※ 注意：請於試卷上「非選擇題作答區」依序作答，並應註明作答之大題及小題題號。

【題一】（共 25 分）

以下的文章是由 2005 年 12 月出刊的 Biometrics 中，選出的一篇文章，題目為 Statistical Issues Arising in the Women's Health Initiative，請在答案紙上先標出行數（每一行最左邊為行數），再寫出錯誤的地方並且寫下正確的答案。例如：行數 1, perhap 改成 perhaps

行數

1. The Women's Health Initiative (WHI) is perhap the most ambitious population research
2. investigation ever undertake. The centerpiece of the WHI program is a randomized, controlled
3. clinic trial (CT) to evaluating the health benefits and risks of four distinct interventions (dietary
4. modification, two postmenopausal hormone therapy [HT] interventions, and calcium/vitamin D
5. supplementation) among 68,132 post-menopausal woman in the age range 50 – 79 at
6. randomization. Participating women identified from the general population live in
7. proximity to any of the 40 participating clinical center throughout the United States. The WHI
8. program also include an observational study (OS) that comprise 93,676 postmenopausal
9. women recruiting from the same population base as the CT. Enrollment into WHI began in
10. 1993 and conclude in 1998. Intervention activities in the estrogen plus progestin HT
11. component of the CT end early in July 8, 2002 when evidence had accumulated that the risks
12. exceeded the benefits. Intervention activities in the estrogen-alone component of the CT also
13. end early, in February 29, 2004. Intervention activities in the another two CT components end
14. in March 31, 2005. Nonintervention follow-up on participating women is plan through 2010,
15. giving a average follow-up duration of about 13 years in the CT and 12 year in the OS.

【題二】（共 25 分）

Please use not less than 150 words to introduce the regression analysis to someone who knows statistics not too much.

接背面

【題三】：填充題(共 25 分)（請填入適當之英文單字或詞句）

A one-way contingency table displays the frequencies of occurrence of two or more mutually exclusive and jointly exhaustive categories in a random sample of total size N. In particular, a 2×1 contingency table displays the frequencies of occurrence of just __(1)__ , and this can be taken to result from a population in which __(2)__ of cases within the two categories are p and (1-p), say.

In the case of 2×1 contingency tables, the probability of obtaining x cases in one category and (N-x) in __(3)__ is specified by __(4)__. For moderately large samples, the probability of obtaining x cases in one category becomes fairly difficult to compute. The traditional solution to this problem has been to rely upon __(5)__. Under the null hypothesis, the quantity x is asymptotically normal with a __(6)__ equal to Np and a __(7)__ equal to Np(1-p).

An alternative strategy for the analysis of contingency tables was put forward in 1900 by K. Pearson whilst seeking to develop a statistical test of goodness of fit. Given asset of categorical data consisting k observed frequencies of the form $O(j)$ and a corresponding set of __(8)__ of the form $E(j)$, subject only to the constraint that $\sum O(j) = \sum E(j)$, Pearson assumed that the $O(j)$ would approximate to a multivariate normal distribution. He then showed that the statistic $X^2 = \sum [O(j) - E(j)]^2 / E(j)$ would be asymptotically distributed as __(9)__.

【題四】（共 25 分）

1. A regression model was used to assess the effect of alcohol drinking on liver cirrhosis after _____ for hepatitis virus infection. (3 分)
(A) collecting
(B) completing
(C) controlling
(D) investigating

2. The likelihood ratio test was performed to _____ model I with model II. (3 分)
(A) assess
(B) compare
(C) analyze
(D) yield

3. The _____ of the component Z into the expression (1) yields the expression (2). (3 分)
(A) display
(B) replacement
(C) use
(D) substitution

接次頁

4. Pair t-test is applied to _____ data in the same individual (3 分)
(A) independent
(B) empirical
(C) correlated
(D) epidemiological

5. Logistic Regression model is _____ for the outcome with binary property. (3 分)
(A) related
(B) adequate
(C) incompetent
(D) confident

6. Chi-square test was related to the assessment of the association between two _____ _____data. (3 分)
(A) empirical
(B) categorical
(C) interval
(D) dependent

7. We adopted likelihood ratio tests to identify a ___(7.1)___ model among a series of ___(7.2)__model.
(7.1) (3 分)
(A) frugal
(B) economic
(C) prudent
(D) parsimonious
(7.2) (4 分)
(A) established
(B) nested
(C) explicit
(D) lucid

試題隨卷繳回