

110學年度高等水文分析作業(二)

第一部分—非線性、多變數迴歸

個人作業請於11月17日(星期三) 3am以前將電子檔繳交至CEIBA

請第六、七組準備簡報11月18日8:10抽籤由一位同學第一節課報告作業成果

1. 利用峇里島某海灘三年的日最高溫 ($T^{\circ}C$)、日降雨量 ($R\text{ mm}$) 和遊客人數 (C 人數) 紀錄檔案為TRC.txt, 請建立海灘週末遊客人數迴歸模式, 以利未來可以根據氣象預報的溫度和降雨量, 預估海灘遊客人數。(15分)
 - A. 首先, 分別以日最高溫 ($T^{\circ}C$)、對數日最高溫 ($\ln T^{\circ}C$)、日降雨量+1後再取對數 ($\ln(R+1)\text{mm}$) 為橫軸, 對數遊客人數 (C) 為縱軸, 繪三張圖。(1分)
 - B. 擬用 $\ln \hat{C} = \beta_0 + \beta_1(\alpha - T) + \beta_2 \ln(R + \gamma)$ 的非線性迴歸模式, 推估遊客人數。最佳化方法是: 先給定 $-\infty < \alpha < \infty$ 、 $\gamma > 0$ 的數值, 求迴歸係數與 (對數) 誤差平方和, 再變動 α 、 γ , 得到所有不同 α 、 γ 數值組合中, (對數) 誤差平方和數值最小的最佳非線性迴歸方程式。(2分)
 - C. 同上小題, 但迴歸模式改為 $\ln \hat{C} = \beta_0 + \beta_1 \ln(T - \delta) + \beta_2 \ln(R + \phi)$, $\delta < t_{\min}$; 或 $\ln \hat{C} = \beta_0 + \beta_1 \ln(\delta' - T) + \beta_2 \ln(R + \phi)$, $\delta' > t_{\max}$ 、 $\phi > 0$ 。(3分)
 - D. 假設日最高溫為三參數對數常態分布, $X = \ln(\tau - T) \propto N[\mu_x, \sigma_x]$, 其中 τ 為上限參數, $\tau > t_{\max}$, 請嘗試不同的 τ 值, 使 $\ln(\tau - T)$ 與20個等機率區間的 χ^2 測試統計值為最小, 決定 τ 、 \bar{x} 、 σ_x 。(2分)
 - E. 假設非零日雨量經Box-Cox轉換後的機率密度分布函數為常態, 利用「偏態係數絕對值最小化」原則, 決定Box-Cox轉換的 λ 參數數值, $y = (R^\lambda - 1)/\lambda$, 以及 μ_y 、 σ_y 。(2分)
 - F. 計算D小題轉換後的 x_i 樣本和E小題轉換後的 y_i 樣本, 兩者的線性相關係數。(1分)
 - G. 擬利用假變數整合有雨、無雨的 x 、 y 變數, 估計 $\ln \hat{C}$, 請寫出線

性迴歸方程式 $\ln \hat{C} = \beta_0 + \beta_1 x + \dots$ ，建模估計迴歸係數。（2分）

H. 計算和比較以上B、C、G三個模式的以下多種估計誤差指標，決定最佳估計模式。指標包括：

- $\sum_{i=1}^{1095} (\ln \hat{C}_i - \ln C_i)^2$ ， $\sum_{i=1}^{1095} (\ln \hat{C}_i - \ln C_i)$ ， $\sum_{i=1}^{1095} |\ln \hat{C}_i - \ln C_i|$ ；
- $\sum_{i=1}^{1095} (\hat{C}_i - C_i)^2$ ， $\sum_{i=1}^{1095} (\hat{C}_i - C_i)$ ， $\sum_{i=1}^{1095} |\hat{C}_i - C_i|$ ；（計算人數估計誤差時，估計值 \hat{C} 不需四捨五入為整數）
- 補償平均值估計的 $\sum_{i=1}^{1095} (\hat{C}_i^s - C_i^s)^2$ 、 $\sum_{i=1}^{1095} (\hat{C}_i^s - C_i^s)$ 、 $\sum_{i=1}^{1095} |\hat{C}_i^s - C_i^s|$

評估三個非線性迴歸模式估計，和對應的「補償平均值估計」法的優劣特性。（2分）