

系所組別： 統計學系

考試科目： 統計學

考試日期： 0220 · 節次： 3

※ 考生請注意：本試題 可 不可 使用計算機

請勿在本試題紙上作答，否則不予計分

1 Multiple Choice $2\% \times 10 = 20\%$

1. Student ID is
(A) a quantitative variable. (B) neither a quantitative nor a qualitative variable. (C) both of a quantitative and qualitative variable. (D) a qualitative variable.
2. Which of the following measure of dispersion would be influenced most by possible extreme value(s)?
(A) variance (B) standard deviation (C) range (D) interquartile range
3. Ruby would like to estimate the average weekly expense of a NCKU student. She plans to collect 1,000 students as her sample in the following manner: At first she will randomly select 50 classes out of about 300 classes in NCKU, and then she will again randomly select 20 students in each selected class. Which of the following design she will be using?
(A) simple random sampling (B) stratified random sampling (C) cluster sampling (D) systematic sampling
4. Jane would like to estimate the proportion of the residents of Taipei City who are satisfied with the traffic situation. She select 1,068 residents from Taipei City as her sample. Which of the following statement is the most appropriate one for the width of the 95% confidence interval of the population proportion?
(A) less than 6%. (B) less than 3%. (C) greater than 3%. (D) cannot be determined since the data is not provided.
5. A and B are two events. Suppose that $P(A \cap B^c) = 0.3$ and $P(A) = 0.7$. What is the value of $P(A \cap B)$?
(A) 0.4 (B) 0.21 (C) 0.3 (D) 0.12
6. Monica built a linear regression model to describe the relationship between the height and weight in her class, in which the weight was the response variable. For some reason, she fitted the model first in the metric system (meter and kilogram), and then fitted again in the British system (foot and pound). What measure was different between these two models?
(A) R^2 (B) The value of test statistic for examining the significance of height. (C) Regression sum of squares. (D) The value of the test statistic for examining the significance of the intercept.

(背面仍有題目,請繼續作答)

系所組別： 統計學系

考試科目： 統計學

考試日期： 0220，節次： 3

※ 考生請注意：本試題 可 不可 使用計算機

7. Refer to Question.6, which of the following assumptions is not required when Monica what to estimate the slope in this model with the least squares method?
(A) Normality (B) Independent observations (C) Equal variances (D) All of the above are required.
8. Which of the followings is equal to $t_{10}^{0.975}$, the 0.975 upper quantile of t -distribution with 10 degrees of freedom? (A) $\sqrt{F_{1,10}^{0.95}}$ (B) $\sqrt{F_{10,1}^{0.05}}$ (C) $\sqrt{(F_{10,1}^{0.975})^{-1}}$ (D) $\sqrt{F_{1,10}^{0.975}}$
9. Which of the following tests is most sensitive to the assumption of normality in general?
(A) Two sample t -test to compare the population means (B) Test the significance of the slope in simple linear regression (C) Two-sample F -test to compare the population variances (D) One sample t -test to examine the population mean.
10. The power curve provides the probability of (A) correctly accepting the null hypothesis. (B) incorrectly accepting the null hypothesis. (C) correctly rejecting the alternative hypothesis. (D) correctly rejecting the null hypothesis.

2 Fill in the Blanks $4\% \times 15 = 60\%$

1. Jack believes that the number of members of a household in Tainan City follows a Poisson distribution. He collect data from 500 randomly selected households, and the data is summarized as Table.1

Number of members	1	2	3	4	5 and more
Number of households	130	280	55	30	5

Table 1: Number of Household members

For simplification reason, the number of household members is considered as five when it is recorded as five or more. He will use the goodness-of-fit test to confirm his assumption. Base on the data in Table1, the expected count of the household with 2 members under the null hypothesis is 1. The test statistic should be compared to 2 distribution, and the conclusion will be 3 under $\alpha = 0.05$.

系所組別： 統計學系

考試科目： 統計學

考試日期： 0220，節次： 3

※ 考生請注意：本試題 可 不可 使用計算機

2. Ross would like to study the relationship between height and weight in his class. Unfortunately, he has never learnt how to perform the simple linear regression analysis. In addition, he cannot provide the raw data but some descriptive statistics summarized as Table.2

	Height	Weight
Sample mean	169	65
Sample Standard Deviation	6.58	7.86

Table 2: Height and Weight

The sample size is 10, and the sample covariance is 39.44. Please help him to fit a regression equation with Weight as the response variable,

$$\text{Weight} = a + b \times \text{Height}.$$

The estimated slope of this model is 4, and the estimated intercept is 5. The total sum of squares is 6, and the residual sum of squares is 7. The data was recorded in metric system (centimeter and kilogram), if one would like to change the data into British system (foot and pound), then the value of the test statistic of the t-test for testing the significance of Height is 8.

3. Jenny analyzed the impact of two explanatory variables X_1, X_2 on the response Y with a multiple regression model,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon,$$

and obtain the following ANOVA tables

Regression Analysis: Y versus X1, X2

The regression equation is

$$Y = - 18.4 + 2.01 X_1 + 4.74 X_2$$

(背面仍有題目,請繼續作答)

系所組別： 統計學系

考試科目： 統計學

考試日期： 0220，節次： 3

※ 考生請注意：本試題 可 不可 使用計算機

Predictor	Coef	SE Coef	T	P
Constant	-18.37	17.97	-1.02	0.341
X1	2.0102	0.2471	8.13	0.000
X2	4.7378	0.9484	5.00	0.002

S = 12.7096 R-Sq = 92.6% R-Sq(adj) = 90.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	14052.2	7026.1	43.50	0.000
Residual Error	7	1130.7	161.5		
Total	9	15182.9			

Source	DF	Seq SS
X1	1	10021.2
X2	1	4030.9

Source	DF	Seq SS
X2	1	3363.4
X1	1	10688.7

The sample size is 10. Based on the output above, the value of the test statistic for testing the hypothesis

$$H_0 : \beta_1 = 0, \text{ without assuming } \beta_2 = 0$$

is 9, and it should be compared to 10 distribution. In addition, the value of the test statistic for testing

$$H_0 : \beta_1 = 0, \text{ assuming } \beta_2 = 0$$

is 11, and it should be compared to 12 distribution. If she would like to fit a simple regression model,

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon,$$

the value of the coefficient of determination would be 13

系所組別： 統計學系

考試科目： 統計學

考試日期：0220，節次：3

※ 考生請注意：本試題 可 不可 使用計算機

4. Mark examined the GRE scores of students from different colleges. He was also interested in whether a student would have better score if he/she attended the supplementary class for GRE test. The sample means of each combination of college and whether a student attended supplementary school are summarized as Table.3

College	Supplementary School	
	Yes	No
Business	1100	1000
Sciences	1300	1100
Engineering	1200	1200

Table 3: GRE score

The overall sample variance of all 60 students is 11864.41, and the residual sum of squares is 200000. The value of the test statistic for testing the interaction effect is 14, and it should be compared to 15 distribution.

3 Problems 10%= 20%

- (10%) In a hypothesis test procedure, why do we usually would pre-specify a confidence coefficient $(1 - \alpha)$? In addition, both of the p-value and α are related to type-I error, please comment on the difference.
- (10%) Jennifer studied if three different procedures would be different in producing a certain product. She selected 15 workers and assigned each of them a procedure randomly, so that she had five workers for each procedure. Each of the 15 workers used the procedure assigned to produce a product and the production time was recorded. Mary would like to conduct the same study, but she plans to select only five workers to provide the data in the same overall same sample size as Jennifer. In Mary's study, each of the five workers will use each of the three procedures, in a random order, to produce a product and the production time will be recorded. Please comment on their research methods.