\mu) = 1 - \Phi(z_{1-\alpha} + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma})$

   (D) $\pi(\mu) = 1 - \Phi(z_{1-\alpha} + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma})$

7. Let $\{x_1, \ldots, x_n\} \sim i.i.d. N(\mu_x, \sigma_x^2)$. Denote sample average $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\tilde{\theta} = (x_1 + x_2)/2$, and $\hat{\theta} \equiv E(\tilde{\theta}|\bar{x})$. Then

   (A) $\hat{\theta} = 1$

   (B) $\hat{\theta}$ is not a BLUE.

   (C) $\hat{\theta}$ is an unbiased estimator of μ_x

   (D) $\hat{\theta}$ is a consistent estimator of μ_x

8. Following previous question, please find $Var(\hat{\theta})$.

   (A) 0

   (B) $\sigma_x^2/2$

   (C) σ_x^2/n

   (D) None of above

9. Let $\{x_1, \ldots, x_n\} \sim i.i.d. N(\mu_x, \sigma_x^2)$. Both μ_x and σ_x^2 are unknown. Denote sample average $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, and $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Then

   (A) $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_x^2} \sim \chi_{n-1}^2$

   (B) The chi-square distribution is used only when n is large enough.

   (C) The confidence interval for σ^2 is symmetric about s_x^2.

   (D) Increasing the sample size n narrows the confidence interval.

10. Specify two regressions estimated on the same dataset; eq(1): $y = a_0 + a_1 DX + a_2(1 - D)X + e$; eq(2): $y = b_0 + b_1 X + u$. X is non-stochastic, and $E(X) = 0$. D is a dummy with values zero or one. e and u are error terms, which satisfy the classical linear regression assumptions. Which of following statement(s) is correct?
   (A) Eq(1) cannot be estimated because perfect collinearity.
   (B) Mean Squared Error (MSE) of eq(1) is smaller than MSE of eq(2).
   (C) b_1 must be ranged between a_1 and a_2.
   (D) OLS estimator of a_0 is equal to OLS estimator of b_0.

11. Consider the structural equation $y = \alpha + \beta x + u$, where X is endogenous that $Cov(x, u) \neq 0$. Suppose we have an instrumental variable z, and estimate $\hat{\beta}_{IV}$ using instrumental variable regression model, where $\hat{\beta}_{IV} = Cov(z, y)/Cov(z, x)$. Which of the following statements are correct?
   (A) For z to be a valid instrument, it must satisfy both relevance ($Cov(z, x) \neq 0$) and the exclusion restriction ($Cov(z, u) = 0$).
   (B) If the instrument is weak, estimator $\hat{\beta}_{IV}$ is approximately unbiased in small samples.
   (C) With a weak instrument, estimator $\hat{\beta}_{IV}$ can have larger bias than the OLS estimator.
   (D) Asymptotic bias of $\hat{\beta}_{IV}$ is $Cov(z, u)/Var(x)$.

12. Consider the linear regression model $y = X\beta + u$, where $E(u|X) = 0$, but the variance of the error term may be heteroskedastic. Which of the following statements are correct?
   (A) The Breusch-Pagan test is based on regressing the squared OLS residuals on the original regressors (or a subset of them).
   (B) The White test allows for heteroskedasticity of unknown functional form and includes cross-product terms of the regressors.

(C) Under the null hypothesis of homoskedasticity, the Breusch-Pagan and White test statistics are asymptotically chi-square distributed.

(D) If heteroskedasticity is present, the OLS estimator becomes biased and inconsistent.

13. Given a function

$$f(x) = \frac{\exp(x-\mu)}{[1+\exp(x-\mu)]^2}, x \in \mathbb{R},$$

please determine which of following statement(s) is correct.

(A) $f(x)$ can be a probability density function.

(B) $E(X) = \mu$.

(C) $E(X-\mu)^3 > 0$.

(D) Denote $odds = \frac{P(X>b)}{[1-P(X>b)]}$ for a given number b, then $\ln(odds) = \mu - b$.

Neneneko recently won the lottery. He plans to spend the money to buy stocks. Neneneko is now studying the property of a certain stock X. Below presents the data of monthly returns of X in the past 8 months (r_x) and the corresponding market returns (r_m) as well as the risk-free rates (r_f). Please answer questions 14 to 16 using the information.

Month_t	r_x (%)	r_m (%)	r_f (%)
1	6	3	1
2	5	3	1
3	1	-2	1
4	0	-2	1
5	-3	1	1
6	-2	1	1
7	8	5	1
8	13	5	1

14. Neneneko is interested in the systematic risk of stock X and would like to estimate the following regression:

$$r_{xt} - r_{ft} = b_0 + b_1(r_{mt} - r_{ft}) + e_t$$

Assume that $e_t \overset{i.i.d.}{\sim} (0, \sigma^2)$ and that $E[e_t | r_{mt}, r_{ft}] = 0$. Which of the followings are correct?

(A) The ordinary least square estimate $\hat{b}_0^{OLS} = 1.533$.

(B) The ordinary least square estimate $\hat{b}_1^{OLS} = 1.672$.

(C) The maximum likelihood estimate $\hat{b}_0^{ML} = 0.014$.

(D) The standard error of $\hat{b}_1^{OLS}$ is 0.513.

15. Following the regression specification in the previous question, Neneneko is interested in understanding the properties of the ordinary least square estimator under different hypotheses for the error term.

A1: $e_t \overset{i.i.d.}{\sim} (0, \sigma^2)$;

A2: $e_t \overset{i.i.d.}{\sim} \Phi(0, \sigma^2)$, where Φ stands for normal distribution;

A3: $e_t \sim (0, \sigma_t^2)$, e_t is independent of e_i, $\forall i \neq t$

With the assumption of $E[e_t | r_{mt}, r_{ft}] = 0$, which of the followings are correct?

(A) The ordinary least square estimator is BUE under A1.

(B) The ordinary least square estimator is $BLUE$ under A2.

(C) Under A_3, the ordinary least square estimator is biased.

(D) Under A_3, given the full sample variance-covariance matrix of the error terms, the generalized least square estimator is $BLUE$.

16. Ideally, Neneneko should perform the following regression:

$$r_{xt} - r_{ft} = b_0 + b_1(r_{mt} - r_{ft}) + e_t$$

Suppose that Neneneko is careless and mess up his code. He performs the following three regressions instead:

$$r_{xt} = a_0 + a_1(r_{mt} - r_{ft}) + \epsilon_t \tag{1}$$

$$r_{xt} - r_{ft} = c_0 + c_1 r_{mt} + \varepsilon_t \tag{2}$$

$$r_{xt} = d_0 + d_1 r_{mt} + \eta_t \tag{3}$$

Which of the followings are correct?

(A) $\hat{c}_1^{OLS} = \hat{b}_1^{OLS}$

(B) $\hat{d}_0^{OLS} = \hat{b}_0^{OLS} + 1\%$

(C) $\hat{a}_0^{OLS} \neq \hat{b}_0^{OLS} + 1\%$

(D) $\hat{d}_0^{OLS} = \hat{c}_0^{OLS} + 1\%$

17. Suppose that a data generating process is as the following:

$$Y_i = a + bX_i^* + \epsilon_i$$

However, X_i^* cannot be directly observed. There are two observable proxies for X_i^*:

$$X_{i_1} = X_i^* + u_{i_1}$$
$$X_{i_2} = X_i^* + u_{i_2}$$

We know that $\epsilon_i \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$, $X_i^* \overset{i.i.d.}{\sim} (\mu_X, 9)$, $u_{i_1} \overset{i.i.d.}{\sim} (0, 1)$, and $u_{i_2} \sim N(0, 3)$. u_{i_1} and u_{i_2} are also uncorrelated to X_i^*, Y_i, ϵ_i, and each other. We now construct an additional regressor:

$$X_{i_3} = \frac{1}{2}X_{i_1} + \frac{1}{2}X_{i_2}$$

For $n = 1, 2,\ \text{or}\ 3$, denote $\hat{\beta}_n$ as the OLS coefficient when we regress Y on X_ϵ. Which of the followings are correct?

(A) $plim(\hat{\beta}_1) < b$

(B) $plim(\hat{\beta}_2) = 0.75b$

(C) $plim(\hat{\beta}_3) = b$

(D) $plim(\hat{\beta}_3) = plim(\hat{\beta}_1)$

18. The number of red balls and blue balls in a bag is unknown, but it is known that the proportion, p, of red is either $\frac{3}{7}, \frac{3}{8},$ or $\frac{1}{3}$. A sample of size 5, drawn with replacement, yields the sequence red, blue, blue, red, and blue. The maximum likelihood estimate for p is:

(A) $\frac{3}{7}$

(B) $\frac{2}{5}$

(C) $\frac{3}{8}$

(D) $\frac{1}{3}$

19. In the two-variable model: $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$, $i = 1, 2, 3, \ldots, 22$ Suppose that $X_1'X_1 = 4, X_2'X_2 = 10, X_1'X_2 = 6, X_1'Y = 4, X_2'Y = 6$, and $Y'Y = 32$, 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 $\epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$. 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 $j > 22$. We can observe that $X_{1j} = 6$ and $X_{2j} = 14$. Please estimate the expected value and variance of Y_j using the formula: $\hat{Y}_j = \hat{\beta}_1 X_{1j} + \hat{\beta}_2 X_{2j}$. Which of the followings are correct?

(A) $E(\hat{\beta}_1) = 0$

(B) $E(\hat{\beta}_2) = 0$

(C) $E(\hat{Y}_j) = 14$

(D) $Var(\hat{Y}_j) = 34$

20. Let the random variable X have probability function

$$f(x)=\begin{cases} p, & X=1,-1 \\ 1-2p, & X=0 \\ 0, & \text{elsewhere} \end{cases}$$

where $0<p<\frac{1}{2}$. Which of the followings are correct?

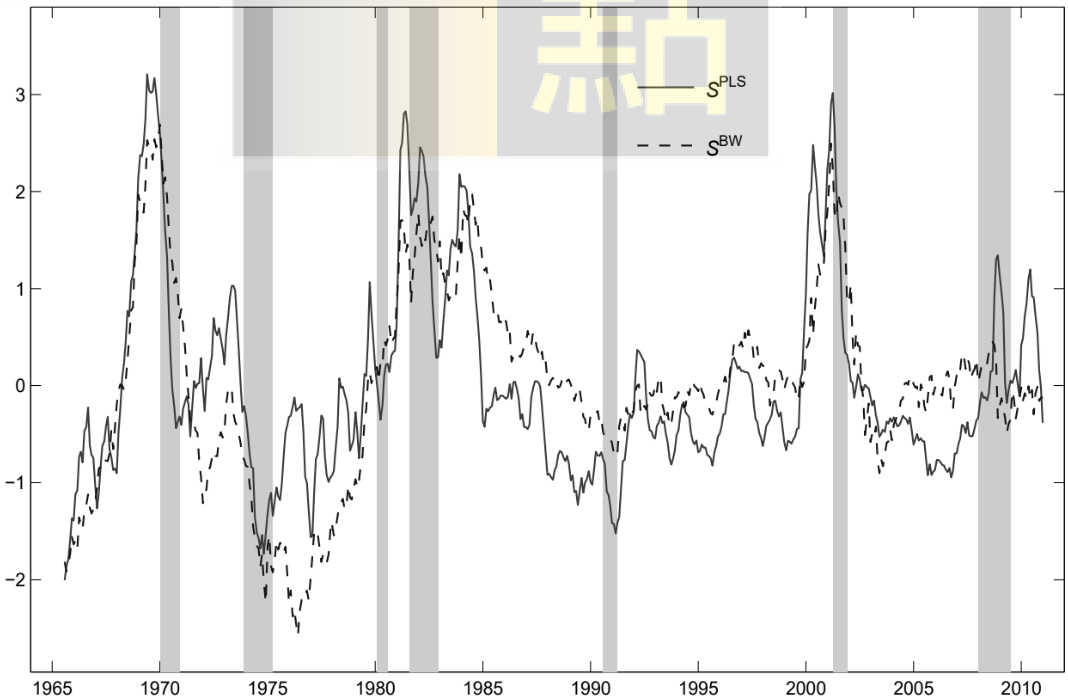
(A) $mean(X)=p$

(B) $variance(X)=2p$

(C) $skewneo(X)=3p$

(D) If $p>\frac{1}{6}$, the distribution of X would have fatter tails than a standard normal distribution.

21. Huang, Jiang, Tu, and Zhou (2015, RFS) constructed a market sentiment index, S^{PLS}, and compared it against the seminal Baker and Wurgler (2006, JF) sentiment index, S^{BW}. They showed that the monthly variable S^{PLS} can negatively predict the market returns in the following month. Which of the followings are correct?



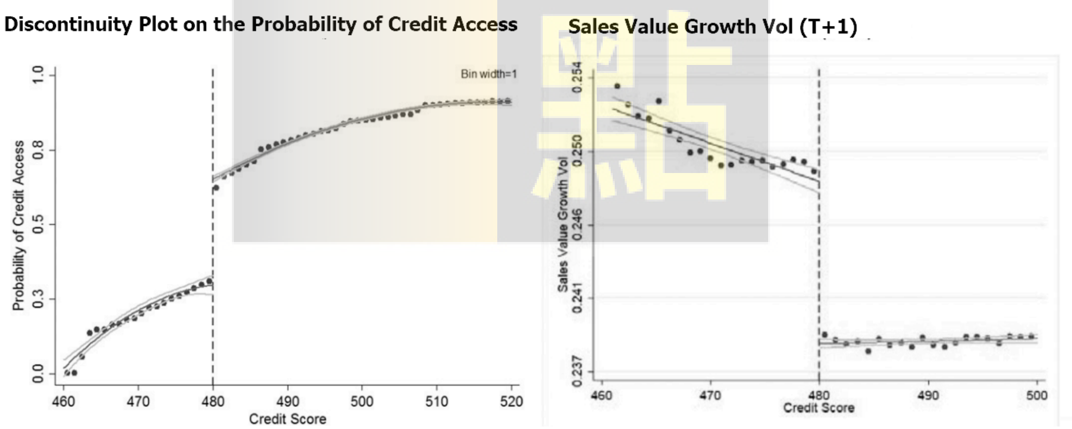
(A) Suppose the $AR(1)$ coefficient of S^{PLS} is 0.98. Since it is smaller than 1, we do not have to worry about the unit root problem.

(B) If a predictor, such as S^{PLS} or S^{BW}, has a unit root, regressing future market returns on it would definitely produce a spurious regression.

(C) If a time series has a unit root, taking a first difference always removes the unit root.

(D) If a time series has a unit root, an exogenous shock to the series in a given time period could have a permanent effect on all the future realizations.

22. Chen, Huang, Lin, and Sheng (2022, MS) studied how access to finance affects the business of small and medium enterprises (SBEs) using the data from Alibaba. The left figure below shows the probability of credit access for SBEs of different credit scores. The right figure below shows the volatility of the SBEs' sales value growth in the following quarter. Which of the followings are correct?



(A) The researchers show that access to finance may causally reduce sales growth volatility.

(B) The discontinuity on the left figure undermines the reliability of the researchers' intention.