一、選擇題（40 分，共 20 小題，每小題 2 分）

探討大腸直腸癌和某飲食因子之相關，研究者使用病例對照研究 (Case-control Study)，由於年齡和性別是重要干擾因子，因此研究者使用年齡-性別-1:1 (一個病例配對一個對照) 配對設計 (Matched Design)。

符號之定義如下

病例（大腸直腸癌）：以 D 表示

對照（無大腸直腸癌）：以 $\overline{D}$ 表示

暴露此飲食因子：以 F 表示

無暴露此飲食因子：以 $\overline{F}$ 表示

據此得到結果如下：

(i) D 和 $\overline{D}$ 皆是 F：30 對

(ii) D 是 F，而 $\overline{D}$ 是 $\overline{F}$：70 對

(iii) D 是 $\overline{F}$，而 $\overline{D}$ 是 F：30 對

(iv) D 和 $\overline{D}$ 皆是 $\overline{F}$：70 對

(以下訊息作為下列運算之用：ln2=0.6931, ln30=3.4012, ln70=4.2485, ln100=4.6052)

其虛無假說(H_0)：病例與對照之暴露(F)機率相同

(1) 任何一配對(Matched Pair)所感興趣的結果是病例及對照暴露之機率，針對上述(i)-(iv)之訊息，何者提供較有用之訊息

(A) (i)+(iv)　(B) (ii)+(iii)　(C) (i)+(iii)　(D) 四者皆具同樣之訊息

見背面

(2) 根據題(1)，您建議使用何種統計分佈來描述此樣本

(A) 二項(Binomial)分佈　(B) 布瓦松(Poisson)分佈　(C) 貝他(Beta)分佈

(D) 以上皆非

(3) 若根據 H_0 及題(1)與題(2)之情境下，則期望配對數(μ)等於

(A) 30　(B) 50　(C) 70　(D) 100

(4) 在 H_0 及題(1)與題(2)之情境下，吾人欲計算有關(ii)訊息的隨機變數 Y 之

$\Pr\left(|Y-\mu|\geq 20\right)$，若使用柴比雪夫不等式(Chebychev inequality)，其機率為

(A) $\frac{1}{2}$　(B) $\frac{1}{4}$　(C) $\frac{1}{16}$　(D) 訊息不足，無法計算

(5) 上述題(4)之柴比雪夫不等式之機率計算，若樣本數增加至∞，則其機率為

(A) $\frac{1}{2}$　(B) $\frac{1}{4}$　(C) $\frac{1}{16}$　(D) 以上皆非

(6) 上述題(5)之柴比雪夫不等式之機率計算是屬於

(A) Converge almost surely　(B) Converge in probability　(C) Converge in distribution　(D) 以上皆非

(7) 吾人欲使用中央極限定理(Central Limit Theorem) 計算上述

$\Pr\left(|Y-\mu|\geq 20\right)$，則會使用何種分佈來達成近似估計?

(A) 二項(Binomial)分佈　(B) 布瓦松(Poisson)分佈　(C) 貝他(Beta)分佈

(D) 以上皆非

(8) 使用中央極限定理下之大樣本理論及所提供樣本資料，在 H_0 之下(若不進行

接次頁

修正)得到檢定統計值約接近

(A) 1 (B) 2 (C) 4 (D) 以上皆非

(9) 根據題(1)及題(2)之情境，若使用廣義線性模式(Generalized Linear Model)來檢定虛無假說及對立假說，須估計幾個迴歸係數?

(A) 1 (B) 2 (C) 3 (D) 以上皆非

(10) 其中有關暴露(主要因子)之迴歸係數最大概似估計值(MLE)為

(A) 0 (B) 0.85 (C) 3.3 (D) 2.3

(11) 若迴歸係數之變異數以大樣本理論計算，則約為

(A) 0.01 (B) 0.03 (C) 0.05 (D) 0.08

(12) 使用中央極限定理下之大樣本理論對上述迴歸係數進行檢定，其檢定統計值約為

(A) 介於 0~1 (B) 介於 1~2 (C) 介於 2~3 (D) 以上皆非

(13) 若將上述檢定統計值取平方,則得到何種近似分佈

(A) 常態(Normal)分佈 (B) 二項(Binomial)分佈 (C) 卡方(Chi-square)分佈 (D) 以上皆非

(14) 若使用大樣本理論計算迴歸係數 MLE 值之 95%CI，其下限為

(A) 0.22 (B) 0.32 (C) 0.42 (D) 以上皆非

(15) 若暴露和疾病之相關支持 H_0，則迴歸係數為

(A) 0.23 (B) 0.85 (C) 1 (D) 以上皆非

見背面

(16) 根據題(15)，進行在 H_0 成立下其參數之檢定，統計檢定值為

(A) 2　(B) 4　(C) 16　(D) 以上皆非

(17) 上述題(16)之檢定統計值之近似分佈為

(A) F 分佈　(B) 卡方(Chi-square)分佈　(C) t 分佈　(D) 以上皆非

(18) 若使用概似函數 (Likelihood function) 計算在此樣本資料所得到 MLE 之下和參數相關之對數函數概似值，取絕對值後約接近

(A) 21　(B) 41　(C) 61　(D) 以上皆非

(19) 若使用概似函數比例檢定 (Likelihood Ratio Test), 將上述題(18)之值與 H_0 支持下之迴歸係數所得到對數概似函數相減再乘以 2 倍所得到數值的絕對值約為

(A) 41　(B) 61　(C) 81　(D) 以上皆非

(20) 上述檢定統計值之分佈為

(A) F 分佈　(B) t 分佈　(C) 常態(Normal)分佈　(D) 以上皆非

二、(40 分，共 10 小題)

一研究收集兩組樣本共有 $n+m$ 筆高血壓數值之資料，此資料以符號表示如下：

	高血壓數值	平均值	標準差
樣本一	$x_{11}, x_{21}, \ldots, x_{n1}$	$\bar{x}_1 = \sum_{i=1}^{n} x_{i1} / n$	$s_1 = \sqrt{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2 / (n-1)}$
樣本二	$x_{12}, x_{22}, \ldots, x_{m2}$	$\bar{x}_2 = \sum_{i=1}^{m} x_{i2} / m$	$s_2 = \sqrt{\sum_{i=1}^{m}(x_{i2} - \bar{x}_2)^2 / (m-1)}$

接次頁

樣本一之資料取自群體(P1)，樣本二之資料則取自另一群體(P2)，欲檢定 (P1) 與 (P2)此兩群體(population)之高血壓分布(distribution)是否相同，建議使用以下之統計檢定方法：

『在任何 n>1 且 m>1 之情況下，使用檢定統計量 T_1，且在虛無假說(null hypothesis)為真時，此 T_1 之分布為 t 分布其自由度為 n+m-2。』

(1) 根據上表之資料符號，寫出此檢定統計量 T_1。(4 分)

(2) 說明該如何收集/選取樣本一與樣本二之資料，使得此建議之檢定統計量 T_1 是一適合的統計檢定方法。(4 分)

(3) 說明(P1) 與(P2)此兩群體之高血壓分布須符合哪些條件(assumptions)，才會使得「在虛無假說(null hypothesis)為真時，此檢定統計量 T_1 之分布是為 t 分布其自由度為 n+m-2」的結果是成立的。(4 分)

(4) 根據題(3)所須符合的條件，寫出所對應之虛無假說(null hypothesis)。(4 分)

(5) 若以簡單線性迴歸模式分析此兩組樣本資料時，請說明應如何重新定義此資料結構的符號表示方式。［提示：須定義出依變數(dependent variable)與自變數(independent variable)］(4 分)

(6) 根據題(5)所定義之新的資料符號，寫出所對應之簡單線性迴歸模式。(4 分)

(7) 根據題(6) 之簡單線性迴歸模式，寫出所需符合的模式假設條件(assumptions)。(4 分)

(8) 根據題(7)之簡單線性迴歸模式，寫出所對應之虛無假說(null hypothesis)。(4 分)

(9) 根據題(5)所定義之新的資料符號，寫出所對應之檢定統計量 T_2。(4 分)

(10) 驗證 T_1 與 T_2 是相同的檢定統計量。(4 分)

見背面

三、（20 分，共 3 小題）

假設隨機變數 T 代表肺癌病人的存活時間。雖然很多研究對於 T 有高度的興趣，但往往在實際研究中，病人因為一些跟疾病不相干的因素離開了研究(例如搬家)，導致吾人無法觀測到 T，這種狀況稱之為設限(censoring)。假設隨機變數 C 代表設限的時間。在實際的研究當中吾人無法觀測到 (T,C)，而只能觀測到 (X,δ)，其中 $X=\min\{T,C\}$，δ 為一個二元(binary)隨機變數，$\delta=1$ 當 $T<C$，$\delta=0$ 當 $C<T$。假設 T 服從參數為 λ 的指數分配（pdf 為 $f_T(t)=\frac{1}{\lambda}e^{-\frac{t}{\lambda}}$)，$C$ 服從參數為 μ 的指數分配(pdf 為 $f_C(c)=\frac{1}{\mu}e^{-\frac{c}{\mu}}$)，且 T 和 C 獨立。

(1)請驗證$P(X\le x,\delta=1)=\frac{\mu}{\lambda+\mu}\left\{1-e^{-\left(\frac{1}{\mu}+\frac{1}{\lambda}\right)x}\right\}$ 以及 $P(X\le x,\delta=0)=\frac{\lambda}{\lambda+\mu}\left\{1-e^{-\left(\frac{1}{\mu}+\frac{1}{\lambda}\right)x}\right\}$。(8 分)

(2)假設吾人蒐集到以下資料 $(X_i,\delta_i), i=1,\ldots,n$。請寫出 (λ,μ) 的概似函數(likelihood function)。(6 分)

(3)承上題，請求出 (λ,μ)的最大概似估計(maximum likelihood estimate)。(6 分)

試題隨卷繳回