

114年公務、關務人員升官等考試、114年
交通事業郵政、港務人員升資考試試題

等 級：薦任

類科(別)：經建行政、工業行政

科 目：統計學

考試時間：2 小時

座號：_____

※注意：(一)可以使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

(三)本科目除專門名詞或數理公式外，應使用本國文字作答。

一、某研究者以樣本數 $n=20$ 的資料建立了一個迴歸模型，已知依變數 Y 的樣本變異數為 $s_y^2=10.53$ ，且該模型的殘差平方和 (Sum of Squared Errors) 為 $SSE=30$ 。

(一)請說明如何利用這些已知的資訊計算出判定係數 (Coefficient of Determination) $R^2=0.85$ ，並請解釋判定係數所代表的統計意涵。(10 分)

(二)一般而言，若模型能解釋約 85% 的變異，常被認為是「非常好」的模型。請就「擬合程度」、「解釋性」與「預測性」三個面向加以討論，說明高 R^2 並不必然代表模型具有良好的整體表現。(10 分)

二、某疾病在一般族群中的盛行率 (發生率) 為 1%。某快速檢驗的敏感度 (即確實患病者被檢測為陽性的機率) 為 95%，特異度 (即未患病者被檢測為陰性的機率) 為 90%。請回答下列問題：

(一)當檢驗結果呈陽性時，個體實際罹患該疾病的條件機率 (即「陽性預測值」，Positive Predictive Value, PPV) 為何？請說明完整的計算過程，並給出具體數值結果。(10 分)

(二)在此例中，若政府大量採用此快速檢驗作為篩檢工具，雖具有快速與低成本的優點，但也可能導致高誤判率與高社會成本。請具體說明造成此現象的統計原因，並討論若採用此快速檢驗兩次 (兩次都陽性才認定為陽性)，是否能有效降低誤判與社會成本，並說明原因。(10 分)

三、政府想知道年資超過 10 年的計程車司機月收入是否高於 4.9 萬。隨機抽取年資超過 10 年的資深司機 20 位，月收入(萬元)如下(已由小到大)：

2.16, 2.54, 2.78, 3.62, 3.88, 4.26, 4.70, 4.80, 4.92, 5.00,
5.08, 5.54, 6.02, 6.12, 6.46, 6.66, 6.94, 7.76, 8.06, 11.00

請根據上述數據，以 $\alpha = 0.05$ 的顯著水準，討論並完成下列分析與檢定：

(一)若接受資料為近似常態，現欲採用母體平均數 μ 檢定：

$$H_0: \mu = 4.9, H_a: \mu > 4.9$$

直觀而言，我們以樣本平均 $\bar{X}$ 作為母體平均數 μ 的估計，因此可設 $\{\bar{X} > c\}$ 為拒絕域。請給出 c 的明確式子(包含樣本變異數 S^2 、樣本數以及分布分位數)，並說明其原理。接著，利用本資料判定此拒絕域是否成立，即是否拒絕 $H_0: \mu = 4.9$ 。(20 分)

(二)若資料分布明顯偏離常態，可考慮採中位數 m 檢定 (sign test)：

$$H_0: m = 4.9, H_a: m > 4.9$$

請說明此檢定所使用的統計量、 p 值的計算，並就本資料說明是否拒絕 H_0 。(20 分)

四、某公司欲了解其品牌產品的平均使用年限，隨機抽取樣本 7 件，其使用年限(單位：年)如：5, 7, 4, 4, 5, 4, 6

(一)請試著以不偏的樣本偏態係數 (skewness) 來說明該分布是否近似常態。(10 分)

(二)根據中央極限定理 (Central Limit Theorem, CLT)，說明在本例樣本數 $n=7$ 的情況下，若直接以常態近似來建立平均數信賴區間，會產生什麼問題？產生的問題可以怎麼處理？(10 分)

註：

若隨機變數 $T \sim t_{n-1}$ 分布，則 $(1-\alpha)$ 分位數 $t_{1-\alpha, n-1}$ 定義為 $P(T \leq t_{1-\alpha, n-1}) = 1-\alpha$ 。常見的數值有 $t_{0.95, 19} = 1.729$ ， $t_{0.975, 19} = 2.093$ ， $t_{0.95, 20} = 1.725$ ， $t_{0.975, 20} = 2.086$

若隨機變數 $Z \sim N(0, 1)$ 分布，則 $(1-\alpha)$ 分位數 $z_{1-\alpha}$ 定義為 $P(Z \leq z_{1-\alpha}) = 1-\alpha$ 。常見的數值有 $z_{0.95} = 1.645$ ， $z_{0.975} = 1.96$