

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 1 頁，共 12 頁

單選題 50 題 (佔 100%)

C	1. 關於 R 語言之特殊值，下列何者「不」正確？ (A) Code：「NA == NULL」；Result：logical(0) (B) Code：「NA == NA」；Result：NA (C) Code：「NULL == NULL」；Result：NULL (D) Code：「NA == FALSE」；Result：NA														
D	2. 關於「資料清洗」的主要目的，下列何者正確？ (A) 讓資料量縮小 (B) 清理過的資料會導致模型無法順利運行 (C) 讓資料無法透過數學運算 (D) 讓計算及分析上，更為方便及降低偏誤														
A	3. 關於資料分析的流程，通常為下列何者？ (A) 資料載入→資料處理→資料分析 (B) 資料載入→資料分析→資料處理 (C) 資料分析→資料載入→資料處理 (D) 資料處理→資料分析→資料載入														
B	4. 關於離群值 (Outlier)，下列敘述何者正確？ (A) 空字串 (B) 在資料中有一個或幾個數值與其他數值差異較大 (C) 編碼錯誤 (D) 遺漏的值														
B	5. 附圖六名男女的身高體重，請問***遺缺的部分，如果以「平均法」來填補，應為何者？ <table border="1"><thead><tr><th>性別</th><th>身高</th></tr></thead><tbody><tr><td>男</td><td>160</td></tr><tr><td>男</td><td>170</td></tr><tr><td>男</td><td>***</td></tr><tr><td>女</td><td>160</td></tr><tr><td>女</td><td>170</td></tr><tr><td>女</td><td>180</td></tr></tbody></table> (A) 163 (B) 165 (C) 168 (D) 172	性別	身高	男	160	男	170	男	***	女	160	女	170	女	180
性別	身高														
男	160														
男	170														
男	***														
女	160														
女	170														
女	180														
D	6. 老闆想看過去十二個月的業績報告，請問使用下列何種圖形最「不」適合？ (A) 圓餅圖 (Pie Chart)														

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 2 頁，共 12 頁

	(B) 長條圖 (Bar Chart) (C) 折線圖 (Line Chart) (D) 樹狀圖 (Tree Diagram)																					
C	<p>7. 附圖為水果店的今日銷售報表，如果想知道每個水果佔總銷售數量百分比，下列 SQL 查詢語句何者正確？</p> <table><tr><th>Userid</th><th>品項</th><th>銷售數量</th></tr><tr><td>1</td><td>蘋果</td><td>2</td></tr><tr><td>2</td><td>芭樂</td><td>3</td></tr><tr><td>3</td><td>香蕉</td><td>1</td></tr><tr><td>4</td><td>蘋果</td><td>1</td></tr><tr><td>5</td><td>蘋果</td><td>2</td></tr><tr><td>6</td><td>芭樂</td><td>1</td></tr></table> <p>(A) select count(銷售數量) as 數量 / (select count(*) from data) as 百分比 from data</p> <p>(B) select 品項, count(銷售數量) as 百分比 from data group by 品項</p> <p>(C) select 品項, count(銷售數量) as 數量 / (select count(*) from data) as 百分比 from data group by 品項</p> <p>(D) select 品項, count(銷售數量) as 數量 / (select count(*) from data) as 百分比 from data</p>	Userid	品項	銷售數量	1	蘋果	2	2	芭樂	3	3	香蕉	1	4	蘋果	1	5	蘋果	2	6	芭樂	1
Userid	品項	銷售數量																				
1	蘋果	2																				
2	芭樂	3																				
3	香蕉	1																				
4	蘋果	1																				
5	蘋果	2																				
6	芭樂	1																				
D	<p>8. 下列何者「不」是實數型資料常見的彙整方式？</p> <p>(A) 平均值 (Mean)</p> <p>(B) 標準差 (Standard Deviation)</p> <p>(C) 最大值 (Maximum)、最小值 (Minimum)</p> <p>(D) 可能值條列 (Possible Value List)</p>																					
A	<p>9. 關於常用的探索性資料分析，下列何者「不」是常用的圖形化方法？</p> <p>(A) 甘特圖 (Gantt Chart)</p> <p>(B) 盒鬚圖 (Box Plot)</p> <p>(C) 趨勢圖 (Run Chart)</p> <p>(D) 柏拉圖 (Pareto Chart)</p>																					

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 3 頁，共 12 頁

C	<p>10. 參考附圖 Python 語言 pandas 套件，關於 describe 函數，下列敘述何者「不」正確？</p> <pre>In [1]: import pandas as pd: import numpy as np: df = pd.DataFrame({: 'sampleID': ['A01', 'A02', 'A03', 'A04'],: 'status': [2, 1, np.nan, 9],: 'measure': [3, 5, 9, 8],: }): df.describe()</pre> <p>Out[1]:</p> <table><thead><tr><th></th><th>status</th><th>measure</th></tr></thead><tbody><tr><td>count</td><td>3.000000</td><td>4.000000</td></tr><tr><td>mean</td><td>4.000000</td><td>6.250000</td></tr><tr><td>std</td><td>4.358899</td><td>2.753785</td></tr><tr><td>min</td><td>1.000000</td><td>3.000000</td></tr><tr><td>25%</td><td>1.500000</td><td>4.500000</td></tr><tr><td>50%</td><td>2.000000</td><td>6.500000</td></tr><tr><td>75%</td><td>5.500000</td><td>8.250000</td></tr><tr><td>max</td><td>9.000000</td><td>9.000000</td></tr></tbody></table> <p>(A) df 資料集有 4 筆資料，3 個變數欄位 (B) status 欄位 75 百分位數為 5.500000 (C) df 資料集沒有 NA 值 (D) describe 函數加上 include='all' 參數，會顯示所有欄位資料摘要</p>		status	measure	count	3.000000	4.000000	mean	4.000000	6.250000	std	4.358899	2.753785	min	1.000000	3.000000	25%	1.500000	4.500000	50%	2.000000	6.500000	75%	5.500000	8.250000	max	9.000000	9.000000
	status	measure																										
count	3.000000	4.000000																										
mean	4.000000	6.250000																										
std	4.358899	2.753785																										
min	1.000000	3.000000																										
25%	1.500000	4.500000																										
50%	2.000000	6.500000																										
75%	5.500000	8.250000																										
max	9.000000	9.000000																										
C	<p>11. 假設 A 商店的客戶年齡為 10, 16, 20, 22, 25, 27, 31, 35, 50。如果我們使用等寬/距（Equal Width/Distance）分割法將客戶切割為四個年齡層，則一位 32 歲的客戶應該屬於第幾個年齡級距？</p> <p>(A) 第一級距 (B) 第二級距 (C) 第三級距 (D) 第四級距</p>																											
D	<p>12. 下列何者「不」是常見的屬性萃取（Feature Extraction）方法？</p> <p>(A) 主成分分析（Principle Component Analysis） (B) 獨立成分分析（Independent Component Analysis） (C) 潛在語意分析（Latent Semantic Analysis） (D) 數值分析（Numerical Analysis）</p>																											
A	<p>13. 假設一個三維資料分布如同瑞士卷的資料集，下列何種方法能較為完善在降維至二維後，依然保持其在三維空間的資料特性？</p> <p>(A) 等距特徵映射（Isometric Feature Mapping, Isomap） (B) 主成分分析（Principle Component Analysis） (C) 線性判別分析（Linear Discriminant Analysis）</p>																											

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 4 頁，共 12 頁

	(D) 獨立成分分析 (Independent Component Analysis)
A	14. 下列何者屬於屬性挑選 (Feature Selection) 的方法？ (A) 逐步迴歸法 (Stepwise Regression) (B) 奇異值分解 (Singular Value Decomposition) (C) 線性判別分析 (Linear Discriminant Analysis) (D) 多元尺度分析 (Multi-Dimensional Analysis)
D	15. 假設您有一組懸浮微粒濃度的資料，若想依照不同等級來辨別空氣污染的程度，須將該組資料轉化為下列何種屬性的資料？ (A) 離散資料 (Discrete) (B) 連續資料 (Continuous) (C) 量化資料 (Quantitative) (D) 類別資料 (Category)
D	16. 關於比較 Hadoop 與 Spark 的差異，下列何者正確？ (A) Hadoop：可透過「hive」以類 SQL 語句操作資料；Spark：目前仍沒有任何工具能以類 SQL 語句操作資料 (B) Hadoop 適合存儲大量的小型檔案 (<1MB)，以妥善利用硬碟空間 (C) Spark 主要是採用硬體內運算技術 (In-disk Analytics) 將運算過程中暫存資料存於硬碟內，因此運算速度比 Hadoop 快 (D) Spark 提供了 DataFrame 資料物件格式，具有良好的資料操作性
D	17. 若希望快速檢查一個元素是否為巨量資料集合中的一員，可以使用下列何種資料結構存放這巨量資料集合？ (A) 堆疊 (Stack) (B) 協同過濾法 (Collaborative Filtering) (C) 佇列 (Queue) (D) 布隆過濾器 (Bloom Filter)
D	18. 為確保巨量資料處理平台上的資訊安全防護，下列做法何者較「不」恰當？ (A) 將巨量資料處理平台放置於獨立網段，登入方法可透過 ACL (Access Control List) 設定或遠端桌面進行跳板連線 (B) 啟用 Kerberos、Ranger、Sentry 等服務設定與驗證，確保連線安全 (C) 採用 RADIUS 服務驗證機制，除了 LDAP (Lightweight Directory Access Protocol) 以外，還包含 PAP (Password Authentication Protocol)、CHAP (Challenge Handshake Authentication Protocol) 與本地端系統密碼 (D) 文件傳輸可採用 FTP 密碼驗證，與複雜的密碼設計

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 5 頁，共 12 頁

C	19. 關於 MapReduce，下列何者「不」是 Map 的操作？ (A) $[1,2,3,4] - (x*2) \rightarrow [2,4,6,8]$ (B) $[1,2,3,4] - (x+2) \rightarrow [3,4,5,6]$ (C) $[1,2,3,4] - (\text{sum}) \rightarrow 10$ (D) $[1,2,3,4] - (3x+1) \rightarrow [4,7,19,33]$
B	20. 關於 Hadoop 分散式檔案系統 (Hadoop Distributed File System, HDFS)，下列敘述何者「不」正確？ (A) HDFS 透過 NameNode 節點與 Client 端進行需求確認，以及管理整體檔案的存寫動作 (B) HDFS 無法處理超過 1TB 大小以上的檔案 (C) HDFS 可部署在硬體設備相對較低的環境上以提供服務 (D) HDFS 可透過 NameNode 和 DataNode 節點儲存資料
D	21. 下列何種情形適合使用單因子變異數分析 (One-way Analysis of Variance)？ (A) 檢驗數據是否服從常態分配 (B) 比較某班級男生與女生數學成績的變異數 (C) 比較兩間輪胎工廠，輪胎平均使用年限是否不同 (D) 比較某工廠 4 部機器由不同人員操作下，其每小時平均產量是否不同
C	22. 班上有 50 位學生，座號為 01 至 50。今要找 5 位同學參加比賽，經討論後決定以簡單隨機抽樣法選取人選，若以附圖亂數表之第一列第一行為起始點由左至右選取，則會抽到哪些座號？ 2 6 4 2 8 4 1 4 9 3 2 5 7 9 0 8 9 8 3 7 5 7 1 7 6 7 8 8 0 7 0 7 9 5 6 2 3 9 0 3 9 8 4 3 5 9 9 4 6 6 1 8 6 9 4 5 4 9 3 3 9 0 9 5 8 0 8 0 5 9 5 0 6 4 8 5 5 2 5 8 2 7 0 9 4 6 7 2 7 4 6 1 1 1 0 8 6 4 6 2 (A) 02, 06, 04, 28, 41 (B) 26, 42, 04, 14, 32 (C) 26, 42, 14, 25, 08 (D) 26, 42, 41, 49, 32
B	23. 下列調查結果何者屬於推論統計？ (A) 由於政府去年鐵腕施行掃毒行動，今年上半年的犯罪率較去年同期減少了 5 個百分點 (B) 由於台灣地區鰻魚生產過剩，我們預計今年外銷鰻魚的平均價格將跌到每台斤不到 100 元台幣

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 6 頁，共 12 頁

	<p>(C) 由於全球金融風暴，台灣地區今年一月份的失業率較去年同期上升了 0.6 個百分點</p> <p>(D) 因應政府拯救房市方案，今年初各大銀行紛紛調降房屋貸款利率，平均利率下降了 2 個百分點</p>																				
A	<p>24. 在一個次數分佈中，樣本數為 10，平均值為 90，標準差為 10。如果在總體中再加上 80、70、60 這三筆資料，則重新計算的平均值將如何變化？</p> <p>(A) 變小</p> <p>(B) 變大</p> <p>(C) 不變</p> <p>(D) 無法判斷</p>																				
A	<p>25. 若某廠牌電池壽命為一常態分配，且已知平均數為 100 小時，標準差為 4 小時。今隨機抽檢一電池，若 Z 為標準常態分配，則下列何者正確？</p> <p>(A) 電池壽命超過 110 小時的機率為 $P(Z>2.5)$</p> <p>(B) 電池壽命超過 100 小時的機率為 $P(Z>1)$</p> <p>(C) 電池壽命介於 100~110 間的機率為 $P(0<Z<2.5)$</p> <p>(D) 電池壽命超過 100 小時的機率為 $P(Z>4)$</p>																				
B	<p>26. 關於資料屬性，下列何者與其他不同？</p> <p>(A) 宗教</p> <p>(B) 身高</p> <p>(C) 性別</p> <p>(D) 國家</p>																				
B	<p>27. 請問附圖中兩變數 A 與 B 的相關係數為何？</p> <table border="1"><tr><td>A</td><td>1</td><td>2</td><td>3</td><td>4</td><td>...</td><td>97</td><td>98</td><td>99</td><td>100</td></tr><tr><td>B</td><td>100</td><td>99</td><td>98</td><td>97</td><td>...</td><td>4</td><td>3</td><td>2</td><td>1</td></tr></table> <p>(A) 1</p> <p>(B) -1</p> <p>(C) 0.5</p> <p>(D) -0.5</p>	A	1	2	3	4	...	97	98	99	100	B	100	99	98	97	...	4	3	2	1
A	1	2	3	4	...	97	98	99	100												
B	100	99	98	97	...	4	3	2	1												
A	<p>28. 附圖為海藻資料集的變數 Chla 的直方圖，請問其平均數與中位數最可能為何者？</p>																				

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 7 頁，共 12 頁

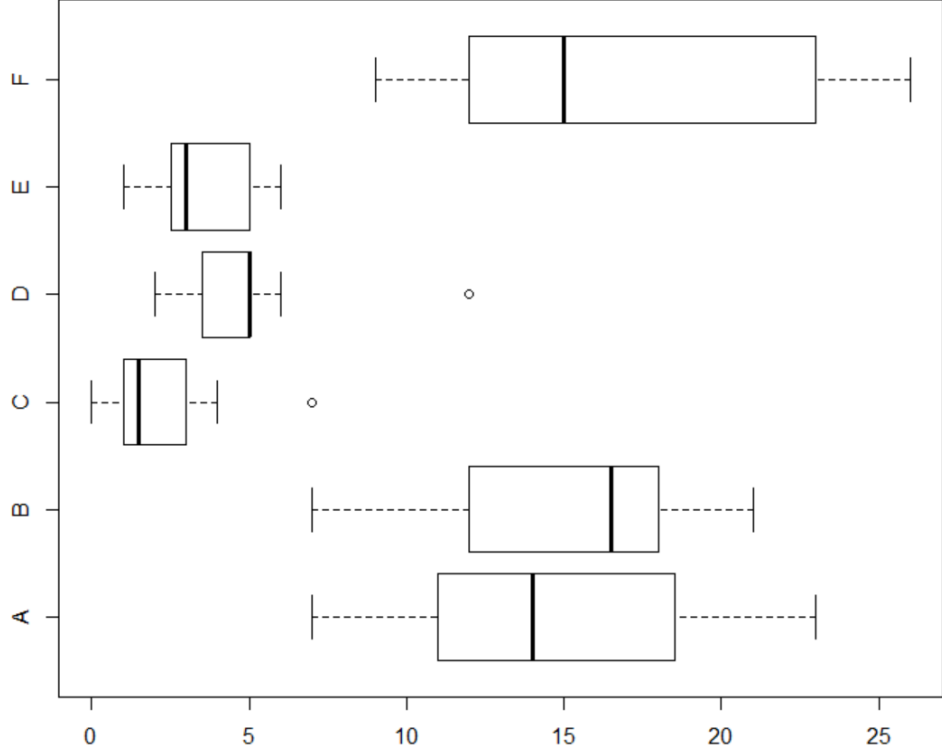
	<p>Histogram of Chla value</p> <p>(A) Mean: 5.5 ; Median: 14 (B) Mean: 5.5 ; Median: 60 (C) Mean: 14 ; Median: 14 (D) Mean: 14 ; Median: 5.5</p>
A	<p>29. 令(X, Y)為二元隨機變數，期望值分別 μ_x 與 μ_y，(X, Y)之共變異數以 $Cov(X, Y)$表示，則下列定義何者正確？</p> <p>(A) $Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$ (B) $Cov(X, Y) = E[(X + \mu_x)(Y + \mu_y)]$ (C) $Cov(X, Y) = E[(\mu_x - \mu_y)]$ (D) $Cov(X, Y) = E[(\mu_x + \mu_y)]$</p>
A	<p>30. 在進行假設檢定的過程中，我們會先將結果分成兩種相反的決策：虛無假設 (Null Hypothesis, H_0) 和對立假設 (Alternative Hypothesis, H_1)。當虛無假設 H_0 為真而拒絕 H_0，稱之為何種錯誤？</p> <p>(A) 型 I 錯誤 (B) 型 II 錯誤 (C) 型 III 錯誤 (D) 型 IV 錯誤</p>
D	<p>31. 關於鄰近函數 (Proximate Function) 中心點的計算，下列何者「不」是利用資料的平均值？</p> <p>(A) 餘弦 (Cosine) (B) 平方歐幾里德 (Squared Euclidean Distance) (C) Bregman 分散度 (Bregman Divergence) (D) 曼哈頓 (Manhattan Distance)</p>
B	<p>32. 下列何種演算法可以將一張高解析度圖片壓縮成較低解析度的圖片？</p>

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 8 頁，共 12 頁

	<p>(A) 支援向量機 (Support Vector Machine, SVM)</p> <p>(B) 奇異值分解 (Singular Value Decomposition, SVD)</p> <p>(C) K 平均法 (K-Means)</p> <p>(D) 最小平方法 (Ordinary Least Square, OLS)</p>
B	<p>33. 參考附圖的結果，其繪圖 R 指令為何？</p>  <p>(A) <code>boxplot(count ~ spray, data = InsectSprays)</code></p> <p>(B) <code>boxplot(count ~ spray, data = InsectSprays, horizontal = TRUE)</code></p> <p>(C) <code>hist(count ~ spray, data = InsectSprays)</code></p> <p>(D) <code>hist(count ~ spray, data = InsectSprays, horizontal = TRUE)</code></p>
D	<p>34. 參考附圖，關於此莖葉圖 (Stem and Leaf Plot)，下列敘述何者「不」正確？</p> <pre> 4 7 5 122445 6 22 7 4899 8 3445558 </pre> <p>(A) 資料呈現非常態分配</p> <p>(B) 最小值為 47</p>

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 9 頁，共 12 頁

	<p>(C) 最大值是 88</p> <p>(D) 眾數是 62</p>
C	<p>35. 關於 K 摺 (K-folds) 法，下列敘述何者「不」正確？</p> <p>(A) 初始將資料隨機區分成 K 個子樣本</p> <p>(B) 一個單獨的子樣本被保留作為測試模型，其他 K-1 個樣本為訓練模型</p> <p>(C) 整個 K 摺法會重複(K+1)-次</p> <p>(D) 一般可採用 K=10 進行 K 摺法</p>
A	<p>36. R 語言中，計算資料百分位數的指令為下列何者？</p> <p>(A) quantile</p> <p>(B) percent</p> <p>(C) median</p> <p>(D) sum</p>
A	<p>37. 參考附圖，R 語言中，關於群集分析的執行結果，下列敘述何者「不」正確？</p> <pre> within cluster sum of squares by cluster: [1] 180641.38 253143.30 54831.62 392747.49 65818.27 (between_SS / total_SS = 12.3 %) Available components: [1] "cluster" "centers" "totss" "withinss" "tot.withinss" [6] "betweenss" "size" "iter" "ifault" > teen.clusters\$betweenss [1] 132781.9 > </pre> <p>(A) 群集個數為 4</p> <p>(B) 群集的總距離變異平方和為 1079964</p> <p>(C) 群集間的總距離變異平方和為 132781.9</p> <p>(D) 群集內的總距離變異平方和為 947182.1</p>
D	<p>38. 關於 Apriori 演算法，下列敘述何者「不」正確？</p> <p>(A) 採用頻繁項目集 (Frequent Itemsets) 的布林值關聯規則方法</p> <p>(B) 計算每一個候選項目集的數目</p> <p>(C) 依據事先設定的最小支持度，以衡量候選項目集的關聯規則是否顯著</p> <p>(D) 如果某候選項目集為頻繁項目集，則其所有子集合一定不是頻繁項目集</p>
C	<p>39. 關於頻繁模式增長 (Frequent-Pattern Growth) 演算法，下列敘述何者「不」正確？</p> <p>(A) 可應用於關聯規則分析</p> <p>(B) 將整個交易中包括頻繁項目集 (Frequent Itemsets) 壓縮至頻繁模式樹 (Frequent-Pattern Tree)</p>

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 10 頁，共 12 頁

	(C) 最多只掃描資料庫一次，可大量減少輸入/輸出 (Input/Output) 時間 (D) 頻繁模式樹僅儲存頻繁項目集 L_1
A	40. 非監督式學習 (Unsupervised Learning) 和監督式學習 (Supervised Learning) 最大的差異為何？ (A) 依變項的有無 (B) 依變項是否為類別變項 (C) 自變項是否為數值型變項 (D) 自變項的多寡
C	41. 關於決策樹 (Decision Tree)，下列何者「不」是其包含三種類型的節點？ (A) 根節點 (Root Node) (B) 內部節點 (Internal Node) (C) 枝節點 (Branch Node) (D) 葉節點 (Leaf Node)
A	42. 進行線性迴歸 (Linear Regression) 時，通常使用何種方法進行模型配適？ (A) 最小平方法 (B) 移動平均法 (C) 歐式距離法 (D) 最大平方法
B	43. 關於羅吉斯迴歸 (Logistic Regression) 與線性迴歸 (Linear Regression) 的比較，下列何者正確？ (A) 羅吉斯迴歸與線性迴歸，都屬於非監督式學習 (Unsupervised Learning) (B) 線性迴歸的輸出是連續的數值 (C) 羅吉斯迴歸常使用於預測房屋價格等數值問題 (D) 線性迴歸常使用於預測分類問題
A	44. 請問在一個接收者操作特徵 (Receiver Operating Characteristic, ROC) 曲線圖中，該圖的 X 軸與 Y 軸分別為何？ (A) X: False Positive Rate ; Y: True Positive Rate (B) X: Precision ; Y: Recall (C) X: Sensitivity ; Y: Specificity (D) X: Detection Rate ; Y: Accuracy
B	45. 下列何者較「不」適合使用線性迴歸 (Linear Regression) 模型作為預測？

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 11 頁，共 12 頁

	<p>(A) 用身高預測高中生的體重</p> <p>(B) 使用球員薪資、投籃數、以及投籃命中率來預測 NBA 總冠軍的機率</p> <p>(C) 使用居住區域、人口密度、以及人口薪資水平來預測房價</p> <p>(D) 使用教育程度、工作年資、以及年齡來預測收入</p>																												
B	<p>46. 下列何者為非線性迴歸 (Non-linear Regression) 之使用時機？</p> <p>(A) 自變數的值成非線性分佈時</p> <p>(B) 依變數的值成非線性分佈時</p> <p>(C) 用來預測的自變項超過一個時</p> <p>(D) 要預測的依變項超過一個時</p>																												
D	<p>47. 關於決策樹學習的過程，下列何者「不」正確？</p> <p>(A) 特徵選擇</p> <p>(B) 決策樹生成</p> <p>(C) 決策樹剪枝</p> <p>(D) 決策樹的生成考慮全局最優，而決策樹的剪枝考慮局部最優</p>																												
C	<p>48. 假設一個六面骰子，擲出各點數的機率如附圖，為各點數建立霍夫曼編碼 (Huffman Coding)。若對應的霍夫曼樹 (Huffman Tree) 中，節點左子樹機率小於右子樹，且左連結編碼為 0，右連結編碼為 1，則點數 5 的編碼為？</p> <table><tr><td>點數</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>機率</td><td>1/63</td><td>2/63</td><td>4/63</td><td>8/63</td><td>16/63</td><td>32/63</td></tr></table> <p>(A) 101</p> <p>(B) 00</p> <p>(C) 01</p> <p>(D) 1010</p>	點數	1	2	3	4	5	6	機率	1/63	2/63	4/63	8/63	16/63	32/63														
點數	1	2	3	4	5	6																							
機率	1/63	2/63	4/63	8/63	16/63	32/63																							
B	<p>49. 附圖為學生讀書時間與成績的關係。假若有一位新學員的評價一與評價二的資料為 3 及 5，使用 KNN 演算法，K 值為 3，請問其評價結果為何？</p> <table><tr><td></td><td>評價一</td><td>評價二</td><td>評價結果</td></tr><tr><td>學員一</td><td>1</td><td>1</td><td>A</td></tr><tr><td>學員二</td><td>2</td><td>2</td><td>A</td></tr><tr><td>學員三</td><td>3</td><td>3</td><td>B</td></tr><tr><td>學員四</td><td>5</td><td>5</td><td>B</td></tr><tr><td>學員五</td><td>6</td><td>6</td><td>C</td></tr><tr><td>學員六</td><td>7</td><td>7</td><td>C</td></tr></table>		評價一	評價二	評價結果	學員一	1	1	A	學員二	2	2	A	學員三	3	3	B	學員四	5	5	B	學員五	6	6	C	學員六	7	7	C
	評價一	評價二	評價結果																										
學員一	1	1	A																										
學員二	2	2	A																										
學員三	3	3	B																										
學員四	5	5	B																										
學員五	6	6	C																										
學員六	7	7	C																										

110 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：110 年 11 月 20 日

第 12 頁，共 12 頁

	<p>(A) A</p> <p>(B) B</p> <p>(C) C</p> <p>(D) 無法評價</p>
D	<p>50. 非監督式學習（Unsupervised Learning）中，K-means 是概念簡單卻實用的分群演算法之一。關於 K-means 方法，下列敘述何者「不」正確？</p> <p>(A) K 值的設定不同，將會影響到分群的結果</p> <p>(B) 同群內的相似度大，群與群的相似度小</p> <p>(C) 隨著 K-means 方法演算的迭代，會不斷計算更新各群的中心位置直到收斂</p> <p>(D) 適合處理有方向性的資料，如螺旋形點分佈、線性點分佈等資料狀況</p>