

編號: G 202 系所: 民航研究所乙組

科目: 統計學

本試題是否可以使用計算機:  可使用,  不可使用 (請命題老師勾選)

## 1. Multiple Choices: (35%, 5% for each part)

(1) If A and B are exclusive events with  $P(A) > 0$  and  $P(B) > 0$ , then

- (a) A and B are independent events                      (b) A and B are dependent events  
 (c)  $P(A|B) > 0$     (d)  $P(A) + P(B) = 1$ .  
 (e) none of the above

(2) Let  $S_A$  ( $CV_A$ ) and  $S_B$  ( $CV_B$ ) denote the standard deviation (coefficient of variation) of the following two data sets,A:  $x_1, x_2, x_3, x_4, x_5$ B:  $10x_1+5, 10x_2+5, 10x_3+5, 10x_4+5, 10x_5+5$ 

then

- (a)  $S_A = S_B$                       (b)  $10S_A = S_B$                       (c)  $10S_A - 5 = S_B$   
 (d)  $10CV_A = CV_B$                       (e)  $10CV_A - 5 = CV_B$

(3) In hypothesis testing, only type I error rate is controlled by setting the significant level

( $\alpha$ ), because

- (a) the other error is not important.  
 (b) the other error never occurs.  
 (c) we can not control both error rates simultaneously.  
 (d) controlling type I error rate is equivalent to controlling the other error rate.  
 (e) none of the above

(4) Central limit theorem implies that as sample size increases,

- (a) the distribution of observations becomes normal.  
 (b) the distribution of sample mean is normal.  
 (c) The distribution of the population becomes normal.  
 (d) the distribution of sample mean is approximately normal.  
 (e) none of the above

(5) Sampling error occurs because

- (a) the investigator chooses the wrong sample.  
 (b) of the operation of chance.  
 (c) of a calculation error in obtaining the sample mean.  
 (d) the measuring device is flawed.  
 (e) all of the above

(6) The box in the center of boxplot marks

- (a) The full range covered by the data.  
 (b) The range covered by the middle three-quarters of the data.  
 (c) The range covered by the middle half of the data.  
 (d) The span one standard deviation on each side of the mean.  
 (e) The span two standard deviation on each side of the mean.

(7) Which statement is true for Pearson correlation coefficient?

- (a) If the Pearson correlation coefficient is zero then the two quantitative variables may still be correlated.  
 (b) If the Pearson correlation coefficient is zero then the two quantitative variables are independent.  
 (c) The Pearson correlation coefficient is not affected by outliers.  
 (d) If there is a strong curvilinear relationship between two quantitative variables, Pearson correlation coefficient will overestimate this relationship  
 (e) none of the above

(背面仍有題目, 請繼續作答)

本試題是否可以使用計算機： 可使用， 不可使用（請命題老師勾選）

2. Let  $\theta_1(\theta_2)$ ,  $Q_{11}(Q_{21})$ ,  $Q_{13}(Q_{23})$  and  $\mu_1(\mu_2)$  denote the median, first quartile, third quartile and mean of two data sets. If  $\theta_1 > \theta_2$ ,  $Q_{11} > Q_{21}$ , and  $Q_{13} > Q_{23}$ , then  
 (a)  $\mu_1 > \mu_2$  (b)  $\mu_1 < \mu_2$  (c) uncertain about the relationship between  $\mu_1$  and  $\mu_2$ .  
 Please justify your answers. (15%)
3. A telephone company employs five information operators who receive requests for information independently of one another, each according a Poisson process with rate  $\lambda=3$  per minute.  
 (1) What is the probability that during a given two-minute period, the first operator receives no requests? (5%)  
 (2) What is the probability that during a given two-minute period, exactly three of the five operators receive no requests? (10%)
4. (1) The time required to complete a college achievement test was found to be normally distributed with a mean of 110 minutes and a standard deviation of 20 minutes. What should the test be terminated to allow just enough time for 90% of the students to complete the test? (5%)  
 (2) If the time required to complete a college achievement test was not normally distributed with a mean of 110 minutes and a standard deviation of 20 minutes, can you still be able to find the time to have at least 90% of the students to complete the test? (10%)
5. Two methods of estimating inventory value used at the end of each month, for a six month period, produced the following 12 estimates:

| Method | Month |      |      |       |     |      |
|--------|-------|------|------|-------|-----|------|
|        | Jan.  | Feb. | Mar. | April | May | June |
| A      | 14    | 12   | 16   | 15    | 10  | 11   |
| B      | 18    | 16   | 17   | 19    | 13  | 13   |

Three results from different statistical analyses are given below, respectively.

| Statistical methods                                | test statistic | degree of freedom | p-value |
|--|----------------|-------------------|---------|
| (a) independent $t$ test<br>(for unequal variance) | $T = -2.12$    | 10                | 0.0600  |
| (b) independent $t$ test<br>(for equal variance)   | $T = -2.12$    | 10                | 0.0599  |
| (c) Paired $t$ test                                | $T = -5.81$    | 5                 | 0.0021  |

Equality of Variances

| Variable value | Method   | Num DF | Den DF | F Value | p-value |
|----------------|----------|--------|--------|---------|---------|
|                | Folded F | 5      | 5      | 1.14    | 0.8871  |

- (1) Which analysis is appropriate for comparing these two estimating methods? Why? Justify your answer. (10%)  
 (2) Use the test you have chosen to make conclusion at  $\alpha = 0.03$ . Do you need any assumptions? (10%)

