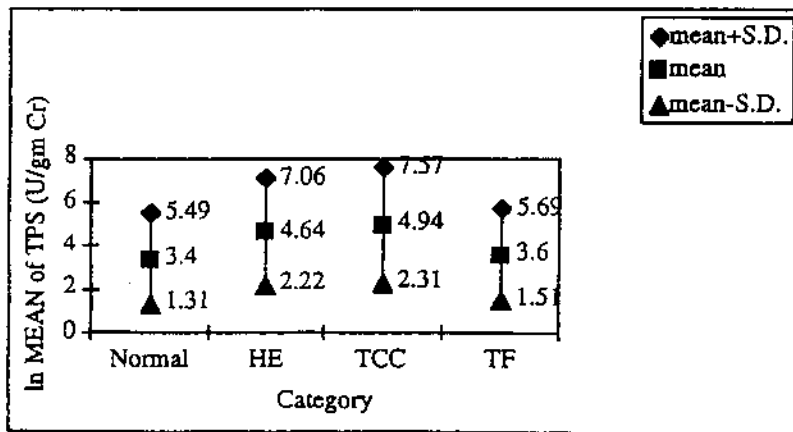


統計學

一、(65分) 在腫瘤診斷上，一個可能的方式是利用病患尿液細胞中的 TPS (Tissue Polypeptide specific Antigen) 值。由於 TPS 的值皆為正，在資料處理上為分析的方便，我們常將其做自然對數的變換而得 $\ln TPS$ 。假設成功大學附屬醫院共蒐集到 Normal, Free of Tumor (FT), Hematuria (HE), and Transitional Cell Carcinoma (TCC) 等四組 TPS 的資料，其個數分別為 75, 36, 44 和 56 人。

A. (35分)

- (1). (4分) 若欲就 $\ln TPS$ (令其為 Y) 來分析上述四群的均數是否相等，一個方式是利用變異數分析 (ANOVA) 法來做。試寫出其模型和基本假設。
- (2). (4分) 試概略說明如何檢定(1)中之假設。
- (3). (4分) 若(1)中之假設符合，原來四組之 TPS 應分別呈何種分配？試列出各組資料所來自之母體的分配形式，你必須以與(1)相同之符號求出 TPS 的均數和變異數。
- (4). (4分) 以(1)之模型為基礎求各組均數和變異數的 maximum likelihood estimators.
- (5). (4分) 各組之樣本均數及標準差如下之圖形所示。



在此 S.D. 表各組之有偏樣本變異數的開方。請問(4)之 MLE 的估計值為何？

- (6). (10分) 設 $T(Y)$ 為由 (4) 所求得之 MLEs 所組成的隨機向量。求全部觀察值向量 Y 在 $T(Y)$ 下的條件密度。
- (7). (5分) 若我們要做變異數分析，(5) 中所求得各 MLE 的值和上述所給之各組的 sample sizes 是否已足？為何？
- B. (10分) 假設就 (A) 之 (1) 檢定各組之 $\ln TPS$ 的均數結果是 reject H_0 ，則下一個步驟是知道 $\ln TPS$ 在那些組上有差異。做 Fisher's least significant difference (LSD) 和 Tukey 的多重比較得如下之結果：

T tests: Fisher's least significant difference (LSD) for $\ln TPS$
experimentwise error rate = 0.05

T Grouping	Mean	N	Group
A	4.94	56	TCC
A			
A	4.64	44	HE
B	3.4	36	FT
B			
B	3.6	75	Normal

Tukey's Studentized Range (HSD) test for $\ln TPS$
comparisonwise error rate = 0.05

Tukey Grouping	Mean	N	Group
A	4.94	56	TCC
A			
B A	4.64	44	HE
B			
B C	3.4	36	FT
C			
C	3.6	75	Normal

- (1). (5分) 何謂 experimentwise error rate? 何謂 comparisonwise error rate?
 (2). (5分) 試解釋上述 Fisher's least significant difference test 和 Tukey's studentized range test 之結果。

C. (20分) 假設現在對上述四組病患除 TPS 外亦有年齡 (AGE) 和性別 (SEX) 等資料，某研究者決定將 AGE 和 SEX 置入 A 之 (1) 的 ANOVA 中，而成 Analysis of Covariance (ANCOVA) 模型。

- (1). (5分) 為何此研究者欲用 ANCOVA? 即 ANOVA 和 ANCOVA 分析的基本目的或差異何在?
 (2). (5分) 試以迴歸模型的方式寫出此 ANCOVA 模型。請勿忘記基本假設。
 (3). (5分) 利用 SAS 之 GLM (General Linear Model) 執行上述之 ANCOVA 得如下之結果：

Source	d.f.	SS	MS	P-value
Model	(a)	(b)		0.0004
Error	(c)	1082.4		
Corrected Total	210	1201.9		

Source	d.f.	TYPE I Sum of squares	P-value
(d)	(e)	(f)	0.0005
AGE d	.	16.3	0.0803
SEX d, AGE	.	5.67	0.3014

P-value 矩陣： $H_0: \mu_i = \mu_j, i, j \in \{Normal, FT, HE, TCC\}, i \neq j.$

	Normal	TCC	FT	HE
Normal	.	0.0006	0.9022	0.0067
TCC	0.0006	.	0.0059	0.6228
FT	0.9022	0.0059	.	0.0296
HE	0.0067	0.6228	0.0296	.

就(2)中模型的假設，以符號或數值填入上述二表中之()。

(4). (5分) 解釋(3)之檢定的結果，並和(B)之結果比較。

二、(35 points, 7 points each) In regular estimation, the ratio of the Cramer-Rao lower bound to the variance of any unbiased estimator $\phi_n(X)$ of a parameter is called the efficiency of that estimator. Suppose we have an i.i.d. random sample X_1, X_2, \dots, X_n form $N(\mu, \sigma^2)$. Let $X = [X_1, X_2, \dots, X_n]^T$. We are interested in σ^2 .

- (1). Find the UMVU estimator $T(X)$ of σ^2 .
- (2). Find the Cramer-Rao lower bound for the parameter σ^2 .
- (3). Find the efficiency of $T(X)$.
- (4). Construct a confidence interval for σ^2 based on the UMVU estimator $T(X)$.
- (5). Find the generalized likelihood ratio test for $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$, where σ_0^2 is a known value.