（請於答案卷作答，勿於試題紙上作答）

一．單選題（20 題，每題 4 分，共 80 分）

1. A tabular summary of a set of data showing classes of the data and the fraction of the items belonging to each class is called:
   a. the class width
   b. a relative frequency distribution
   c. a cumulative relative frequency distribution
   d. none of the above

2. Regarding chi-square test:
   a. no cell in the table should have an expected count less than 1
   b. no more than 20% of the cells should have an expected count less than 5
   c. it is not symmetric
   d. all of the above

3. Two events X and Y are independent if:
   a. $P(Y|X)=P(X)$
   b. $P(Y|X)=P(Y)$
   c. both a and b are satisfied
   d. none of the above

4. If a penny is tossed six times and comes up heads all six times, the probability of heads on the seventh trial is:
   a. less than the probability of trials
   b. 1/64
   c. 0.5
   d. larger than the probability of trails

5. The probability distribution for a discrete random variable which is used to compute the probability of x successes in n trials is known as the:
   a. normal probability distribution
   b. standard normal distribution
   c. binomial probability distribution
   d. none of above

6. Which of the following is not a characteristic of a binomial experiment?
   a. at least 2 outcomes are possible
   b. the experiment consists of a sequence of n identical trials
   c. probabilities remain the same as successive trials are made
   d. the trials are independent of each other

7. In simple linear regression, when explanatory variable is closer to its mean:
   a. the standard deviation of the population decreases
   b. the population mean increases
   c. the confidence interval of the Y decreases
   d. the standard error of the mean increases

8. The difference between the point estimate, such as the sample mean $\overline{X}$, and the value of the population parameter it estimates, such as the population mean μ, is known as the:
   a. confidence level
   b. sampling error
   c. parameter estimate
   d. interval estimate

9. Whenever the population standard deviation is unknown and the population has a normal or near-normal distribution, which distribution is used in developing an interval estimation?
   a. t distribution
   b. chi-square distribution

試題隨卷繳回

背面有題

   c. beta distribution
   d. standard distribution

10. Method of model selection of multiple regression includes:
   a. forward selection
   b. backward elimination
   c. stepwise selection
   d. all of the above

11. When the null hypothesis has been true, but the sample information has resulted in the rejection of the null, a _____ has been made.
   a. level of significance
   b. Type II error
   c. critical value
   d. Type I error

12. For setting the decision rule in a small-sample case, if it is reasonable to assume that the population is normal, we use:
   a. the Z distribution
   b. the t distribution with n-1 degrees of freedom
   c. the t distribution with n+1 degrees of freedom
   d. none of the above

13. An important application of the chi-square distribution is:
   a. making inferences about a single population variance
   b. testing for goodness of fit
   c. testing for the independence of two variables
   d. all of the above

14. Symptom of collinearity in multiple regression model includes:
   a. two or more of the explanatory variables are correlated
   b. instability of the estimated coefficients
   c. increase the standard errors
   d. all of the above

15. The coefficient of determinations is:
   a. The square root of the correlation coefficient
   b. Usually less than zero
   c. the correlation coefficient squared
   d. none of the above

16. The probability distribution for a discrete random variable which is used to compute the probability of x occurrences of an event over a specified interval is known as:
   a. the Poisson probability distribution
   b. the standard normal distribution
   c. a discrete random variable
   d. a linear function

17. Which of the following is not a limitation of coefficient of correlation?
   a. it quantifies only the strength of the linear relationship between two variables
   b. it will be affected by data outliers
   c. a high correlation between two variables implies a cause-and-effect relationship
   d. none of the above

18. The residual refers to:
   a. $\bar{y}_i - \hat{y}_i$
   b. $y_i - \hat{y}_i$
   c. $\hat{y}_i - \bar{y}_i$
   d. None of the above

19. The characteristics of adjusted R^2 include:
   a. it is an estimator of the population correlation
   b. it can explain the proportion of variability by the regression model
   c. it decreases when the inclusion of a variable improves the model ability to predict Y
   d. none of the above

20. The ANOVA procedure is a statistical approach for determining whether or not:
   a. the means of two or more populations are equal
   b. the means of more than two samples are equal
   c. the means of two samples are equal
   d. none of the above

二. 計算題（兩題，每題 10 分，共 20 分):請寫出計算過程。

1. A study tries to evaluate the association between education years and income. The study collects 100 observations with education years from 6 to 23 years, and the annual income range is between $200,000 and $1,500,000. The median income is $600,000, SE is $50,000. The study conducts Pearson's Correlation test and got Pearson's Correlation coefficient equals 0.7.

   When education year increases one year, how much income will be increased? (10%)

2. An International Company claims that the batteries it produces have useful lives of more than 90 hours, with a known standard deviation (of the population) of 27 hours. A test is undertaken to test the validity of this claim.

   a. The correct set of hypotheses for this test is? (5%)
   b. A sample of 81 batteries had an average useful life of 105 hours. The test statistic has a value of ? (5%)