題號： 296
科目： 統計學(H)
節次： 2

國立臺灣大學 113 學年度碩士班招生考試試題

題號：296
共 5 頁之第 1 頁

Please choose the most suitable answer among all alternatives for each question and clearly mark it on the answer sheet. Each question is worth 5 points.

1. Which descriptive statistic is least affected by extremely small or extremely large data values?
   A) Mean
   B) Median
   C) Range
   D) Mean and Median
   E) Median and Range

2. What is the relationship among the mean, median, and mode in a right-skewed (or positively skewed) distribution?
   A) The mean is typically the smallest value.
   B) The median is typically the smallest value.
   C) The mode is typically the smallest value.
   D) They are all equal.
   E) Cannot conclude without further information.

3. Suppose there are two distributions, A and B. A is a normal distribution with a mean of 4 and a standard deviation of 3. B is a normal distribution with a mean of 5 and a standard deviation of 2. Which of the following statements is false?
   A) The width of distribution A is wider than that of distribution B.
   B) The mean, median, and mode are equal to 4 for distribution A and 5 for distribution B.
   C) The frequency of values peaks at the mean for both distributions.
   D) 100% of the values fall between ±3 standard deviations for both distributions A and B.
   E) All of the statements above are correct.

4. The wait time at a specific stoplight follows a uniform distribution ranging from zero to five minutes. What is the probability of having to wait more than 120 seconds at the stoplight?
   A) 0.2
   B) 0.4
   C) 0.6
   D) 0.8
   E) 1.0

5. Why is the central limit theorem important in statistics?
   A) Because for a large sample size n, it asserts the population is approximately normal.
   B) Because for any population, it asserts the sampling distribution of the sample mean is approximately normal, regardless of the shape of the population.
   C) Because for a large sample size n, it asserts the sampling distribution of the sample mean is approximately normal, regardless of the shape of the population.
   D) Because for any sample size n, it asserts the sampling distribution of the sample mean is approximately normal.
   E) None of the above.

題號： 296
科目： 統計學(H)
節次： 2

國立臺灣大學 113 學年度碩士班招生考試試題

題號：296
共 5 頁之第 2 頁

6. A company is recruiting management trainees for entry-level marketing positions. Based on historical data, approximately 20% are expected to remain employed after six months. If the company has recently hired five trainees, what is the approximate probability that exactly three of them will still be employed at the end of six months?
   A) 0.005
   B) 0.008
   C) 0.05
   D) 0.2
   E) 0.5

7. For a given population proportion, which of the following statements is true regarding the width of a confidence interval for the population proportion?
   A) It is narrower for a lower confidence level than for a higher confidence level.
   B) It is wider for a larger sample size than for a smaller sample size.
   C) It is narrower for a larger sample proportion than for a smaller sample proportion.
   D) It is wider for a larger sample proportion than for a smaller sample proportion.
   E) None of the above is correct.

8. The head of a research team claims that he can accurately determine whether a person has a medical background or an engineering background based on their problem-solving approach. When presented with one individual and asked to identify their background (either medical or engineering), the team lead treats this as a hypothesis test with the null hypothesis being that the person has a medical background and the alternative that the person has an engineering background. Which of the following statements illustrates a Type II error?
   A) Identifying the person as having an engineering background when, in fact, the person has a medical background.
   B) Identifying the person as having a medical background when, in fact, the person has a medical background.
   C) Identifying the person as having a medical background when, in fact, the person has an engineering background.
   D) Identifying the person as having an engineering background when, in fact, the person has an engineering background.
   E) None of the above.

9. Which of the following does not influence the standard error of the regression slope?
   A) The spread around the regression line
   B) The spread of x values
   C) The sample size
   D) The critical value
   E) All of these affect the standard error

10. To evaluate the effectiveness of a new exercise program, a random sample of 25 individuals is chosen from a population of adults engaged in the program. Each participant's weight is measured both before and after the program. Assuming that the population of differences in weight before and after the program follows a normal distribution, the mean decrease in weights for the 25 participants is determined to be 2.5 pounds, with a standard deviation of differences of 3.2 pounds. What conclusion can be drawn about the effectiveness of the exercise program based on the test at the 0.05 level of significance?
    A) The exercise program is not effective.
    B) The exercise program is effective.
    C) The sample size is relatively small to make a conclusion.
    D) A conclusion cannot be drawn without calculating the p-value.
    E) None of the above.

接 次 頁

題號： 296
科目： 統計學(H)
節次： 2

國立臺灣大學 113 學年度碩士班招生考試試題

題號：296
共 5 頁之第 3 頁

11. Which of the following is a violation of one of the major assumptions of the simple regression model?
    A) The error terms are independent of each other.
    B) A histogram of the residuals forms a bell-shaped, symmetrical curve.
    C) A plot of the residual versus x forms a horizontal band pattern.
    D) As the value of x increases, the value of the error term also increases.
    E) The error terms show no pattern.

12. Which, if any, of the following statements about the chi-square test of independence is false?
    A) If r_i is the row total for row i and c_j is the column total for column j, then the estimated expected cell frequency corresponding to row i and column j equals $(r_i)(c_j)/n$.
    B) The test is valid if all of the estimated cell frequencies are at least five.
    C) The chi-square statistic is based on $(r-1)(c-1)$ degrees of freedom, where r and c denote, respectively, the number of rows and columns in the contingency table.
    D) The alternative hypothesis states that the two classifications are statistically independent.
    E) All of the other statements about the chi-square test of independence are true.

13. A marketer aims to test the effectiveness of the ABC filter designed to reduce harmful chemicals in drinking water. The marketer measured harmful chemical levels in 30 households' water before installing the product. Next, the same 30 households installed the ABC filter. After two weeks of continuous use, the harmful chemical levels were measured again. The comparison of harmful chemical levels before versus after using the product is an example of testing the difference between _____.
    A) two means from independent populations.
    B) two population variances from independent populations.
    C) two population proportions.
    D) two population medians.
    E) matched pairs from two dependent populations.

For Questions 14-16, please read the scenario and output below.

To compare the performance characteristics of four brands of space heaters, a researcher bought all the heaters they could find from these brands and tested them. For performance, the researcher measured the amount of time it took to warm up a specific-sized room by 5 degrees from a set temperature. They also coded the characteristics of the space heaters. The researcher first conducted a one-way ANOVA on the space heaters' performance among the four brands. The output is given below.

ANOVA table

Source	SS	df	MS	F	p-value
Treatment	6.000	3	1.9998	18.85	3.46E-05
Error	1.713	16	0.1061		
Total	7.713	19			

Post hoc analysis
Tukey simultaneous comparison t-values ($d.f. = 16$)

見背面

題號： 296
科目： 統計學(H)
節次： 2

國立臺灣大學 113 學年度碩士班招生考試試題

題號：296

共 5 頁之第 4 頁

		Brand 2	Brand 3	Brand 4	Brand 1
		1.40	2.28	2.58	2.95
Brand 2	1.40				
Brand 3	2.28	4.27			
Brand 4	2.58	5.38	1.35		
Brand 1	2.95	7.09	3.07	1.63	
Critical values for experimentwise error rate:					
		0.05	2.91		
		0.01	3.76		

14. Based on the output, at a significance level of .05, we would conclude that _____.
    A) brand 1 differs from brand 3, and brand 3 differs from brand 2, while the rest of the space heater pairs do not differ from each other in terms of their performance.
    B) brand 1 differs from brand 3, and brand 2 differs from brands 1, 3, and 4, while the rest of the space heater pairs do not differ from each other in terms of their performance.
    C) only brand 2 differs from the other three brands, while the rest of the space heater pairs do not differ from each other in terms of their performance.
    D) all four brands of space heaters differ from each other in terms of their performance.
    E) none of the four brands of space heaters differ from each other in terms of their performance.

15. How many heaters did the researcher test in this research?
    A) 17
    B) 19
    C) 20
    D) 21
    E) 38

16. To further understand the space heaters' performance, in addition to the brand, the researcher would like to add voltage as a predictor. Hence, they decided to conduct a multiple regression model. How many independent variables should they create for the brand in this multiple regression model?
    A) One
    B) Two
    C) Three
    D) Four
    E) Five

For Questions 17-20, please read the scenario below.
A consultant hired by a chain restaurant focusing on eco-friendly practices has developed a multiple regression model for the chain restaurants' performances for their branches. In this multiple regression model, the dependent variable, performances, is the branch sales in thousands of dollars. For example, a data entry of 25 for the dependent variable indicates a sales of $25,000. The dataset the consultant was given also included branch and market factors, such as the adaptation of eco-friendly practices, branch size, employee tenure, and population density of surrounding areas.

接 次 頁

題號： 296
國立臺灣大學 113 學年度碩士班招生考試試題
題號：296
科目： 統計學(H)
節次： 2
共 5 頁之第 5 頁

17. The consultant first tested a multiple regression model with the eco-friendly variable as one of the indicators (X1 = 0 if not adopted and X1 = 1 if adopted). The output of this multiple regression model shows that the coefficient for this variable (X1) is −1.6. The t test showed that X1 was significant at $\alpha = .05$. This result implies that for companies adopting and not adopting sustainable practices, _____.
   A) on average, the branches that adopted eco-friendly practices made $1,600 less than those that did not adopt eco-friendly practices.
   B) on average, the branches that did not adopt eco-friendly practices made $1,600 less than those that adopted eco-friendly practices.
   C) on average, the branches that adopted eco-friendly practices made 1.6 times less than those that did not adopt eco-friendly practices.
   D) on average, the branches that did not adopt eco-friendly practices made 1.6 times less than those that adopted eco-friendly practices.
   E) on average, the performances were not affected by whether the branches adopted eco-friendly practices or not.

18. The result on eco-friendly practice motivates the consultant to dig deeper into the dataset. While exploring, the consultant accidentally adds a very insignificant independent variable (an independent variable that has a very weak relationship with the dependent variable) to a multiple regression equation. As a result of this change, the value of the explained variation (SSR) will _____, the value of the multiple coefficient of determination (R^2) will _____, and the calculated value of the F statistic will most likely _____.
   A) decrease, increase, decrease
   B) increase, decrease, decrease
   C) increase, increase, increase
   D) increase, increase, decrease
   E) decrease, decrease, decrease

19. After a couple of days, the consultant developed this regression model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \varepsilon$. If the consultant wish to test the significance of higher-order terms (x_1^2 and x_2^2), which test would they use?
   A) overall F test
   B) partial F test
   C) Durbin-Watson test
   D) t test
   E) Cook's distance measure

20. Finally, the consultant would like to visualize some of the models they developed for the client. The graph of the prediction equation obtained from the model $y = \beta0 + \beta1X1 + \beta2X2 + \varepsilon$ is a(n) _____.
   A) line
   B) parabola
   C) exponential curve
   D) ellipse
   E) plane

試題隨卷繳回