題組（每題組 6 分）

請依順序回答下面題目（若不依順序回答，不予計分），每一題組有二小題，全對才給分（請將題號標示清楚，如 (A) 1, (A) 2）。下面陳述若有錯誤，請寫上 False; 若陳述完全正確，請寫上 True，均不用多做說明。

(A)

1. 平均數 (mean) 是一種位置測度 (location measures)，會受極端值 (extreme value) 所影響。中位數 (media) -- 即第 50th 百分位數 (percentile) -- 也是一種位置測度，但卻不會受到極端值所影響。綜合而言，平均數會受極端值所影響，但分位數 (quantile) 及百分位數（如百分之 25，百分之 95 分位數）均不會受極端值所影響。
2. 變異數 (variance) 是一種變異性測度 (variation measure)，在財務學中我們常把它當成是風險 (risk) 的測度。全距 (range)，變異係數 (coefficient of variation) 以及 $1/(n-1)\sum_{i=1}^{n}|x_i-\bar{x}|$ 也是變異性測度，因此也可以當成另一種風險的測度。

(B)

1. 令 A,B 為兩互斥事件 (mutually exclusive events)，$\mathrm{P}(A\cap B)=0$，且 $\mathrm{P}(A)\succ 0$，$\mathrm{P}(B)\succ 0$。因此 A 事件發生時，B 事件一定不會發生。換句話說，A,B 事件沒有什麼關係，即 A,B 為兩相互獨立 (independent) 的事件。
2. 已知 $\mathrm{E}(X-\mathrm{E}(X))^2\le \mathrm{E}(X-c)^2,\forall c$；其中 c 為任何實數，此為均方誤 (mean squared errors) 的一個重要特性。令 X,Y 為兩獨立的隨機變數，$\varphi(Y)$ 為 Y 的函數，則 $\mathrm{E}(X-\mathrm{E}(X))^2\le \mathrm{E}(X-\varphi(Y))^2$ 亦成立。

(C)

1. 令 X 為一連續隨機變數 (continuous random variable)，其機率密度函數 (probability density function) $f_X(a), c<a<d$ 存在。由於 X 在 (c,d) 區間為連續，根據連續隨機變數的定義，$X=a$ 的機率必定為 0，因此 $\mathrm{P}(X=a)=f_X(a)=0$。

2. 令 X,Y 的聯合分配 (joint distribution) 存在，其連續的聯合機率密度函數為 $f_{X,Y}(a,b)=1/4,$ 當 $0\le a\le 2, 0\le b\le 2$；$f_{X,Y}(a,b)=0$ 當 a,b 為其它值。則 $\mathrm{P}(X\le 1, Y\le 1)=1/4$，$\mathrm{P}(Y=X^2)\succ 0$。

(D)

1. 若 $X\sim N(a,1)$，$Y\sim N(b,1)$ 均為常態分配，則 X,Y 的聯合機率分配亦為常態分配，且線性轉換後的變數 $aX+b$ 及 $a+bY$ 也服從常態分配，其平均數 (mean) 分別為 a^2+b 及 $a+b^2$。

2. 若 $X\sim N(a,1)$，$Y\sim N(b,1)$ 均為常態分配，則線性轉換後的變數 $aX-bY$ 亦為常態分配，其平均數恰為 $(a-b)(a+b)$。

(E)

1. 令 Student's t 分配為 $t(n)$，其中 n 為其自由度；令 $F(m,n)$ 為 F 分配，其自由度分別為 m,n。當 $n=3$ 時，$t(n)$ 的第一及第二階中央動差存在 (平均數為 0，變異數為 3)，但其第三階中央動差不存在。此外，將 $t(3)$ 做非線性轉換，則 $t(3)^2=F(1,3)$。不同的是，$t(3)^2$ 的第一及階中央動差存在，但第二階中央動差 (即變異數) 及第三階中央動差則不存在。

2. 當樣本大於 30 以上時稱做大樣本，樣本小於 30 則稱做小樣本。並且不論樣本大於 30 還是小於 30，我們的重點都是找出統計量的實際分配，只有當無法導出實際分配時，才退而求其次去推導統計量的極限分配 (asymptotic distribution)。

(F)

1. 令 $\{X_1,\ldots,X_n\}$ 為一組 i.i.d. (independent and identical distributed) 隨機變數，其分配為 $N(\mu,\sigma^2)$，其中假設 σ^2 為已知參數。令 $\bar{X}=1/n\sum_{i=1}^{n}X_i$，則在信賴係數為 90% 之下 μ 的信賴區間為 $\left(\bar{X}-1.645\sigma/\sqrt{n},\bar{X}+1.645\sigma/\sqrt{n}\right)$，此一信賴區間為一隨機區間。換句話說，此區間並非固定 (是隨機的)，其上下界並非是實數。

2. 令 $\hat{\theta}$ 代表 $\theta_0=1$ 的不偏估計式，而 X 為一隨機變數，$\mathrm{E}(X)=1,\mathrm{var}(X)=5$，且 $\mathrm{cov}(\hat{\theta},X)=2$。則新估計式 $\tilde{\theta}_1=\hat{\theta}-X$ 與 $\tilde{\theta}_2=\hat{\theta}+X$ 均為偏誤估計式 (biased estimator)，但 $\tilde{\theta}_1$ 比較有效 (efficient)，其次是 $\tilde{\theta}_2$，最後才是 $\hat{\theta}$。

(G)

1. 令 $\{X_1,\ldots,X_n\}$ 為一組 i.i.d. 白奴里 (Bernoulli) 隨機變數，其值等於 0 或 1；等於 1 的機率為 p，則 $\mathrm{E}(X_i)=p,\mathrm{var}(X_i)=p(1-p),\forall i$。已知 $\bar{X}=1/n\sum_{i=1}^{n}X_i$ 為 p 的不偏估計式，則 $\mathrm{var}(X_i)$ 的不偏估計式為 $\bar{X}(1-\bar{X})$。

2. 令 $\{X_1,\ldots,X_n\}$ 為一組 i.i.d. 白奴里 (Bernoulli) 隨機變數，其值等於 0 或 1；等於 1 的機率為 p，則 $\mathrm{E}(X_i)=p,\mathrm{var}(X_i)=p(1-p),\forall i$。已知 $\bar{X}=1/n\sum_{i=1}^{n}X_i$ 為 p 的不偏估計式，則 $\mathrm{var}(X_i)$ 的一致估計式 (consistent estimator) 為 $\bar{X}(1-\bar{X})$。

國 立 清 華 大 學 命 題 紙

97 學年度計量財務金融系（所）甲組碩士班入學考試
乙組

科目統計學 科目代碼 5103、5203 共 7 頁第 4 頁 *請在【答案卷卡】內作答

(H)

1. 令 $\{X_1,\ldots,X_n\}$ 為一組 i.i.d. (independent and identical distributed) 隨機變數，其平均數為 μ，變異數為 $\sigma^2<\infty$。因此檢定 $H_0:\mu=3$ 時，可以利用 $\sqrt{n}(\bar{X}-3)/S\sim t(n-1)$ 的特性建構檢定統計分析，其中 $S^2=1/(n-1)\sum_{i=1}^{n}(X_i-\bar{X})^2$。

2. 令 $\{X_1,\ldots,X_n\}$ 為一組 i.i.d. (independent and identical distributed) 隨機變數，其平均數為 μ，變異數為 $\sigma^2<\infty$。又假設 $\{X_1,\ldots,X_n\}$ 分配為 $N(\mu,\sigma^2)$，σ^2 已知。在檢定 $H_0:\mu=3$ 時，我們可以利用 $\sqrt{n}(\bar{X}-3)/\sigma\sim N(0,1)$ 來建構檢定統計分析，我們也可以利用 $\sqrt{n}(\bar{X}-3)/S\sim t(n-1)$ 來建構檢定統計結果，其中 $S^2=1/(n-1)\sum_{i=1}^{n}(X_i-\bar{X})^2$。利用前者做假設檢定會比後者來得更有效 (efficient)。

(I)

1. 假設簡單迴歸模型為 $y_t=\beta x_t+\varepsilon_t$，$\varepsilon_t\sim N(0,\sigma^2)$。令 $\hat{\varepsilon}_t$ 為此迴歸模型的殘差 (residual)，此時我們可計算出 $\sum_{t=1}^{n}\hat{\varepsilon}_t=\sum_{t=1}^{n}x_t\hat{\varepsilon}_t=0$。

2. 假設迴歸模型 1 為 $y_t=\beta x_t+\varepsilon_t$，其 $R^2=0.95$；迴歸模型 2 為 $\ln(y_t)=\beta x_t+\varepsilon_t$，$R^2=0.5$。若我們利用 R^2 來判定那個模型比較好，則可知模型 1 比模型 2 好。若我們利用調整後的 R^2 (adjusted R^2) 來比較時，也會得到同樣的答案。

(J)

1. 考慮一個簡單迴歸模型 $y_t = \beta x_t + \varepsilon_t$，$\varepsilon_t \sim N(0,\sigma^2)$。在古典條件 (classical conditions) 下，最小平方估計式 $\hat{\beta}$ 及 $\hat{\sigma}^2$ 為迴歸模型的最佳線性不偏估計式 (best linear unbiased estimator; BLUE)；其中 $\hat{\beta}$ 及 $\hat{\sigma}^2$ 為 β 及 σ^2 的最小平方估計式。

2. 考慮一個簡單迴歸模型 $y_t = \beta x_t + \varepsilon_t$。模型 1 假設變異數具有齊一性 (homoskedasticity): $\varepsilon_t \sim N(0,\sigma^2)$。模型 2 假設變異數具不齊一性 (heteorskedasticity): $\varepsilon_t \sim N(0,\sigma_t^2)$，$\sigma_i^2 \neq \sigma_j^2$。假設 $\{(y_1,x_1),\ldots,(y_n,x_n)\}$ 為研究者所收集到的資料。此時不論是模型 1 還是模型 2，所計算出的最小平方估計值 (OLS estimates) 均一樣。

Just write down the answer (only the answer) in the answer sheet.

1. (10%) Let $X_1, X_2, \ldots, X_6$ be a sample from a uniform distribution on $[0,\theta]$, where $\theta \in [1,2]$ is an unknown parameter. Find an unbiased estimator for θ of variance less than 1/10.

2. (10%) Let $\{-0.8,-0.4,-0.2,0.6\}$ be a sample from a uniform distribution on [-1,1]. Based on this sample, please generate (simulate) a sample which is from an exponential distribution $0.5e^{-x/2}$.

3. (10%) Let $X_1, X_2, \ldots, X_n$ be a sample from a distribution (not necessary Normal distribution) with mean μ and variance $\sigma^2 < \infty$. Let $\ddot{\mu} = \frac{1}{[n/3]}\sum_{i=1}^{[n/3]} X_i$, where [.] represents the largest integer less than or equal to n/3. Please construct a suitable 95% confidence interval for μ based on the estimator $\ddot{\mu}$.

4. (10%) Use the following information to answer Questions 4.1 through 4.5.

Multiple regression was used to explain stock returns using the following variables:

Dependent variable:

Dependent variable:

RET = annual stock returns (%)

Independent variables:

MKT = Market capitalization = Market capitalization / \$1.0 million

IND = Industry quartile ranking (IND = 4 is the highest ranking)

FORT=Fortune 500 firm, where {FORT = 1 if the stock is that of a Fortune 500 firm, FORT=0 if not a Fortune 500 stock}

The regression results are presented in the tables below

		Coefficient	*Standard Error*	*t-Statistic*	*p-Value*
Intercept		0.5220	1.2100	0.430	0.681
Market Capitalization		0.0460	0.0150	3.090	0.021
Industry Ranking		0.7102	0.2725	2.610	0.040
Fortune 500		0.9000	0.5281	1.700	0.139
ANOVA	*df*	*SS*	*MSS*	*F*	*Significance F*
Regression	3	20.5969	6.8656	12.100	0.006
Error	6	3.4031	0.5672		
Total	9	24.0000			

Test	*Test-Statistic*
Breusch-Pagan	9.3
Durbin-Watson	1.8

4.1 The expected amount of the stock return attributable to it being a Fortune 500 stock is *closest to:*

A . 0.522.

B. 0.046.

C. 0.710.

D. 0.900.

4.2 The expected return on the stock of a firm that is not in the Fortune 500, has a market capitalization of \$5 million, and is in an industry with a rank of 3 is *closest* to:

A. 2.88%.

B. 3.98%.

C. 1.42%.

D. 2.10%.

4.3 Does being a Fortune 500 stock contribute significantly to stock returns?

A. Yes, at a 10% level of significance.

B. Yes, at a 5% level of significance.

C. No, not at a reasonable level of significance.

D. No, not at a 15% level of significance.

4.4 The p-value of the Breusch-Pagan test is 0.0005. The lower and upper limits for the Durbin-Watson test are 0.40 and 1.90, respectively. Based on this data and the information in the tables, there is evidence of:

A. only multicollinearity.

B. only serial correlation.

C. serial correlation and heteroskedasticity.

D. only heteroskedasticity.

4.5 Ron Working incorrectly uses the standard error of estimate instead of the standard error of the forecast in his calculation of the confidence interval for the predicted value from a simple linear regression with 26 observations. All else equal, the confidence interval he calculates will be:

A. the same as the correct confidence interval.

B. wider than the correct confidence interval.

C. narrower than the correct confidence interval.

D. wider than the correct confidence interval only if the R^2 is greater than 0.50.