

數據科學方法期中考

1. (56 分) 是非題。答是請寫理由, 答非請寫理由或舉反例。只有答案不給分。
 - (a) 在其他條件相同的狀況下, 信賴區間的信賴水準增加時, 區間會變寬
 - (b) 在某個固定的信心水準下, 樣本大小增加時, 犯型 1 錯誤的機率也會增加
 - (c) 若事件 A 的發生會增加事件 B 發生的機率, 則事件 B 的發生也會增加事件 A 發生的機率
 - (d) 如果兩個事件為互補事件, 則其中一個事件會發生的機率為 1
 - (e) p -value 即為拒絕 H_0 時會犯錯的機率
 - (f) p -value 越小代表假設檢定的結果越顯著
 - (g) 流行病學研究結果顯示 $\text{pm} 2.5$ 濃度越高的地方心臟病的發生率也會隨之提高, 因此我們可以推斷 $\text{pm} 2.5$ 為心臟病的致病因子

2. 科學家相信風速資料服從 Weibull distribution。請利用氣象局“新竹市東”自動氣象站之風速資料
 - (a) (20 分) 利用 MLE 或 MoM 估計 Weibull distribution 的未知參數 λ 與 k
 - (b) (15 分) 利用 Kolmogorov-Smirnov test 檢定該資料是否來自上題估計出的 Weibull distribution
 - (c) (20 分) 利用 bootstrapping 建立參數 λ 與 k 各自在 95% 信賴水準時的信賴區間