

**Problem 1**

The following data were obtained from a survey of thyroid dysfunction among an elderly population aged 65 or over. This sample was drawn by a simple random sampling scheme from the elderly population.

Sex	Thyroid dysfunction		Total
	Yes	No	
Female	28 (3.05%)	368(40.13%)	396(43.18%)
Male	35(3.82%)	486(53%)	521(56.82%)
Total	63(6.87%)	854(93.13%)	917(100%)

Inside the parentheses are percents of the total sample.

- (1) There are four cells in this 2 by 2 table. Please determine the joint distribution of the counts of the 4 cells when repeated samples are drawn (i.e. the sampling distribution of the counts). (5 分)
- (2) Please construct a 95% confidence interval on the proportion of thyroid dysfunction among the elderly population (so-called prevalence of thyroid dysfunction) using normal approximation to a binomial distribution when the sample size is sufficiently large. (5 分)
- (3) Please use Pearson chi-square statistics to test whether the prevalence of thyroid dysfunction is sex-dependent. A p-value less than 0.05 indicates statistical significance. (5 分)
- (4) Please construct a 95% confidence interval on the odds ratio measuring the strength of association. Assume that the sample size be sufficiently large. (5 分)

Note: The upper 2.5 percentile of a standard normal distribution is 1.96 and the upper 5 percentile of a chi-square distribution with 1 degree of freedom is 3.84.

**Problem 2**

In the planning stage of the above survey, the sample size was determined by either of the following approaches.

- (1) If the sample prevalence of thyroid dysfunction were to be within 3 percentage point of the population prevalence (i.e., the margin of error is  $\pm 3\%$ ) at the confidence level of 95%, what would be the required sample size? (5 分)
- (2) If the population prevalence were 4% (i.e., Assume that under the null hypothesis

(背面仍有題目,請繼續作答)

the prevalence be 4%), the probability of rejecting it would not exceed 5% (i.e., The probability of type I error is not larger than 0.05); If the population prevalence were 7% (i.e., Assume under the alternative hypothesis the prevalence be 7%), the probability of rejecting it would not exceed 10% (i.e., The probability of type II error is no larger than 10%). Using the sample prevalence as the test statistics, what would be the required sample size? (5 分)

- (3) Please compare the sample sizes in (1) and (2), and comment on the comparison. (10 分)

Note: The upper 5 percentile of a standard normal distribution is 1.645 and the upper 10 percentile of a standard normal distribution is 1.28.

### Problem 3

目前造成世界恐慌的 SARS 致病微生物仍未明。學者提出假說有可能為單一因子也可能是多重因子(multifactorial)，請設計一流行病學研究來探討其致病因(single factor and multifactorial)。請以一虛擬的數據，來解釋此一研究之發現，並說明如何探究因果關係。(30 分)

### Problem 4

冠心病 (coronary heart disease) 的臨床表現相當多元，包括心肌梗塞、心絞痛、心律不整、傳導障礙、心因性猝死等。冠心病為國人重要的死因之一，台灣每年約有近六千人因冠心病而死亡，粗死亡率每十萬人口約 25 人左右。衛生署的統計資料顯示，過去三四十年來台灣冠心病的年齡標準化死亡率逐年上升，但近幾年來 (1992 之後) 似乎開始呈現下降趨勢，然而有關其發生率或盛行率的趨勢調查卻仍十分有限。

- 1) 請說明年齡標準化的目的。(5 分)
- 2) 若僅以衛生署統計的死亡率作為疾病防治之依據，可能會有什麼限制?(5 分)
- 3) 近幾年冠心病死亡率的下降，你認為可能的原因為何？(至少提出兩種不同的原因)(10 分)
- 4) 針對上題你所提出的可能原因，請分別說明你將如何設計研究來檢驗這些想法?(10 分)