

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 Choice 5%×9

1. Which of the following statements is/are true?

i. Under Simple Random Sampling Without Replacement, the sample mean is unbiased for estimating the population mean for any kind of population.

ii. Exponential distribution is right-skewed.

iii. Type-I error is generally more serious than type-II error.

A.i B.ii C.iii D.i,ii E.i,iii F. ii,iii G.i,ii,iii H. None

2. Which of the following statements is/are true?

i. ANOVA (Analysis of Variance) is a commonly used statistical tool to examine if the variances of several populations' are equal.

ii. If the data appears to be seriously skewed, then Chebyshev's theorem cannot be used, or we need to transform the data beforehand.

iii. The variance of the sample mean is a population parameter.

A.i B.ii C.iii D.i,ii E.i,iii F. ii,iii G.i,ii,iii H. None

3. Which of the following statements is/are true?

i. For fitting a regression model $y = ax + b$, the assumption of normality is not required for the least square estimates of a and b .

ii. The sum of the frequencies for all classes will always equal 1 in a contingency table.

iii. z-score does not change with the change of the measurement unit.

A.i B.ii C.iii D.i,ii E.i,iii F. ii,iii G. i,ii,iii H. None

(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）

4. Which of the following statements is/are true?
- The type-I error rate of a test can be controlled beforehand.
 - If the population is symmetric, then the population mean is equal to the population mode.
 - Central Limit Theorem guaranties that, if a fair coin was flipped for many times, then the proportion of the times to get head is about one half.
- A.i B.ii C.iii D.i,ii E.i,iii F. ii,iii G. i,ii,iii H. None
5. Which of the following statements is/are true?
- Under a regression model $y = a + bx_1 + cx_2$, we mean that one unit change of x_1 would lead to the change of b unit on y .
 - One calculates a 95% confidence interval for the population mean μ as $3.5 \leq \mu \leq 7.3$, which means the probability of μ falling into the interval of $[3.5, 7.3]$ is 0.95 given all of the assumptions being held.
 - If two events are independent, then they are mutually exclusive.
- A.i B.ii C.iii D.i,ii E.i,iii F. ii,iii G. i,ii,iii H. None
6. Which of the following statements is/are true?
- Type II error happens when the null hypothesis is true but rejected.
 - Coefficient of variance does not change with the change of the measurement unit.
 - The primary purpose of considering a cluster sampling instead of the simple random sampling is for the sampling convenience but not a better estimation.
- A.i B.ii C.iii D.i,ii E.i,iii F. ii,iii G. i,ii,iii H. None

編號: G 364 系所: 統計學系

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)

7. Which of the following statements is/are true?

- i. If a statistical procedure is developed for the ordered data, then it would be appropriate for the nominal data as well.
- ii. If two variables are independent, then the correlation coefficient between them has to be zero.
- iii. If there is no outlier exists in a regression analysis, then the absolute values of all the standardized residuals have to be less than or equal to 3.

A. i B. ii C. iii D. i, ii E. i, iii F. ii, iii G. i, ii, iii H. None

8. To fit a simple regression model, $y = ax + b$, one examined the normality plot of the response y first and obtained Figure 1 as following:

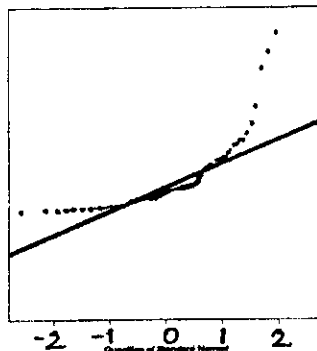


Figure 1: Normality Plot

Which of the following transformations might be able to improve the normality or/and stabilize the variance?

- i. y^2
- ii. $\log(y)$
- iii. $\frac{1}{\sqrt{y}}$

A. i B. ii C. iii D. i, ii E. i, iii F. ii, iii G. i, ii, iii H. None

(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）

9. Suppose that we used the original y in Question.8 to fit a linear regression model, which of the plots in Fig 2 could be the possible residual plots?

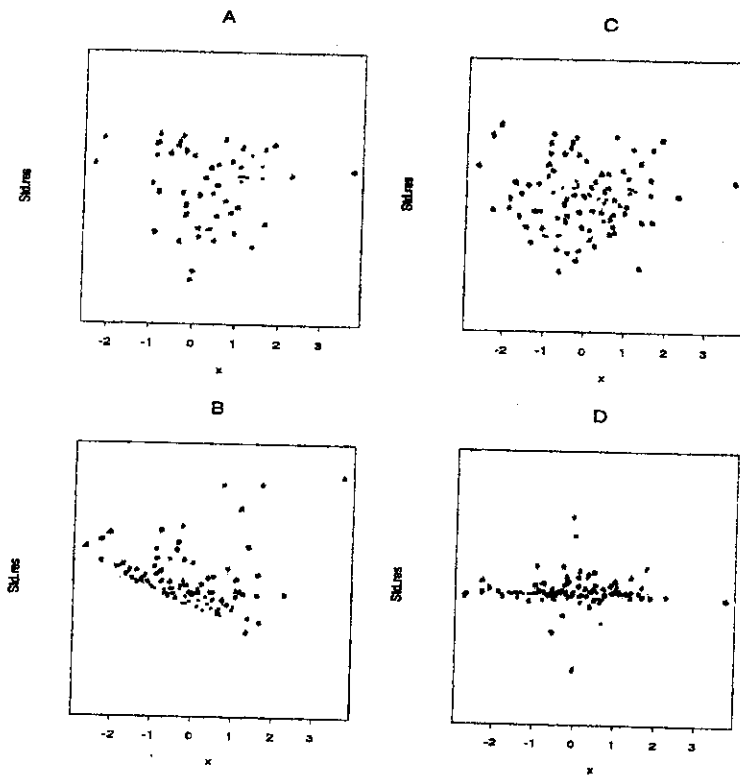


Figure 2:

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 Problems

1. (20%) It is expected that certain relationship exists between the education levels of the parents', that is, if the father has higher education level, then the education level of mother's tends to be high. To verify this, John collected the following data from the parents' education level from 75 high students. Let Z represent "Father's Education level" and Y represents "Mother's education level", and

$$Y, Z = \begin{cases} 1, \text{ Less than High School} \\ 2, \text{ High School} \\ 3, \text{ College and higher} \end{cases}$$

		Z		
		$Z = 1$	$Z = 2$	$Z = 3$
Y	$Y = 1$	11	6	6
	$Y = 2$	7	6	9
	$Y = 3$	6	7	17

He used the usual χ^2 test to examine the relationship between Y and Z . The value of the test statistic was 6.11, which gave the p-value of 0.19. That is, he failed to reject the null hypothesis of Y and Z being independent. Comment on his analysis (If you agree with him, please explain your reason(s). If you do not agree with him, please explain your reason(s) and also describe what you would do if you do not agree with him.)

(背面仍有題目,請繼續作答)

編號: G 364 系所: 統計學系

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)

2. (10%) We often use R^2 , which is defined as

$$R^2 = \frac{SS_{\text{reg}}}{SS_{\text{tot}}}$$

to describe how well the explanatory variable(s) could explain the response variable in a linear regression analysis. Comment on this criterion.

3. (10%) In the usual basic statistical course, you have learned how to examine if two populations have the equal variance by the ratio of their sample variances

$$\frac{S_1^2}{S_2^2}$$

based on F distribution with appropriate degrees of freedom. Comment on this test.

4. (15%) One would like to evaluate the efficacy of two brands of aspirin (denoted as A1, A2) and one brand of non-aspirin-based pain killer (denoted as NonA). He conducted an experiment in which 32 Patients were chosen to participate in this experiment and they were randomly divided into four groups. Each group received one of the following treatments; A1, A2, NonA, and placebo. The response variable is the self-reported degree of improvement/reduction of pain in the patient's headache, on a scale of 1 (no improvement) to 15 (no pain remaining). The following is the ANOVA table to test if there is any significant effects among these four treatments, and some summarized statistics obtained from MINITAB

編號: G 364 系所: 統計學系

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)

Descriptive Statistics: Placebo, A1, A2, NonA

Variable	N	Mean
Placebo	8	5.000
A1	8	6.000
A2	8	7.000
NonA	8	10.000

MTB > ADVOneway C1 C2 C3 C4.

One-way ANOVA: Placebo, A1, A2, NonA

Analysis of Variance

Source	DF	SS	MS	F	P
Factor	3	112.00	37.33	7.47	0.001
Error	28	140.005.00			
Total	31	252.00			

In this study, we would like to know the following questions:

- Are the real drugs (non-placebo) different from the placebo?
- Are A1 and A1 different?
- Are the aspirin-based pain killers different from the non-aspirin-based pain-killer?

Give the an appropriate ANOVA table (without p-values) with necessary calculation for these three questions, and describe how you will answer these three questions.