



企業電子化人才能力鑑定

# 巨量資料分析與處理

110/111 雲端指南

( 內附試題彙編 )



中華民國電腦技能基金會  
Computer Skills Foundation

## 企業電子化人才能力鑑定

2021/1/1

DPA

「巨量資料處理與分析」學科試題及答案

## 第一類：資料科學與巨量資料

- ABC      1-01.      下列關於 Hadoop 生態圖譜系統的敘述，哪些為真？(複選)
- ( A ) HDFS：分散式檔案儲存系統
  - ( B ) MapReduce：資料分散式計算框架
  - ( C ) Sqoop：與關聯式資料庫資料交換
  - ( D ) HBase：關聯式資料庫
- A      1-02.      HDFS 會先將原始檔案切割成多個區塊 (block)，並將每一個區塊複製三份，分別存入三個不同的節點，假設每一個區塊的大小是 128MB，若今日要將一個 256MB 的檔案放入 HDFS，請問會產生幾個區塊？使用了 HDFS 多少空間？
- ( A ) 6 個區塊，使用了 768MB
  - ( B ) 2 個區塊，使用了 256MB
  - ( C ) 3 個區塊，使用了 384MB
  - ( D ) 4 個區塊，使用了 512MB
- ABCD      1-03.      下列哪些是跨產業資料探勘標準作業程序 (CRISP-DM) 的流程？(複選)
- ( A ) 商業理解、資料理解
  - ( B ) 資料準備
  - ( C ) 模式建立
  - ( D ) 評估與測試
- BD      1-04.      關於資料探勘流程，下列哪些敘述正確？(複選)
- ( A ) 資料探勘通常是一次性任務，一般不需要重複循環這六個步驟
  - ( B ) 資料探勘一般需要六個流程：商業理解、資料理解、資料準備、建模、評估、發佈
  - ( C ) 資料探勘項目只需要資料探勘工程師即可完成
  - ( D ) 資料探勘的結果評估需要從商業價值方面考慮

### 第三章 試題彙編

- B      1-05.      根據 IBM 的定義，下列哪一項不是巨量資料 (Big Data) 的特性？
- ( A ) Volume
  - ( B ) Visualization
  - ( C ) Variety
  - ( D ) Veracity
- A      1-06.      GB, PB, TB, ZB 為四種電腦記憶體容量的單位，請問它們的大小排序為何？
- ( A ) ZB > PB > TB > GB
  - ( B ) PB > TB > ZB > GB
  - ( C ) TB > ZB > GB > PB
  - ( D ) PB > ZB > GB > TB
- ABC      1-07.      下列哪些是量化模型的特質？(複選)
- ( A ) 模型是現實的抽象表徵
  - ( B ) 所有的模型無法完全代表真相，但總有一些模型是有用的
  - ( C ) 有效的模型必須正確地表達欲研究事物的相關特質
  - ( D ) 不需透過電腦進行實作
- ACD      1-08.      2001 年由 Gartner 提出巨量資料 (Big Data) 概念包括哪些特性？(複選)
- ( A ) Volume
  - ( B ) Verification
  - ( C ) Velocity
  - ( D ) Variety
- C      1-09.      Hadoop 生態圖譜系統中執行資料分散式計算框架為哪一技術？
- ( A ) HDFS
  - ( B ) HBase
  - ( C ) MapReduce
  - ( D ) Hive

## 企業電子化人才能力鑑定

2021/1/1

- B 1-10. Hadoop 生態圖譜系統中常用哪一種資料庫系統？  
（ A ） Mahout  
（ B ） HBase  
（ C ） MapReduce  
（ D ） HDFS
- D 1-11. CRISP-DM 資料探勘流程包括下列六項，其正確流程步驟為？（1）資料準備、（2）資料理解、（3）評估與測試、（4）模式建立、（5）商業理解、（6）部署應用  
（ A ） 521346  
（ B ） 512436  
（ C ） 512346  
（ D ） 521436
- B 1-12. 應用統計學的範疇包括敘述統計學與哪一項目？  
（ A ） 資料探勘  
（ B ） 推論統計學  
（ C ） 機器學習  
（ D ） 作業研究
- B 1-13. Big Data 顧名思義形容資料很 "大"， 以下哪一項不是描述其 "大" 的主要特徵？  
（ A ） 資料總量很大  
（ B ） 資料計算的時間很長  
（ C ） 資料產生的速度很快  
（ D ） 資料的來源很多
- BD 1-14. 下列哪些不是 Hadoop 的基本組件？(複選)  
（ A ） 檔案存儲系統 HDFS  
（ B ） 資料分析程式庫 Mahout  
（ C ） 分散式計算框架 MapReduce  
（ D ） 資料存儲系統 HIVE

## 第三章 試題彙編

- ABD      1-15.      在資料準備時，下列描述哪些正確？(複選)
- (A) 資料準備時，經過資料整合、清理、轉換、減少等步驟架構良好的資料
  - (B) 資料整合包括搜集資料、選擇資料、整合資料
  - (C) 資料清理包括減少變數數目、消除不一致、平衡偏斜資料
  - (D) 資料轉換包括正常化資料、分散／整合資料、建構新屬性
- D      1-16.      以下選項中，哪一項不是資料科學的基礎組成部分？
- (A) 領域知識
  - (B) 電腦科學
  - (C) 統計學
  - (D) 演算邏輯
- B      1-17.      以下哪一項不是統計學與電腦科學在商業資料分析中的差異？
- (A) 統計學通常以資料為導向，電腦科學通常以方法為導向
  - (B) 統計學分析樣本資料，電腦科學分析全量資料
  - (C) 統計學專注於分析數值與類別資料，電腦科學還能分析自然語言、圖像影音等資料
  - (D) 統計學偏重於假設和模型，電腦科學偏重於演算和程式
- C      1-18.      美國某百貨根據客戶消費資訊分析出星期五的下班時段購買尿布的消費者多會也購買啤酒，因此其改變了貨架的擺放位置，把啤酒和尿布擺在一起，甚至把比較貴的尿布放在啤酒旁邊，成功增加了這兩項商品的銷售金額。請問上述情境屬於商業問題中資料科學任務的何者？
- (A) 相似度分析 (similarity matching)
  - (B) 分類及類別的機率估計 (classification)
  - (C) 共同出現群組 (co-occurrence or affinity grouping)
  - (D) 集群 (clustering)
- B      1-19.      跨產業資料探勘標準作業流程 (Cross Industry Standard Process for Data Mining) 的步驟順序為何？(1) 資料理解、(2) 商業理解、(3) 評估與測試、(4) 資料準備、(5) 模式建立、(6) 部署應用
- (A) (1) → (2) → (4) → (5) → (6) → (3)
  - (B) (2) → (1) → (4) → (5) → (3) → (6)
  - (C) (2) → (4) → (1) → (5) → (3) → (6)
  - (D) (1) → (4) → (2) → (5) → (6) → (3)

## 企業電子化人才能力鑑定

2021/1/1

- ABD      1-20.      2001 年 Gartner 提出 Big Data 定義之 3 個 V 分別為何？  
(複選)  
    (A) Volume  
    (B) Velocity  
    (C) Vector  
    (D) Variety
- C      1-21.      以下何者不屬於 Hadoop 生態圖譜系統的一部分？  
    (A) MapReduce  
    (B) HDFS  
    (C) VMWare  
    (D) HBase
- A      1-22.      資料採礦社群網站 KDNuggets 對使用者所做的調查中，以下何者  
不是處理 Big data 的常見軟體工具？  
    (A) Word  
    (B) SQL  
    (C) R  
    (D) Python
- D      1-23.      商業問題中，希望能將顧客群中屬性相似的顧客聚集在一起，但事  
先並不清楚各群的樣貌時，資料科學家通常可以採用以下何種分析  
方法？  
    (A) 分類及類別的機率估計 (classification)  
    (B) 迴歸 (regression)  
    (C) 相似度分析 (similarity matching)  
    (D) 集群 (clustering)
- ABD      1-24.      目前下列哪些軟體工具在處理 Big Data 時較為常見？(複選)  
    (A) SQL  
    (B) R  
    (C) Julia  
    (D) Python

### 第三章 試題彙編

BCD

1-25.

以下何者屬於跨產業資料探勘標準作業流程（Cross Industry Standard Process for Data Mining）的一部分？（複選）

- （A）問卷調查
- （B）商業理解
- （C）資料準備
- （D）模式建立

DPA

「巨量資料處理與分析」學科試題及答案

## 第二類：資料處理的基礎

- D            2-01.        下列何者非為 SQL 語言格式？  
                  ( A ) 資料定義語言 (DDL)  
                  ( B ) 資料操作語言 (DML)  
                  ( C ) 資料控制語言 (DCL)  
                  ( D ) 資料分析語言 (DAL)
- ACD        2-02.        下列對主鍵 (Primary Key) 的敘述， 哪些為真？(複選)  
                  ( A ) 不考慮複合主鍵的情況下，該表格中該欄位內的值不得重複  
                  ( B ) 主鍵只能包含一個欄位  
                  ( C ) 包含多個欄位的主鍵，稱為複合主鍵 (Composite Key)  
                  ( D ) 一張表內僅能有一個主鍵
- A            2-03.        在 SQL 語言中，INSERT 指令的功用為何？  
                  ( A ) 新增資料  
                  ( B ) 修改資料  
                  ( C ) 刪除資料  
                  ( D ) 沒有此指令
- C            2-04.        如附圖為一 SQL 語法敘述，試問別名 a, b 分別指哪一張表格？  
                  (請參閱附圖作答)
- ```
Select a.empno, a.ename, a.sal, a.deptno, b.deptno, b.dname  
From emp a join dept b  
on a.deptno =b.deptno;
```
- ( A ) a 是 dept 資料表的別名、b 是 emp 資料表的別名  
                  ( B ) a 是 empno 資料表的別名、b 是 deptno 資料表的別名  
                  ( C ) a 是 emp 資料表的別名、b 是 dept 資料表的別名  
                  ( D ) a 是 sal 資料表的別名、b 是 dname 資料表的別名



### 第三章 試題彙編

B

2-05.

附圖中之 SQL 語法敘述，下列選項何者最貼近全意？（請參閱附圖作答）

```
Update emp
Set sal=sal*1.05
Where deptno=(select deptno
                From dept
                Where dname='Research');
```

- (A) 找出 emp 資料表內 dname 欄位是 Research 的資料列，並對該些資料列的 sal 欄位進行更新的動作，調整為原 sal 的 1.05 倍
- (B) 子查詢內先在 dept 資料表中找出 dname 為 Research 的 deptno 值，再將這些 deptno 值回傳到父查詢，父查詢為更新 emp 資料表，找出 emp 資料表的 deptno 欄位符合子查詢回傳結果的資料列，並將該些資料列的 sal 欄位值調整成原值的 1.05 倍
- (C) 更新資料表 emp，對其所有資料列中的 sal 欄位值都進行更新動作，調整幅度為原 sal 的 1.05 倍
- (D) 子查詢內先在 emp 資料表中找出 dname 為 Research 的 deptno 值，再將這些 deptno 值回傳到父查詢，父查詢為更新 dept 資料表，找出 dept 資料表的 deptno 欄位符合子查詢回傳結果的資料列，並將該些資料列的 sal 欄位值調整成原值的 1.05 倍

ACD

2-06.

結構化查詢語言 (SQL) 包含哪些種類？(複選)

- (A) 資料控制語言 (Data Control Language, DCL)
- (B) 資料查詢語言 (Data Query Language, DQL)
- (C) 資料定義語言 (Data Definition Language, DDL)
- (D) 資料操作語言 (Data Manipulation Language, DML)

C

2-07.

下列何者為非結構化資料庫儲存資料的方式？

- (A) 矩陣方式儲存
- (B) 表格方式儲存
- (C) 資料以鍵值對 (key-value pair) 的方式儲存，無需事先定義綱要
- (D) XML 格式儲存

## 企業電子化人才能力鑑定

2021/1/1

- ABCD      2-08.      關於比例（ratio）屬性值擁有下列哪些運算性質？（複選）  
（A）不等性（distinctness）  
（B）順序性（order）  
（C）可加性（addition）  
（D）可乘性（multiplication）
- ACD      2-09.      影響資料分析技術的重要資料集特質有哪些？（複選）  
（A）維度（dimensionality）  
（B）連續性（continuity）  
（C）稀疏性（sparsity）  
（D）分辨率（resolution）
- B      2-10.      關於資料庫的描述，下列何者錯誤？  
（A）在許多領域裡，如：商業、電子商務、工程和圖書館等，資料庫扮演重要的角色  
（B）資料庫內所儲存的資料，一般而言為不相關的資料  
（C）使用者可以對資料庫中的檔案資料執行新增、擷取、更新、刪除等操作  
（D）資料庫為儲存電子檔案的場所
- D      2-11.      SQL 語法中用來修改資料表的指令為何？  
（A）UPDATE TABLE  
（B）CREAT TABLE  
（C）DROP TABLE  
（D）ALTER TABLE
- ABC      2-12.      下列哪些指令屬於資料操作語言（Data Manipulation Language）？（複選）  
（A）INSERT  
（B）UPDATE  
（C）DELETE  
（D）CREATE
- D      2-13.      下列何者不屬於非結構化資料？  
（A）語音檔  
（B）影像檔  
（C）電子郵件  
（D）XML 檔

### 第三章 試題彙編

- B 2-14. 關於 JSON (JavaScript Object Notation) 格式的敘述，下列何者為非？  
（A）易於人們閱讀和編寫  
（B）陣列可用大括弧 { } 來寫入資料  
（C）是一種資料交換格式源自於 JavaScript  
（D）與其他程式相容性高
- BC 2-15. 關於 NoSQL 資料庫的敘述，下列哪些選項是正確的？(複選)  
（A）NoSQL 適用於需要使用關聯模型的資料  
（B）對於行動裝置用戶，NoSQL 資料庫的反應速度較快  
（C）泛指非關聯式資料庫管理系統模型的資料庫管理系統  
（D）無法同時多人連線
- ABC 2-16. 下列哪些資料庫屬於 NoSQL 資料庫？(複選)  
（A）MongoDB 資料庫  
（B）CouchDB 資料庫  
（C）Redis 資料庫  
（D）MySQL 資料庫
- D 2-17. 通過屬性對應數值的性質，可以分為四類：a 名目 (nominal)、b 順序 (order)、c 區間 (interval) 和 d 比率 (ratio)，請判斷下列屬性屬於哪種類型：(1) 長度、(2) 鑽石硬度 (好，較好，最好)、(3) 郵政編碼、(4) 攝氏溫度，下列配對何者正確？  
（A）(1)-b、(2)-c、(3)-b、(4)-c  
（B）(1)-d、(2)-c、(3)-a、(4)-b  
（C）(1)-b、(2)-c、(3)-a、(4)-c  
（D）(1)-d、(2)-b、(3)-a、(4)-c
- D 2-18. 結構化查詢語言 (SQL) 的刪除資料指令為何？  
（A）DROP  
（B）REMOVE  
（C）CLEAR  
（D）DELETE
- B 2-19. 結構化查詢語言 (SQL) 的更新資料指令為何？  
（A）RENEW  
（B）UPDATE  
（C）INSERT  
（D）REVISE

## 企業電子化人才能力鑑定

2021/1/1

- D 2-20. 如附圖所示，已知產品資料表：Product (ProductID, DesignDate, DesignerID)，找出產品資料表中設計日期為 2014 年 9 月 13 日 (含) 以前所有資料的 SQL 指令為何？(請參閱附圖作答)

| ProductID | DesignDate | DesignerID |
|-----------|------------|------------|
| 1         | 2014/09/12 | 1          |
| 2         | 2014/09/13 | 1          |
| 3         | 2014/10/02 | 2          |
| 4         | 2014/10/02 | 3          |
| 5         | 2014/10/11 | 2          |

- (A) SELECT \* FROM Product WITH DesignDate < '9/13/2014'  
(B) SELECT \* FROM Product WHERE DesignDate < '9/13/2014'  
(C) SELECT \* FROM Product WITH DesignDate <= '9/13/2014'  
(D) SELECT \* FROM Product WHERE DesignDate <= '9/13/2014'

- C 2-21. 如附圖所示，已知產品資料表：Product (ProductID, DesignDate, DesignerID)，找出產品資料表中設計師編號為 2 或 3 所有資料的 SQL 指令為何？(請參閱附圖作答)

| ProductID | DesignDate | DesignerID |
|-----------|------------|------------|
| 1         | 2014/09/12 | 1          |
| 2         | 2014/09/13 | 1          |
| 3         | 2014/10/02 | 2          |
| 4         | 2014/10/02 | 3          |
| 5         | 2014/10/11 | 2          |

- (A) SELECT \* FROM Product WHERE DesignerID = 2 AND DesignerID = 3  
(B) SELECT \* FROM Product WHERE DesignerID = 2 IN DesignerID = 3  
(C) SELECT \* FROM Product WHERE DesignerID = 2 OR DesignerID = 3  
(D) SELECT \* FROM Product WHERE DesignerID = 2 BOTH DesignerID = 3

### 第三章 試題彙編

- D 2-22. 如附圖所示，已知產品資料表：Product (ProductID, DesignDate, DesignerID)，找出產品資料表中不重覆的 DesignerID 資料的 SQL 指令為何？（請參閱附圖作答）

| ProductID | DesignDate | DesignerID |
|-----------|------------|------------|
| -----     | -----      | -----      |
| 1         | 2014/09/12 | 1          |
| 2         | 2014/09/13 | 1          |
| 3         | 2014/10/02 | 2          |
| 4         | 2014/10/02 | 3          |
| 5         | 2014/10/11 | 2          |

- (A) SELECT DIFFERENT DesignerID FROM Product  
 (B) SELECT SINGLE DesignerID FROM Product  
 (C) SELECT PRECISE DesignerID FROM Product  
 (D) SELECT DISTINCT DesignerID FROM Product

- C 2-23. 已知產品資料表：Product (ProductID, DesignDate, DesignerID)，以下 SQL 指令結果為何？（請參閱附圖作答）

| ProductID | DesignDate | DesignerID |
|-----------|------------|------------|
| -----     | -----      | -----      |
| 1         | 2014/09/12 | 1          |
| 2         | 2014/09/13 | 1          |
| 3         | 2014/10/02 | 2          |
| 4         | 2014/10/02 | 3          |
| 5         | 2014/10/11 | 2          |

SELECT COUNT (DesignerID) FROM Product

- (A) 3  
 (B) 4  
 (C) 5  
 (D) 9

## 企業電子化人才能力鑑定

2021/1/1

- D 2-24. 下列非結構化與結構化資料的敘述何者正確？  
（A）非結構化資料其訊息形式通常是相對固定  
（B）網際網路上出現的海量資料，只有非結構化一種  
（C）結構化訊息是不可以數字化的  
（D）非結構化資料相對於結構化資料，從巨觀上看也是結構化訊息的一種形式
- D 2-25. 下列語句中，與語句 " WHERE vend\_id = 'DLL01' OR vend\_id = 'BRS01' " 等價的為下列何者？  
（A）WHERE vend\_id = 'DLL01' AND 'BRS01'  
（B）WHERE vend\_id IS 'DLL01' OR 'BRS01'  
（C）WHERE vend\_id %I N% ('DLL01''BRS01')  
（D）WHERE vend\_id I N ('DLL01','BRS01')
- B 2-26. SQL 可以使用 SELECT 語句的 GROUP BY 子句進行分組，下列關於子句敘述，何者有誤？  
（A）使用聚集函數（如 SUM 求和函數）時，藉助 GROUP BY 子句可以實現對每個小組的資料進行計算  
（B）如果分組行中有 NULL 值，那麼 NULL 對應的列會被跳過，不會放入任何一組  
（C）GROUP BY 子句可以包含任意數目的行（column），實現嵌套分組  
（D）GROUP BY 子句中列出的每個行（column）都必須是檢索行或有效的表達式
- A 2-27. 下列有關 NoSQL 資料庫的敘述何者正確？  
（A）MongoDB 可以採用 BSON (Binary Format Of JSON) 形式的儲存文件  
（B）Redis 無法採用鍵值對 (Key-Value Pair) 的方式儲存  
（C）鍵值儲存資料需事先定義綱要  
（D）Neo4j 是以文件為基礎的儲存方式
- A 2-28. 以下關於資料屬性尺度的範例中，哪一項是正確的？  
（A）攝氏溫度是區間尺度  
（B）華氏溫度是比例尺度  
（C）郵遞區號是順序尺度  
（D）年級是名目尺度

### 第三章 試題彙編

- C      2-29.      以下關於屬性運算性質的說法，哪一項是錯誤的？
- (A) 我們可以說“2 米是 1 米的 2 倍”，所以這個資料具有比例屬性
  - (B) 我們可以說“30 攝氏度比 20 攝氏度高 10 度”，所以這個資料具有區間屬性
  - (C) 我們可以說“代表男性的 2 大於代表女性的 1”，所以這個資料具有順序屬性
  - (D) 我們可以說“黑色與白色不同”，所以這個資料具有名目屬性
- A      2-30.      以下關於資料集類型的論述，哪一項是錯誤的？
- (A) 明星的照片是圖形資料 (graph data)
  - (B) 財務報表是紀錄資料 (record data)
  - (C) 地理訊息資料是有序資料 (sequence data)
  - (D) 基因序列是有序資料 (sequence data)
- C      2-31.      以下常用資料庫中，哪一項不是關聯式資料庫？
- (A) MySQL
  - (B) Oracle
  - (C) MongoDB
  - (D) SQL Server
- AC      2-32.      以下關於關聯式資料庫中資料表的說法哪些正確？(複選)
- (A) 資料表是二維表
  - (B) 資料表的行列可以互換
  - (C) 資料表的單欄主索引鍵是唯一的
  - (D) 資料表的主索引鍵必須是數值型
- D      2-33.      在 SQL 語言中，以下哪一個命令並不會修改資料庫？
- (A) INSERT
  - (B) UPDATE
  - (C) DELETE
  - (D) SELECT

## 企業電子化人才能力鑑定

2021/1/1

- BC            2-34.        以下 SQL 語句中，哪些項目是錯誤的？(複選)
- ( A ) SELECT \* FROM Product
  - ( B ) SELECT \* FROM Product WHERE DesignDate <= 10/2/2006
  - ( C ) SELECT \* FROM Product WHERE DesignerID > 3 JOIN DesignerID < 6
  - ( D ) SELECT DesignerID, COUNT(DesignerID) AS 總設計量 FROM Product GROUP BY DesignerID
- ACD           2-35.        關於 SQL 語言，下列描述中哪些是正確的？(複選)
- ( A ) SQL 是美國國家標準局訂定的語言標準
  - ( B ) 資料操作語言 (DML) 用來對資料進行定義和擷取
  - ( C ) SQL 語言在不同的資料庫版本中可能會不同
  - ( D ) 資料控制語言 (DCL) 用來對資料的存取權進行控制
- A            2-36.        下列關於 XML 的說法中哪一項是正確的？
- ( A ) XML 的格式非常彈性
  - ( B ) XML 預設了大量的標籤
  - ( C ) XML 繼承了 HTML 中的格式和標籤
  - ( D ) XML 比其普通文本佔用了更少的存儲空間
- A            2-37.        以下哪一項不是非結構化的多媒體資料之應用？
- ( A ) 通過營收與成本資料進行利潤分析
  - ( B ) 通過電影的色調來分析不同類型電影的偏好
  - ( C ) 通過音頻分析對音樂進行分類
  - ( D ) 通過人臉識別實現自動對焦
- B            2-38.        以下何項並非將非結構化資料轉換成結構化資料？
- ( A ) 將斷詞前的文字資料轉換成文件詞項矩陣
  - ( B ) 將矩陣資料轉換成熱圖
  - ( C ) 將 JSON 文件轉換成表格形式
  - ( D ) 將圖像 (image) 資料轉換成顏色代碼矩陣
- AC           2-39.        關於 NoSQL 資料庫的描述，以下哪些項目不正確？(複選)
- ( A ) NoSQL 的意思是 " 不使用 SQL "
  - ( B ) NoSQL 資料庫具有很好的縮放性
  - ( C ) NoSQL 資料庫通過資料表建構而成
  - ( D ) NoSQL 資料庫可以使用文件存儲、鍵值存儲、圖形存儲等多種方式



### 第三章 試題彙編

- B      2-40.      下列何者屬於量化屬性？  
 (A) 公司部門  
 (B) 一批產品的不良數  
 (C) 學校學生的性別  
 (D) 商品付款方式
- D      2-41.      下列針對屬性衡量尺度的描述何者正確？  
 (A) 名目尺度 (nominal scale) 僅表示群或類別，例如身分證字號、瞳孔顏色、居住地、期中考排名等  
 (B) 區間尺度 (interval scale) 無絕對零點，但差異或距離有意義。如日期、攝氏或華氏溫度  
 (C) 順序尺度 (ordinal scale) 通常用於順序有別時，例如身高高度 (低、中、高)、河水流速、年收入排名等  
 (D) 比例尺度 (ratio scale) 其有自然零點，且比率有意義，例如克式 (Kelvin) 溫度、長度、時間、次數等
- A      2-42.      針對屬性的分類方式中，下列連續屬性的描述何者有誤？  
 (A) 通常以整數變數呈現  
 (B) 以實數值為其屬性值  
 (C) 常見的如溫度、高度或重量等  
 (D) 雖是實數值，實務上仍以有限的數字來衡量與呈現
- B      2-43.      下列選項中何者不屬於 SQL 的三大種類？  
 (A) 資料定義語言 (Data Definition Language, DDL)  
 (B) 資料驗證語言 (Data Verification Language, DVL)  
 (C) 資料控制語言 (Data Control Language, DCL)  
 (D) 資料操作語言 (Data Manipulation Language, DML)
- A      2-44.      請選出符合下述情境的 SQL 指令「找出 Product 資料表中設計日期為 103 年 10 月 2 日之前 (包括 103 年 10 月 2 日) 所有資料，建議日期格式以 (‘月/日/西元年’) 為主」  
 (A) `SELECT * FROM Product WHERE DesignDate <= '10/2/2014'`  
 (B) `SELECT DesignDate FROM Product WHERE '10/2/2014'`  
 (C) `SELECT * FROM Product WHERE DesignDate => '10/2/2014'`  
 (D) `UPDATE * FROM Product WHERE DesignDate => '10/2/2014'`

## 企業電子化人才能力鑑定

2021/1/1

- B            2-45.        請選出符合下述情境的 SQL 指令「找出 Product 資料表中設計日期為 2014 年 10 月 2 日且設計者編號是 1001 的所有資料」
- (A) SELECT \* FROM Product WHERE DesignDate = '10/2/2014' OR DesignerID =1001
  - (B) SELECT \* FROM Product WHERE DesignDate = '10/2/2014' AND DesignerID =1001
  - (C) SELECT DesignDate AND DesignerID FROM Product WHERE='10/2/2014' AND 1001
  - (D) SELECT DesignDate='10/2/2014' AND DesignerID =1001 FROM Product
- D            2-46.        下列對非結構化訊息的描述何者錯誤？
- (A) 訊息形式相對不固定，通常是各種格式的文件（text）、影像（image）、音頻（audio）、視頻（video）
  - (B) 非結構化相對於結構化訊息而言，從巨觀上看也是結構化訊息的一種形式
  - (C) 結構化訊息是可以數字化的，方便地通過電腦與資料庫技術進行管理
  - (D) 網際網路上出現的海量訊息，大概分為結構化和非結構化兩種
- D            2-47.        下列何項不是文字資料探勘的工作？
- (A) 語料庫（corpus）製作
  - (B) 詞幹還原（words stemming）
  - (C) 建構關鍵詞-文件矩陣（term-document matrix）
  - (D) 擷取畫素、顏色編碼、頻率、幀率、赫茲、畫面格數等特徵
- A            2-48.        NoSQL 適用於不需要使用關聯模型的資料，近年來因大型網頁應用程式（例如：Google、Facebook、Yahoo、Twitter 等公司的系統）興起，下列何者不為其特色？
- (A) 採用綱要式（schema）的資料表達方式，彈性且無須太多的事前規劃，但仍可與時俱進
  - (B) 可處理大量資料
  - (C) 有較高的可用性，亦即可以同時多人連線，克服關聯式資料庫（RDBMS）資料鎖住（data lock）的問題
  - (D) 縮放性（scalability）佳

## 第三章 試題彙編

- B 2-49. 下列何者不為常見的 NoSQL 資料儲存方式？  
 (A) JSON 文件儲存方式  
 (B) 實體關聯儲存方式  
 (C) 鍵值對 (key-value pair) 儲存方式  
 (D) 節點、邊、關係 (relation) 等屬性的圖形 (graph) 結構儲存方式
- D 2-50. 關於 HTML，下列敘述何者有誤？  
 (A) HTML 文檔也被稱為網頁  
 (B) HTML 元素都要以開始標籤起始，以結束標籤終止  
 (C) <h3> 與 </h3> 之間的文本為三級標題  
 (D) <body> 與 </body> 之間的文本是段落
- ABC 2-51. 區間尺度 (interval scale) 可做以下哪幾種運算？(複選)  
 (A) 不等性 (distinctness) 運算  
 (B) 順序性 (order) 運算  
 (C) 可加性 (addition) 運算  
 (D) 可乘性 (multiplication) 運算
- B 2-52. 每日股票交易價格，可視為以下何種有序資料？  
 (A) 時間 (temporal) 資料  
 (B) 次序 (sequential) 資料  
 (C) 空間 (spatial) 資料  
 (D) 基因序列 (genetic sequence) 資料
- D 2-53. 以下何種資料庫為非關聯型資料庫 (Not Only SQL，簡稱 NoSQL)？  
 (A) SQL Server  
 (B) Oracle  
 (C) MySQL  
 (D) MongoDB
- BC 2-54. 校務系統的關聯式資料庫中，下述資料欄位 (或欄位組合) 哪些可以設定為學生資料表之主索引鍵 PK？(複選)  
 (A) 姓名、生日  
 (B) 身分證字號  
 (C) 學號  
 (D) 電話、地址

# 企業電子化人才能力鑑定

2021/1/1

- B            2-55.        關於 XML，下列敘述何者有誤？
- ( A ) XML 與 HTML 都是非結構化資料
  - ( B ) HTML 用來傳輸和儲存資料，而 XML 被用來顯示資料
  - ( C ) 通過 XML，可以在不同的系統之間輕鬆地交換資料
  - ( D ) XML 允許創作者定義自己的標籤和文檔結構
- D            2-56.        關於 JSON，下列敘述何者有誤？
- ( A ) JSON 是 JavaScript 的物件表示法 (JavaScript Object Notation)
  - ( B ) JSON 解析器和 JSON 資料集支援許多不同的程式語言
  - ( C ) JSON 非常易於解析
  - ( D ) JSON 文本格式在語法上與創建 JavaScript 物件的代碼不同

### 第三章 試題彙編

A

2-57.

如附圖何者 SQL 語法可以找出所有產品編號、設計日期、以及設計人員姓名？（請參閱附圖作答）

- 產品資料表：Product(產品編號, 設計日期, 設計者編號).PK：產品編號
- 設計人員資料表：Designer(設計者編號, 部門編號, 姓名, 戶籍).PK：設計者編號
- 部門資料表：Department(部門編號, 部門名稱).PK：部門編號

產品資料表

| ProductID | DesignDate | DesignerID |
|-----------|------------|------------|
| 1         | 2006/09/12 | 1          |
| 2         | 2006/09/13 | 1          |
| 3         | 2006/10/02 | 2          |
| 4         | 2006/10/02 | 3          |
| 5         | 2006/10/11 | 2          |

設計人員資料表

| DesignerID | Department | Name   | Residence |
|------------|------------|--------|-----------|
| 1          | 1          | ALAN   | 台北市       |
| 2          | 1          | SERENA | 台北市       |
| 3          | 2          | ANDY   | 台中市       |
| 4          | 3          | ALPHA  | 高雄市       |

部門資料表

| DepartmentID | DepartmentName |
|--------------|----------------|
| 1            | 設計一課           |
| 2            | 設計二課           |
| 3            | 設計三課           |

A 選項

```
SELECT ProductID, DesignDate, Designer.Name
FROM Product
INNER JOIN Designer
ON Product.DesignerID = Designer.DesignerID
```

B 選項

```
SELECT ProductID, DesignDate, Name
FROM Product, Designer
```

C 選項

```
SELECT ProductID, DesignDate, Designer.Name
FROM Product
INNER JOIN Designer
WHERE Product.DesignerID = Designer.DesignerID
```

D 選項

```
SELECT ProductID, DesignDate, Designer.Name
FROM Product
INNER JOIN Designer
ON Product.ProductID = Designer.ProductID
```

(A) 選項 (A)

(B) 選項 (B)

(C) 選項 (C)

(D) 選項 (D)

## 企業電子化人才能力鑑定

2021/1/1

A

2-58.

如附圖何者 SQL 語法可以找出所有設計者之最新設計產品的日期，並顯示欄位三個：設計者編號、設計者姓名、最新設計產品的日期？（請參閱附圖作答）

|                                                                                                                                                                                              |                                                                                                                                                                                                             |                             |                                      |                                                                                                                                                                       |  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|--------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| <ul style="list-style-type: none"><li>產品資料表：Product(產品編號, 設計日期, 設計者編號).PK：產品編號</li><li>設計人員資料表：Designer(設計者編號, 部門編號, 姓名, 戶籍).PK：設計者編號</li><li>部門資料表：Department(部門編號, 部門名稱).PK：部門編號</li></ul> |                                                                                                                                                                                                             |                             |                                      |                                                                                                                                                                       |  |
| <b>產品資料表</b>                                                                                                                                                                                 |                                                                                                                                                                                                             | <b>設計人員資料表</b>              |                                      |                                                                                                                                                                       |  |
| ProductID                                                                                                                                                                                    | DesignDate                                                                                                                                                                                                  | DesignerID                  | DesignerID Department Name Residence |                                                                                                                                                                       |  |
| 1                                                                                                                                                                                            | 2006/09/12                                                                                                                                                                                                  | 1                           | 1 ALAN 台北市                           |                                                                                                                                                                       |  |
| 2                                                                                                                                                                                            | 2006/09/13                                                                                                                                                                                                  | 1                           | 2 SERENA 台北市                         |                                                                                                                                                                       |  |
| 3                                                                                                                                                                                            | 2006/10/02                                                                                                                                                                                                  | 2                           | 3 ANDY 台中市                           |                                                                                                                                                                       |  |
| 4                                                                                                                                                                                            | 2006/10/02                                                                                                                                                                                                  | 3                           | 4 ALPHA 高雄市                          |                                                                                                                                                                       |  |
| 5                                                                                                                                                                                            | 2006/10/11                                                                                                                                                                                                  | 2                           |                                      |                                                                                                                                                                       |  |
|                                                                                                                                                                                              |                                                                                                                                                                                                             | <b>部門資料表</b>                |                                      |                                                                                                                                                                       |  |
|                                                                                                                                                                                              |                                                                                                                                                                                                             | DepartmentID DepartmentName |                                      |                                                                                                                                                                       |  |
|                                                                                                                                                                                              |                                                                                                                                                                                                             | 1 設計一課                      |                                      |                                                                                                                                                                       |  |
|                                                                                                                                                                                              |                                                                                                                                                                                                             | 2 設計二課                      |                                      |                                                                                                                                                                       |  |
|                                                                                                                                                                                              |                                                                                                                                                                                                             | 3 設計三課                      |                                      |                                                                                                                                                                       |  |
| A 選項                                                                                                                                                                                         | SELECT d.DesignerID, d.Name, a.NewestDate<br>FROM (SELECT DesignerID,<br>MAX(DesignDate) AS NewestDate<br>FROM Product<br>GROUP BY DesignerID) a<br>INNER JOIN Designer d<br>ON a.DesignerID = d.DesignerID |                             | C 選項                                 | SELECT d.DesignerID, d.Name,<br>MAX(p.DesignDate) AS NewestDate<br>FROM Product p<br>INNER JOIN Designer d<br>ON p.DesignerID = d.DesignerID<br>GROUP BY d.DesignerID |  |
| B 選項                                                                                                                                                                                         | SELECT d.DesignerID, d.Name, a.NewestDate<br>FROM (SELECT DesignerID,<br>MIN(DesignDate) AS NewestDate<br>FROM Product<br>GROUP BY DesignerID) a<br>INNER JOIN Designer d<br>ON a.DesignerID = d.DesignerID |                             | D 選項                                 | SELECT d.DesignerID, d.Name,<br>MIN(p.DesignDate) AS NewestDate<br>FROM Product p<br>INNER JOIN Designer d<br>ON p.DesignerID = d.DesignerID<br>GROUP BY d.DesignerID |  |

(A) 選項 (A)

(B) 選項 (B)

(C) 選項 (C)

(D) 選項 (D)

### 第三章 試題彙編

C

2-59.

如附圖何者 SQL 語法可以找出所有居住於台中市的設計者，並依設計者姓名的字典順序降冪排列？（請參閱附圖作答）

- 產品資料表：Product(產品編號, 設計日期, 設計者編號), PK：產品編號
- 設計人員資料表：Designer(設計者編號, 部門編號, 姓名, 戶籍), PK：設計者編號
- 部門資料表：Department(部門編號, 部門名稱), PK：部門編號

產品資料表

| ProductID | DesignDate | DesignerID |
|-----------|------------|------------|
| 1         | 2006/09/12 | 1          |
| 2         | 2006/09/13 | 1          |
| 3         | 2006/10/02 | 2          |
| 4         | 2006/10/02 | 3          |
| 5         | 2006/10/11 | 2          |

設計人員資料表

| DesignerID | Department | Name   | Residence |
|------------|------------|--------|-----------|
| 1          | 1          | ALAN   | 台北市       |
| 2          | 1          | SERENA | 台北市       |
| 3          | 2          | ANDY   | 台中市       |
| 4          | 3          | ALPHA  | 高雄市       |

部門資料表

| DepartmentID | DepartmentName |
|--------------|----------------|
| 1            | 設計一課           |
| 2            | 設計二課           |
| 3            | 設計三課           |

A 選項

```
SELECT Name
FROM Desinger
ORDER BY Name
WHERE Residence = ' 台中市'
```

B 選項

```
SELECT Name
FROM Desinger
WHERE Residence = ' 台中市'
ORDER BY Name
```

C 選項

```
SELECT Name
FROM Desinger
WHERE Residence = ' 台中市'
ORDER BY Name DESC
```

D 選項

```
SELECT Name
FROM Desinger
WHERE Residence = ' 台中市'
GROUP BY Name
```

(A) 選項 (A)

(B) 選項 (B)

(C) 選項 (C)

(D) 選項 (D)

## 企業電子化人才能力鑑定

2021/1/1

B

2-60.

如附圖何者符合規範之 XML 文件？（請參閱附圖作答）

|      |                                                                                                                                                                                                                                                                                                                                                                                                  |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A 選項 | <pre> &lt;personname&gt;   &lt;firstname&gt;Jane&lt;/firstname&gt;   &lt;surname&gt;Smith Doe&lt;/surname&gt; &lt;/personname&gt; &lt;email&gt;janesd@gmail.com&lt;/email&gt; &lt;affiliation&gt;   &lt;orgname&gt;University of California&lt;/orgname&gt;   &lt;orgdiv&gt;Department of Statistics&lt;/orgdiv&gt; &lt;/affiliation&gt; </pre>                                                  |
| B 選項 | <pre> &lt;author&gt;   &lt;personname&gt;     &lt;firstname&gt;Jane&lt;/firstname&gt;     &lt;surname&gt;Smith Doe&lt;/surname&gt;   &lt;/personname&gt;   &lt;email&gt;janesd@gmail.com&lt;/email&gt;   &lt;affiliation&gt;     &lt;orgname&gt;University of California&lt;/orgname&gt;     &lt;orgdiv&gt;Department of Statistics&lt;/orgdiv&gt;   &lt;/affiliation&gt; &lt;/author&gt; </pre> |
| C 選項 | <pre> &lt;Author&gt;   &lt;Personname&gt;     &lt;Firstname&gt;Jane&lt;/firstname&gt;     &lt;Surname&gt;Smith Doe&lt;/surname&gt;   &lt;/personname&gt;   &lt;Email&gt;janesd@gmail.com&lt;/email&gt;   &lt;Affiliation&gt;     &lt;Orgname&gt;University of California&lt;/orgname&gt;     &lt;Orgdiv&gt;Department of Statistics&lt;/orgdiv&gt;   &lt;/affiliation&gt; &lt;/author&gt; </pre> |
| D 選項 | <pre> &lt;author&gt;   &lt;personname&gt;     &lt;firstname&gt;Jane&lt;firstname&gt;     &lt;surname&gt;Smith Doe&lt;surname&gt;   &lt;/personname&gt;   &lt;email&gt;janesd@gmail.com&lt;email&gt;   &lt;affiliation&gt;     &lt;orgname&gt;University of California&lt;orgname&gt;     &lt;orgdiv&gt;Department of Statistics&lt;orgdiv&gt;   &lt;/affiliation&gt; &lt;/author&gt; </pre>      |

(A) 選項 (A)

(B) 選項 (B)

(C) 選項 (C)

(D) 選項 (D)



### 第三章 試題彙編

- C      2-61.      以下何者不為 NoSQL 資料庫的特性？
- (A) 採用無綱要 (schemaless) 的資料表達方式，彈性且無須太多的事前規劃
  - (B) 無需撰寫複雜的 SQL 查詢指令，因此基於 NoSQL 的應用系統開發速度較快
  - (C) 資料皆經過正規化，因此節省儲存空間
  - (D) 採用 NoSQL 資料庫可為應用系統的規模縮放性 (scalability) 預做準備，彈性的應用系統可以應付突來的人流高峰
- B      2-62.      關於 NoSQL 資料庫，下列敘述何者有誤？
- (A) NoSQL 可以理解為 " NotOnlySQL "，是對不同於傳統的關聯式資料庫 (relational database) 的統稱
  - (B) NoSQL 使用 SQL 作為查詢語言，因此具有靈活的資料模型
  - (C) NoSQL 資料庫都去掉了關聯式資料庫的關聯式特性，這使得資料庫更容易擴展
  - (D) NoSQL 資料庫都具有非常高的讀寫性能
- D      2-63.      考慮下列數據：6, 47, 49, 15, 42, 41, 7, 39, 43, 40, 36，請問第 3 四分位數 (Q3) 為何？
- (A) 40
  - (B) 41
  - (C) 42
  - (D) 43
- D      2-64.      將非結構化資料轉換為結構化格式，下列何者為非？
- (A) 文件轉為各文件與詞項的橫列縱行矩陣
  - (B) 剖析 XML 的語法後轉換為列表 (list) 或表格 (table) 的資料格式
  - (C) 擷取圖片、音頻與視頻中的特徵製成表格的形式
  - (D) 將資料以 XML 格式儲存
- A      2-65.      若欲使用 SQL 語言，找出 Product 資料表中設計日期 (designDate) 為 95 年 10 月 2 日之前 (包括 95 年 10 月 2 日) 所有資料，建議日期格式以 '10/2/2006' 為主 ('月/日/西元年')
- (A) SELECT \* FROM Product WHERE designDate<='10/2/2006'
  - (B) SELECT \* FROM designDate WHERE Date > 951002
  - (C) SELECT \* FROM designDate WHERE Product <='10/2/2006'
  - (D) SELECT \* FROM Product

## 企業電子化人才能力鑑定

2021/1/1

- C            2-66.        以下何種資料不是非結構化資料？  
                  ( A ) HTML 格式的文本資料  
                  ( B ) XML 格式的文本資料  
                  ( C ) 空間資料  
                  ( D ) 影像資料
- B            2-67.        下列關於資料的描述何者錯誤？  
                  ( A ) 屬性是資料物件的性質或特徵，例如河川的流速、學生的身高、瞳孔的顏色是否為棕色等  
                  ( B ) 一群資料物件 (object, 通常為縱行) 及其屬性 (attribute, 通常為橫列) 所形成的集合  
                  ( C ) 通常以多個屬性描述資料物件，屬性又稱為變數 (variable)、欄位 (field) 或特徵 (feature)  
                  ( D ) 資料物件又稱為紀錄 (record)、資料點 (point)、樣本 (sample)、實體 (entity) 或樣例 (instance, case)
- ACD        2-68.        下列對於資料格式的描述，哪些選項是正確的？(複選)  
                  ( A ) JSON 格式 (JavaScript Object Notation) 是一種輕量級的資料交換格式，源自於 JavaScript  
                  ( B ) HTML 格式現已成為網頁及其他領域中，在進行各種資料交換的依循標準  
                  ( C ) XML 格式是一個簡單且非常彈性的文字資料格式  
                  ( D ) XML 並不是用來編排內容，而是用來描述資料，它並沒有如同 HTML 一般的預設標籤，事實上使用者需要定義描述資料的各種標籤
- C            2-69.        下列何者為名目尺度 (nominal scale) ？  
                  ( A ) 長度  
                  ( B ) 日期  
                  ( C ) 郵遞區號  
                  ( D ) 年級
- C            2-70.        下列對於訊息的敘述何者為錯誤？  
                  ( A ) 非結構化訊息形式相對較不固定  
                  ( B ) 結構化訊息是可以數字化的  
                  ( C ) 非結構化訊息通常包括 Excel 檔案  
                  ( D ) 結構化訊息包括訂單資料

### 第三章 試題彙編

- B            2-71.        下列哪一項資料庫不是關聯式資料庫？  
                  ( A ) SQL Server  
                  ( B ) MongoDB  
                  ( C ) MariaDB  
                  ( D ) MySQL
- D            2-72.        衣服尺寸包括大、中、小是屬於哪一種衡量尺度？  
                  ( A ) 比例尺度 <ratio scale>  
                  ( B ) 名目尺度 <nominal scale>  
                  ( C ) 區間尺度 <interval scale>  
                  ( D ) 順序尺度 <ordinal scale>
- ABCD        2-73.        下列 SQL 語句中，哪些能夠檢索出 prod\_name (產品名稱) 為 "Fish bean bag toy" 的資料？(複選)  
                  ( A ) WHERE prod\_name LIKE 'Fish%'  
                  ( B ) WHERE prod\_name LIKE '%bean%'  
                  ( C ) WHERE prod\_name LIKE 'F%y'  
                  ( D ) WHERE prod\_name LIKE '%F%y'
- C            2-74.        關於 SQL 語句，下列敘述何者有誤？  
                  ( A ) 在 SELECT 語句中，如果給定一個萬用字符 (\*)，則返回所有行 (column)  
                  ( B ) 檢索不需要的行 (column) 通常會降低檢索和程序性能，所以一般情況下最好不要使用萬用字符 (\*)  
                  ( C ) SQL 語句區分大小寫  
                  ( D ) 藉助萬用字符 (\*)，還能檢索出名字未知的行 (column)

## 企業電子化人才能力鑑定

2021/1/1

- C            2-75.        現有資料表 Products，其中 prod\_name，prod\_price 分別代表產品的名稱和價格，如附圖該如何找出沒有登記價格的產品名稱？（請參閱附圖作答）

|      |                                                                     |
|------|---------------------------------------------------------------------|
| A 選項 | <pre>SELECT prod_name FROM Products WHERE prod_price IS 0;</pre>    |
| B 選項 | <pre>SELECT prod_name FROM Products BY prod_price IS 0;</pre>       |
| C 選項 | <pre>SELECT prod_name FROM Products WHERE prod_price IS NULL;</pre> |
| D 選項 | <pre>SELECT prod_name FROM Products BY prod_price IS NULL;</pre>    |

- (A) 選項 A  
(B) 選項 B  
(C) 選項 C  
(D) 選項 D

### 第三章 試題彙編

DPA

「巨量資料處理與分析」學科試題及答案

#### 第三類：統計分析基礎

- ABC      03-01.      下列哪些為平均數的優點？(複選)
- (A) 資料無極端值或偏態時，具有代表性
  - (B) 考慮所有觀察值，敏感度較高
  - (C) 適合代數與統計推論計算
  - (D) 不受極端值影響
- B      03-02.      下列何者為平均絕對離差 (mean absolute deviation, MAD) 的定義？
- (A) 一組統計資料中，出現次數最多之觀測值
  - (B) 計算各觀察值與平均數的距離總和，再取其算術平均數，而不考慮各均差之正負號
  - (C) 將資料按其數值大小順序排列，取該數列的中間數值，或中間兩項數值的平均數
  - (D) 資料中的最大值與最小值的差
- C      03-03.      有八個數字分別為 11, 11, 12, 13, 13, 13, 13, 13，試問此八個數字之全距、眾數、平均數分別為何？
- (A) 全距為 13，眾數為 2，平均數為 11
  - (B) 全距為 3，眾數為 13，平均數為 11
  - (C) 全距為 2，眾數為 13，平均數為 12.375
  - (D) 全距為 13，眾數為 13，平均數為 2

## 企業電子化人才能力鑑定

2021/1/1

B

03-04.

附圖為某商品在各時期的銷售價格時，所成交的次數，請問成交次數出現最多次的組別是哪一組？（請參閱附圖作答）

| 商品銷售價格 成交次數 |       |    |
|-------------|-------|----|
| 1           | 5-10  | 12 |
| 2           | 10-15 | 21 |
| 3           | 15-20 | 29 |
| 4           | 20-25 | 10 |
| 5           | 25-30 | 9  |
| 6           | 30-35 | 5  |
| 7           | 35-40 | 4  |
| 8           | 40-45 | 1  |
| 9           | 45-50 | 1  |
| 10          | 50-55 | 0  |
| 11          | 55-60 | 0  |
| 12          | 60-65 | 1  |

- (A) 第一組銷售價格為 5-10 時，成交次數最多  
 (B) 第三組銷售價格為 15-20 時，成交次數最多  
 (C) 第四組銷售價格為 20-25 時，成交次數最多  
 (D) 第五組銷售價格為 30-35 時，成交次數最多

ACD

03-05.

如附圖有關  $f(x)$  為一卜瓦松機率函數 (Poisson distribution)，下列哪些敘述為真？（請參閱附圖作答）(複選)

|                                                                              |                            |
|------------------------------------------------------------------------------|----------------------------|
| $f(x) = \frac{e^{-\lambda x} \cdot \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$ |                            |
| A 選項                                                                         | $\lambda$ 為單位時間內，事件發生的平均次數 |
| B 選項                                                                         | 期望值為 $2\lambda$            |
| C 選項                                                                         | 變異數為 $\lambda$             |
| D 選項                                                                         | 隨機變數 $x$ 表示事件發生的次數         |

- (A) 選項 A  
 (B) 選項 B  
 (C) 選項 C  
 (D) 選項 D

### 第三章 試題彙編

- ABCD      03-06.      如附圖針對下列二項分配公式， $p$  為單一試驗中成功的機率，請問以下選項哪些是正確的？（請參閱附圖作答）（複選）

$$f(x) = C_x^n p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

- (A)  $n$  為試驗次數  
 (B)  $x$  為  $n$  次試驗中，成功的次數  
 (C)  $n-x$  為  $n$  次試驗中，失敗的次數  
 (D)  $f(x)$  為  $n$  次試驗中發生  $x$  次成功的機率
- B      03-07.      已知一組樣本大小為 3 的資料，其樣本平均數為 2，樣本標準差為 1，如果要評斷此組資料與母體平均數 0 之間的差異程度，其統計量  $t$  值為何？（計算至小數點後二位，四捨五入）
- (A) 1.33  
 (B) 3.46  
 (C) 2.12  
 (D) -1.44
- B      03-08.      美國大型連鎖超商曾對顧客購物清單進行探勘，發現啤酒與尿布同時出現的比例很高，難道尿布的使用者不喝奶粉，這份調查反映了兩者間為何？
- (A) 因果關係  
 (B) 相關性  
 (C) 毫無關係  
 (D) 差異性
- B      03-09.      迴歸模式之整體配適情況，可由下列何種係數得知？
- (A) 相關係數  
 (B) 判定係數  
 (C) 共變異數  
 (D) 中位數
- D      03-10.      執行迴歸分析時的前置作業，以下何者為非？
- (A) 驗證常態性  
 (B) 驗證一致性  
 (C) 驗證獨立性  
 (D) 驗證相關性

## 企業電子化人才能力鑑定

2021/1/1

AC

03-11.

下列關於卡方分配的敘述哪些正確？(複選)

- (A) 在大於等於 0 (正數) 範圍的右偏分配，不同的自由度決定不同的卡方分配
- (B) 自由度趨近無窮大，其分配不會趨近於常態分配
- (C) 自由度趨近無窮大，其分配會趨近於常態分配
- (D) 其機率密度曲線呈左右對稱的鐘型曲線

B

03-12.

如附圖某資料的分配呈卡方分配，在顯著水準 ( $p$ ) 為 0.05，且自由度 ( $df$ ) 為 11 的情況下，請查下列附表，找出卡方分配的臨界值為何？(請參閱附圖作答)

|      | $p < 0.1$ |  | $p < 0.05$ | $p < 0.01$ | $p < 0.001$ |
|------|-----------|--|------------|------------|-------------|
| $df$ | .10       |  | .05        | .01        | .001        |
| 1    | 2.71      |  | 3.84       | 6.64       | 10.83       |
| 2    | 4.60      |  | 5.99       | 9.21       | 13.82       |
| 3    | 6.25      |  | 7.81       | 11.34      | 16.27       |
| 4    | 7.78      |  | 9.49       | 13.28      | 18.47       |
| 5    | 9.24      |  | 11.07      | 15.09      | 20.52       |
| 6    | 10.64     |  | 12.59      | 16.81      | 22.46       |
| 7    | 12.02     |  | 14.07      | 18.48      | 24.32       |
| 8    | 13.36     |  | 15.51      | 20.09      | 26.12       |
| 9    | 14.68     |  | 16.92      | 21.67      | 27.88       |
| 10   | 15.99     |  | 18.31      | 23.21      | 29.59       |
| 11   | 17.28     |  | 19.68      | 24.72      | 31.26       |
| 12   | 18.55     |  | 21.03      | 26.22      | 32.91       |
| 13   | 19.81     |  | 22.36      | 27.69      | 34.53       |
| 14   | 21.06     |  | 23.68      | 29.14      | 36.12       |
| 15   | 22.31     |  | 25.00      | 30.58      | 37.70       |

- (A) 21.03
- (B) 19.68
- (C) 22.36
- (D) 23.68

D

03-13.

四分位距 (Interquartile Range, IQR) 指的是中間多少比例的資料？

- (A) 20 %
- (B) 30 %
- (C) 40 %
- (D) 50 %



## 第三章 試題彙編

- A      03-14.      已知某資料服從卡方分配，自由度為 20，試算該資料期望值與變異數？
- (A) 期望值為 20，變異數為 40
  - (B) 期望值為 40，變異數為 20
  - (C) 期望值為 40，變異數為 40
  - (D) 期望值為 20，變異數為 20
- A      03-15.      下列何者不是用來衡量集中趨勢的統計量？
- (A) 變異數
  - (B) 眾數
  - (C) 平均數
  - (D) 中位數
- ABCD      03-16.      下列哪些是用來衡量分散程度的統計量？(複選)
- (A) 變異係數
  - (B) 四分位差
  - (C) 全距
  - (D) 標準差
- A      03-17.      某校高二有四個班級，甲班同學身高平均值為 168 公分，標準差為 6.8 公分；乙班同學身高平均值為 169 公分，標準差為 6.5 公分；丙班同學身高平均值為 160 公分，標準差為 5 公分；丁班同學身高平均值為 165 公分，標準差為 6 公分。請問哪一個班級學生的身高較為懸殊？
- (A) 甲班
  - (B) 乙班
  - (C) 丙班
  - (D) 丁班
- D      03-18.      下列抽樣方法，何項不屬於隨機抽樣？
- (A) 簡單隨機抽樣
  - (B) 系統抽樣
  - (C) 分層抽樣
  - (D) 滾雪球抽樣

## 企業電子化人才能力鑑定

2021/1/1

- B      03-19.      下列抽樣方法，何項不屬於離散機率分配？  
 (A) 二項分配  
 (B) 指數分配  
 (C) 卜瓦松分配  
 (D) 超幾何分配
- C      03-20.      關於信賴區間的描述，下列敘述何者正確？  
 (A) 當樣本數越大，信賴區間長度越長  
 (B) 重複兩次蒐集資料的過程，在相同樣本數的條件下，這兩次所求得信賴區間相同  
 (C) 當樣本數固定的情況下，信心水準越大，信賴區間長度越長  
 (D) 信賴區間長度與樣本數無關
- B      03-21.      某輪胎工廠宣稱其生產之輪胎平均壽命至少 50,000 公里。今某一汽車修理廠欲採購其輪胎，抽取 25 個輪胎檢測其平均壽命為 49,500 公里，標準差為 500 公里，試利用附圖資料計算，輪胎平均壽命在 95% 的信心水準下之信賴區間為何？（請參閱附圖作答）

假設輪胎的壽命為常態分配：

$$t_{24,0.05} = 1.711, t_{24,0.025} = 2.064, t_{25,0.05} = 1.708, t_{25,0.025} = 2.060$$

- (A) (49328.9, 49671.1)  
 (B) (49293.6, 49706.4)  
 (C) (49329.2, 49670.8)  
 (D) (49294, 49706)

- A      03-22.      關於統計假設檢定，如附圖敘述哪一項有誤？（請參閱附圖作答）

|      |                                                          |
|------|----------------------------------------------------------|
| A 選項 | 樣本大小不會影響檢定結果                                             |
| B 選項 | 當虛無假設 ( $H_0$ ) 為真，檢定結果為拒絕 $H_0$ ，則犯下型 I 誤差              |
| C 選項 | 發生型 I 誤差的最高機率稱之為顯著水準                                     |
| D 選項 | 在顯著水準 $\alpha = 0.1$ 下，若 P 值為 0.008，則結論為拒絕虛無假設 ( $H_0$ ) |

- (A) 選項 (A)  
 (B) 選項 (B)  
 (C) 選項 (C)  
 (D) 選項 (D)

### 第三章 試題彙編

- B      03-23.      關於變異數分析 (analysis of variance, ANOVA) 的假設，下列何者錯誤？
- (A) 各組資料抽自常態母體
  - (B) 各組資料的母體平均數均相等
  - (C) 各組資料的母體變異數均相等
  - (D) 各組資料的母體分配彼此相互獨立
- D      03-24.      進行變異數分析可以得知若干個母體平均數是否相等，若檢定結果拒絕虛無假設時，我們可以進行多重比較 (multiple comparisons)，以瞭解哪些母體平均數不同，請問下列哪一項不是多重比較的技術？
- (A) Bonferroni test
  - (B) Dunnett's test
  - (C) Tukey's test
  - (D) Goodness of fit test
- B      03-25.      某位老師想瞭解四種教學法與學生成績的關係，試問你（妳）會建議使用下列何種統計方法？
- (A) 迴歸分析
  - (B) 變異數分析
  - (C) Z 檢定
  - (D) 卡方檢定
- D      03-26.      在變異數分析問題中，包含 4 個處理且每個處理有 10 個觀察值，若誤差平方和 (SSE) 為 352.8，則誤差均方和 (MSE) 為何？
- (A) 88.2
  - (B) 9.05
  - (C) 35.28
  - (D) 9.8
- AC      03-27.      關於假設檢定中的 P 值，下列描述中哪些選項是正確的？(複選)
- (A) 如果 P 值很小，可以接受對立假設
  - (B) 如果 P 值很小，可以接受虛無假設
  - (C) P 值衡量了 型 I 誤差的概率
  - (D) P 值衡量了 型 II 誤差的概率

## 企業電子化人才能力鑑定

2021/1/1

- C      03-28.      隨著經濟增長，居民收入也在增長，同時貧富差距也增大了，此種現象可以使用第 20 百分位數和第 80 百分位數來表示，假設某年的居民收入資料如下：975，8040，57720，13457，7990，3810，1709，2400，785，2679；那麼對應的第 20 百分位數和第 80 百分位數分別為何？
- (A) 8040，785
  - (B) 785，8040
  - (C) 975，13457
  - (D) 1790，13457
- C      03-29.      顯著性檢驗常用來判斷某現象是否因為機遇而碰巧發生，關於顯著性檢驗，下列敘述何者有誤？
- (A) 做檢驗前先寫出虛無假設 (null hypothesis)，這個假設是表明，你所尋求的現象在總體中並不存在，而對立假設 (alternative hypothesis) 陳述的是該現象存在
  - (B) P 值是在虛無假設 (null hypothesis) 為真的情況下計算得到的
  - (C) 如果設置顯著水準 (significance level) 為 0.05，而計算出來的 P 值為 0.01，那麼我們接受虛無假設 (null hypothesis)，認為虛無假設正確
  - (D) 如果設置顯著水準 (significance level) 為 0.05，而計算出來的 P 值為 0.02，那麼我們拒絕虛無假設 (null hypothesis)，認為對立假設 (alternative hypothesis) 正確
- C      03-30.      在 5 個始祖鳥的化石標本中，記錄了股骨與肱骨的長度資料分別是，股骨：38、56、59、64、74，肱骨：41、63、70、72、84；使用相關係數分析這組資料，下列敘述何者正確？
- (A) 相關係數為 0.794，說明股骨和肱骨長度正相關
  - (B) 相關係數為 0.794，說明股骨和肱骨長度負相關
  - (C) 相關係數為 0.994，說明股骨和肱骨長度正相關
  - (D) 相關係數為 0.994，說明股骨和肱骨長度負相關
- B      03-31.      下列平均絕對離差 (Mean Absolute Deviation, MAD) 的敘述何者正確？
- (A) 其值越大，則資料的分散情況也越小
  - (B) 觀念簡單，易於了解
  - (C) 當資料數較多時，計算十分容易
  - (D) 不易受極端值的影響

### 第三章 試題彙編

- BCD      03-32.      下列有關變異係數 (Coefficient of Variance, CV) 的敘述哪些正確？(複選)
- (A) 變異係數等於標準差乘以平均數
  - (B) 變異係數是無單位係數
  - (C) 變異係數適用於單位相同但數值相差懸殊的資料欲比較分散程度
  - (D) 變異係數適用於不同單位的兩組或兩組以上資料欲比較分散程度
- BD      03-33.      盒鬚圖 (box plot) 能顯示以下哪些統計量？(複選)
- (A) 平均數
  - (B) 中位數
  - (C) 變異數
  - (D) 第 3 四分位數 (Q3)
- C      03-34.      下列抽樣調查的敘述何者正確？
- (A) 隨機抽樣中每組樣本彼此不是獨立的
  - (B) 若對母體的每一個元素進行調查，稱之為抽樣
  - (C) 抽樣是指利用具有代表性的樣本資料，藉此分析欲研究母體的一些特性
  - (D) 抽樣的重要性不包括節省成本
- D      03-35.      若隨機變數  $X$  具有二項分配 (Binomial Distribution)，成功的機率為 0.2，試驗次數為 10，則期望值與變異數分別為？
- (A) 期望值 8，變異數 2.5
  - (B) 期望值 8，變異數 1.6
  - (C) 期望值 2，變異數 2.5
  - (D) 期望值 2，變異數 1.6
- B      03-36.      若隨機變數  $X$  具有卜瓦松分配 (Poisson Distribution)，單位時間內，事件發生的平均次數為 5，則期望值與變異數分別為？
- (A) 期望值 5，變異數 10
  - (B) 期望值 5，變異數 5
  - (C) 期望值 10，變異數 5
  - (D) 期望值 10，變異數 10

## 企業電子化人才能力鑑定

2021/1/1

- BD      03-37.      下列推論統計敘述哪些正確？(複選)
- (A) 虛無假設 (null hypothesis)：對母體的未知參數所作之假設，以  $H_1$  表示之
- (B) 型 I 誤差：當  $H_0$  為真，檢定結果為拒絕  $H_0$ ，所犯的錯誤
- (C) 發生型 II 誤差之最大機率，稱為顯著水準 (significance level)
- (D) 型 II 誤差：當  $H_0$  為誤，檢定結果為接受  $H_0$ ，所犯的錯誤

- C      03-38.      請參閱附圖作答：

|                                                                                               |                                              |
|-----------------------------------------------------------------------------------------------|----------------------------------------------|
| 假設 $X_1, X_2, \dots, X_n$ 為一組抽自 $N(\mu, \sigma^2)$ 的隨機樣本，則 $(n-1) s^2 / \sigma^2$ 統計量的敘述何者正確？ |                                              |
| A 選項                                                                                          | 自由度為 $n$ 之 $t$ 分配 ( $t$ Distribution)        |
| B 選項                                                                                          | 自由度為 $(n-1)$ 之 $t$ 分配 ( $t$ Distribution)    |
| C 選項                                                                                          | 自由度為 $(n-1)$ 之卡方分配 (Chi-Square Distribution) |
| D 選項                                                                                          | 自由度為 $n$ 之卡方分配 (Chi-Square Distribution)     |

- (A) 選項 (A)
- (B) 選項 (B)
- (C) 選項 (C)
- (D) 選項 (D)
- A      03-39.      下列有關變異數分析的敘述，何者正確？
- (A) 變異數分析主要為探討連續型資料型態之因變數與類別型資料型態之自變數的關係
- (B) 變異數分析是檢定各類別間變異數是否相等的問題
- (C) 平方和會隨試驗單位的個數增加而減少
- (D) 總變異 = 組內變異 - 組間變異
- CD      03-40.      下列有關迴歸分析的敘述，哪些正確？(複選)
- (A) 判定係數 =  $1 - SSR / SST$
- (B)  $-1 \leq \text{判定係數} \leq 1$
- (C) 判定係數越大，表示迴歸模式之配適度越高
- (D) 迴歸分析可採用最小平方法

## 第三章 試題彙編

BD

03-41.

依附圖有關迴歸分析的敘述，哪些正確？（請參閱附圖作答）  
（複選）

|      |                                                                        |
|------|------------------------------------------------------------------------|
| 題幹   | 以下迴歸模型<br>$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i=1,2,\dots,n$ |
| A 選項 | 如果隨機變數 $Y$ 是分類變數，可以使用線性迴歸模式                                            |
| B 選項 | $\beta_1$ 統計假設是： $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$          |
| C 選項 | 隨機變數 $Y$ 是自變數 (Independent Variable)                                   |
| D 選項 | $\varepsilon$ 項目是誤差項                                                   |

(A) 選項 (A)

(B) 選項 (B)

(C) 選項 (C)

(D) 選項 (D)

C

03-42.

如附圖有關迴歸分析的敘述，何者正確？（請參閱附圖作答）

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                               |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|
| <pre>Call: lm(formula = rating ~ complaints, data = attitude)  Residuals:     Min       1Q   Median       3Q      Max -12.8799  -5.9905   0.1783   6.2978   9.6294  Coefficients:             Estimate Std. Error t value Pr(&gt; t ) (Intercept) 14.37632     6.61999   2.172  0.0385 * complaints   0.75461     0.09753   7.737 1.99e-08 *** --- Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  Residual standard error: 6.993 on 28 degrees of freedom Multiple R-squared:  0.6813, Adjusted R-squared:  0.6699 F-statistic: 59.86 on 1 and 28 DF, p-value: 1.988e-08</pre> |                                                               |
| A 選項                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 截距項達到 $\alpha=0.01$ 的顯著水準                                     |
| B 選項                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 迴歸模型為 $\text{complaints} = \beta_0 + \beta_1 * \text{rating}$ |
| C 選項                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 樣本個數為 30                                                      |
| D 選項                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | $R^2$ 等於 0.6699                                               |

(A) 選項 (A)

(B) 選項 (B)

(C) 選項 (C)

(D) 選項 (D)

## 企業電子化人才能力鑑定

2021/1/1

B 03-43. 有關判定係數 (coefficient of determination) 的描述何者正確？

- (A) 判定係數介於 -1 跟 1 之間
- (B) 判定係數越高，表示迴歸模式之預測值越接近觀察值
- (C) 當模型中解釋變數增加時，判定係數會降低
- (D) 判定係數越接近 1 代表迴歸線的解釋力越小

D 03-44. 簡單隨機抽樣的樣本結果，其單個觀測值與樣本平均數都有對應的抽樣分佈，關於它們的抽樣分佈，下列敘述何者有誤？

- (A) 許多樣本計算出的平均值都是以母體平均值  $\mu$  為中心
- (B) 一些觀測值平均數的分佈，要比個別觀測值的分佈更接近常態分佈
- (C) 如果從母體隨機抽取愈來愈多的觀測值，則這些觀測值平均數的分佈，遲早會接近常態分佈
- (D) 只有在樣本數足夠大的情況下，樣本平均數的抽樣分佈才會是近似常態分佈

BC 03-45. 有關附圖模型及選項，下列哪些敘述正確？（請參閱附圖作答）  
（複選）

|      |                                             |
|------|---------------------------------------------|
| 題幹   | 迴歸模型 $y = -0.1x + 1.4$<br>判定係數 $r^2 = 0.87$ |
| A 選項 | x 值為 1 時，可以預測 y 值為 2.3                      |
| B 選項 | x 每變動 1 個單位，預測的 y 反方向變動 0.1 個單位             |
| C 選項 | $r^2 = 0.87$ ，x 構成的迴歸模型可以解釋 87% 的 y 的變異     |
| D 選項 | 斜率為負，說明變量間正相關                               |

- (A) 選項 (A)
- (B) 選項 (B)
- (C) 選項 (C)
- (D) 選項 (D)

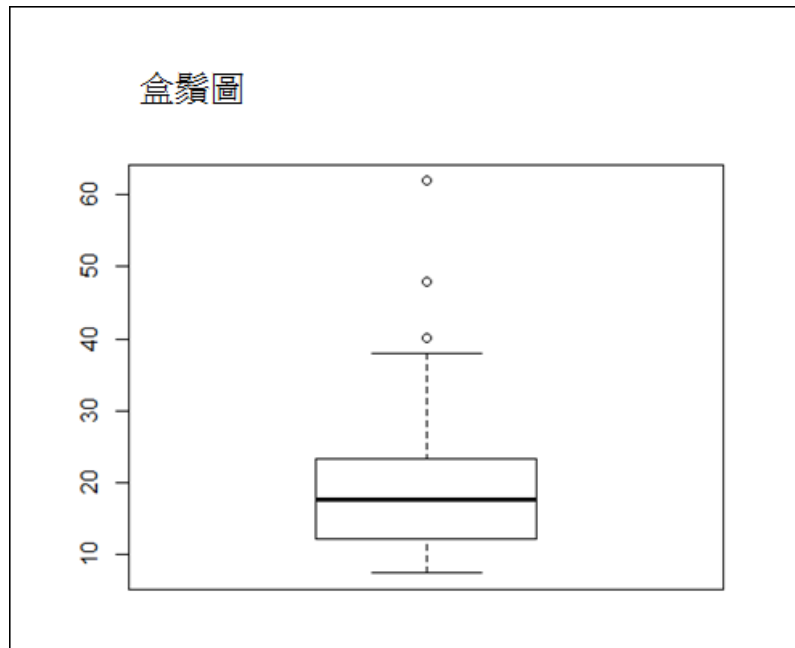
A 03-46. 下列相關分析的敘述，何者為正確？

- (A) 二隨機變數具有相關性，但是不一定存在因果關係
- (B) 散佈圖 (scatter plot) 無法反映兩變數之間的相互關係
- (C) 共變異數 (covariance) 範圍會介於 -1 與 1 之間
- (D) 相關係數為零，表示兩變數間無任何相關關係



## 第三章 試題彙編

- D      03-47.      以下哪個選項不是平均數的優點？  
 (A) 考慮了所有觀察值，敏感度較高  
 (B) 計算簡單易於了解  
 (C) 適合代數與統計推論計算  
 (D) 其結果具有代表性
- C      03-48.      下列四個數值 1、2、5、6 的變異係數為多少？  
 (A) 3.5  
 (B) 2.38  
 (C) 0.68  
 (D) 5.67
- A      03-49.      下列盒鬚圖中，哪一項描述不正確？（請參閱附圖作答）



- (A) 該資料的均值約為 18  
 (B) 該資料具有 3 個極端值  
 (C) 該資料的最小值小於 10  
 (D) 該資料的最大值約為 60
- ABD      03-50.      下列關於直方圖與長條圖的敘述哪些正確？（複選）  
 (A) 直方圖的長條是連續且緊密貼近的，長條圖相反  
 (B) 長條圖的位置沒有一定的順序，直方圖相反  
 (C) 長條圖是量化資料的視覺化方法，直方圖是類別資料的視覺化方法  
 (D) 直方圖的 Y 軸可以是比例

## 企業電子化人才能力鑑定

2021/1/1

- C            03-51.    下述常用的抽樣方法，哪一項不是隨機抽樣？
- （A）系統抽樣：將母體編號隨機抽取一個號碼後，隔若干號取出下一個樣本，直到所需樣本數為止
- （B）集群抽樣：將母體分群然後隨機抽取出若干群，抽到之群每個都要被調查
- （C）配額抽樣：將母體分成若干層根據比例決定各層樣本數，研究者在每層配額內選擇樣本
- （D）分層抽樣：將母體分為若干層再對各層進行簡單隨機抽樣
- D            03-52.    丟擲三個公正銅板，其中有兩個是正面的機率為多少？
- （A）0.5
- （B）0.125
- （C）0.25
- （D）0.375
- C            03-53.    關於常態分配，如附圖以下哪一項描述不正確？（請參閱附圖作答）

|      |                               |
|------|-------------------------------|
| A 選項 | 常態分配通常以 $N(\mu, \sigma^2)$ 表示 |
| B 選項 | 常態分配的平均數、中位數和眾數相等             |
| C 選項 | $\mu$ 越大，分配圖越右偏               |
| D 選項 | $\sigma$ 越大，分配圖越扁             |

- （A）選項（A）
- （B）選項（B）
- （C）選項（C）
- （D）選項（D）
- B            03-54.    對於 " 95% 信賴區間 " 的理解，哪一項為正確的？
- （A）真實值乘以 1.05 不會大於信賴區間的上限，真實值乘以 0.95 不會小於信賴區間的下限
- （B）計算 100 個信賴區間，有大約 95 個包含真實值
- （C）說明了抽樣誤差是 5%
- （D）樣本數目越多，95% 信賴區間的範圍越大

### 第三章 試題彙編

- D      03-55.      以下有關相關性的描述，哪一項為因果關係？
- （A）買了啤酒要買尿布
- （B）天蠍座的人更容易成功
- （C）喝紅酒可以延緩衰老
- （D）下雨要撐傘

- C      03-56.      如附圖選項，下列何者不是簡單線性迴歸分析的基本統計假設？  
（請參閱附圖作答）

|      |                         |
|------|-------------------------|
| A 選項 | $\epsilon_i$ 服從常態分配     |
| B 選項 | $\epsilon_i$ 的期望值為 0    |
| C 選項 | $\epsilon_i$ 的變異數為 1    |
| D 選項 | $\epsilon_i$ 彼此之間不存在相關性 |

- （A）選項（A）
- （B）選項（B）
- （C）選項（C）
- （D）選項（D）
- A      03-57.      關於迴歸分析的判定係數（coefficient of determination），哪一項描述是正確的？
- （A）判定係數越高表示迴歸模式之配適度越高
- （B）判定係數高表示模型未來的預測會很準確
- （C）判定係數低表示模型的參數不顯著
- （D）判定係數低表示模型誤差比較大
- A      03-58.      下列有關變異數分析的描述，哪一項是正確的？
- （A）變異數分析需要假定獨立性、常態性、變異數齊一性
- （B）變異數分析檢定各類別間中位數是否相等
- （C）假設存在三個類別，如果 F 檢定 P 值小於 0.05，說明在該顯著性水準下，三個類別之間沒有顯著差異
- （D）假設存在三個類別，如果 F 檢定 P 值小於 0.05，說明在該顯著性水準下，三個類別分別存在顯著差異

## 企業電子化人才能力鑑定

2021/1/1

C 03-59. 附圖模型中，哪個選項為羅吉斯迴歸？（請參閱附圖作答）

|      |                                                                          |
|------|--------------------------------------------------------------------------|
| A 選項 | $Y \sim a \cdot X_1 + b \cdot X_2 + c \cdot X_1 \cdot X_2 + \varepsilon$ |
| B 選項 | $Y \sim a \cdot X_1^2 + b \cdot X_2 + \varepsilon$                       |
| C 選項 | $Y \sim a \cdot X_1 + b \cdot X_2 + \varepsilon$ ，其中 $Y=0$ 或者 $Y=1$      |
| D 選項 | $Y \sim a \cdot X_1 + b \cdot X_2 + \varepsilon$ ，其中 $X_1=0$ 或者 $X_1=1$  |

（A）選項（A）

（B）選項（B）

（C）選項（C）

（D）選項（D）

C 03-60. 我們常用統計推論中的 P 值做為資料分析的一種結果，下列關於 P 值的敘述何者正確？

（A）P 值檢定的結果和臨界值檢定的結果不會一致

（B）信賴區間不能用在假設檢定問題中

（C）P 值檢定的結果和臨界值檢定的結果是一致的

（D）在假設檢定問題的計算中，Z 值檢定和 t 值檢定是完全一樣的

D 03-61. 下列關於卡方分配的描述，哪一項是錯誤的？

（A）卡方分配是右偏分配

（B）卡方分配只有一個參數：自由度

（C）卡方分配的自由度接近無窮大時，卡方分配趨近於常態分配

（D）常態分配的隨機變數平方之後服從卡方分配

A 03-62. 一支股票的總獲利是用買入時股價的百分比來表示，要了解一年中紐約證券交易所中所有掛牌的 1,528 支股票總獲利的分佈情況，應該使用下列哪個圖形？

（A）直方圖

（B）莖葉圖

（C）長條圖

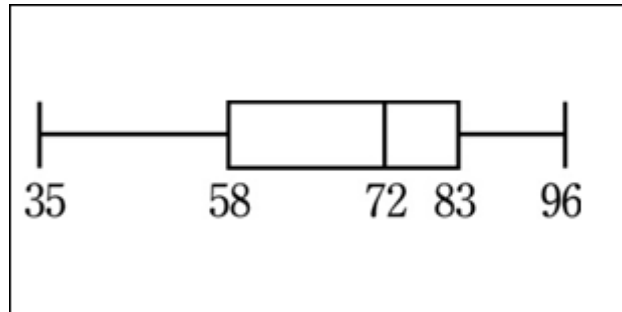
（D）圓餅圖

### 第三章 試題彙編

ABD

03-63.

某位老師將班上同學的數學期中考成績繪出如附圖的盒鬚圖 (Box Plot)，藉由此盒鬚圖我們可以得知此班級數學期中考成績的哪些資訊？（請參閱附圖作答）（複選）



- (A) 班上同學數學考最高者為 96 分
- (B) 數學期中考分數的第 3 四分位數  $Q_3 = 83$  分
- (C) 數學期中考分數的中位數為 70.5 分
- (D) 數學期中考分數的四分位距為 25 分

ABD

03-64.

下列哪些敘述說明類別資料通常如何產生？（複選）

- (A) 順序尺度的資料中獲得
- (B) 區間尺度的資料正規化後獲得
- (C) 對連續資料分組化簡獲得
- (D) 通過離散機率分配來隨機產生

BC

03-65.

關於中位數的描述，下列哪些選項有誤？（複選）

- (A) 不受極端值影響
- (B) 中位數對觀察值敏感度較高
- (C) 若有極端值存在時則不具代表性
- (D) 中位數對觀察值缺乏敏感度

B

03-66.

下列何者不為常見的分散度衡量統計量？

- (A) 變異係數
- (B) 眾數
- (C) 平均絕對差
- (D) 全距

## 企業電子化人才能力鑑定

2021/1/1

- D 03-67. 有關於平均絕對離差（MAD）的描述，下列何者正確？
- （A）敏感度較低，代表性較佳
  - （B）會考慮各均差之正負號
  - （C）其值越小，則資料的分散情況也越大
  - （D）當資料數多時，較不易計算

- BC 03-68. 下列對於長條圖的敘述哪些正確？（複選）
- （A）長條圖是表示數值資料最常用的統計圖
  - （B）長條圖是表示質性資料最常用的統計圖
  - （C）長條圖中的位置並沒有一定的順序
  - （D）長條圖的長條是連續且緊密貼近的

- D 03-69. 一個公正的骰子，連續且獨立的丟擲 18 次，隨機變數  $X$  代表出現點數為 3 的次數，附圖選項何者正確？（請參閱附圖作答）

|      |                |
|------|----------------|
| A 選項 | $X$ 的期望值為 6    |
| B 選項 | $X^2$ 的期望值為 36 |
| C 選項 | $X$ 的變異數為 1.58 |
| D 選項 | $X$ 的標準差為 1.58 |

- （A）選項（A）
  - （B）選項（B）
  - （C）選項（C）
  - （D）選項（D）
- C 03-70. 下列對於卜瓦松分配（Poisson distribution）的敘述何者有誤？
- （A）在特定區間內，事件發生的平均次數固定
  - （B）事件發生的機率與區間的起點無關
  - （C）在任何區間內，事件發生彼此會互相影響
  - （D）在特定區間內，事件發生兩次以上的機率為 0
- D 03-71. 以下何者不是隨機抽樣方法？
- （A）集群抽樣
  - （B）系統抽樣
  - （C）分層抽樣
  - （D）滾雪球抽樣

### 第三章 試題彙編

- C 03-72. 如附圖在統計檢定中，該理論的虛無假設為下列何者？（請參閱附圖作答）

若一研究人員認為在某特定行業的平均薪水( $\mu$ )超過 4 萬元，並希望證明他的理論是正確的。

|      |                        |
|------|------------------------|
| A 選項 | $H_0: \mu > 40,000$    |
| B 選項 | $H_0: \mu \geq 40,000$ |
| C 選項 | $H_0: \mu \leq 40,000$ |
| D 選項 | $H_0: \mu < 40,000$    |

- (A) 選項 A  
(B) 選項 B  
(C) 選項 C  
(D) 選項 D

- C 03-73. 關於顯著水準的敘述，下列何者有誤？
- (A) 顯著水準越大，越可能拒絕虛無假設  
(B) 顯著水準是發生型 I 誤差的最大機率  
(C) 顯著水準又稱為型 II 誤差  
(D) 給定的顯著水準下，樣本量增加會造成型 II 誤差的機率降低

- A 03-74. 執行迴歸分析時理想上須先驗證資料是否遵守先決條件，下列敘述何者錯誤？
- (A) 以 Normal probability plot 驗證獨立性  
(B) 以 Q-Q plot 驗證常態性  
(C) 以 Modified Levene Test 驗證一致性  
(D) 以 Residual time sequence plot 驗證獨立性

- B 03-75. 下列有關變異數分析的敘述何者正確？
- (A) 用以檢定各類別間的變異數是否相等  
(B) 平方和會隨試驗單位的個數增加而增加  
(C) 樣本觀察值算出的統計量之值大於 F 分配時，接受虛無假設  
(D) 主要探討連續型資料型態之自變數與類別型資料型態之因變數的關係

## 企業電子化人才能力鑑定

2021/1/1

AB 03-76. 假設檢定的兩種誤差為附圖中哪些？（請參閱附圖作答）（複選）

|      |                                 |
|------|---------------------------------|
| A 選項 | 當 $H_0$ 為真，檢定結果為拒絕 $H_0$ ，所犯的錯誤 |
| B 選項 | 當 $H_0$ 為誤，檢定結果為接受 $H_0$ ，所犯的錯誤 |
| C 選項 | 當 $H_0$ 為真，檢定結果為拒絕 $H_1$ ，所犯的錯誤 |
| D 選項 | 當 $H_0$ 為誤，檢定結果為接受 $H_1$ ，所犯的錯誤 |

（A）選項（A）

（B）選項（B）

（C）選項（C）

（D）選項（D）

BCD 03-77. 圖形是傳達資料資訊最有效的方式，下列哪些原則能夠畫出更好的圖？（複選）

（A）使用象形圖

（B）讓資料很醒目

（C）在標示和說明裡面將圖內的變量、單位以及資料來源表述清楚

（D）謹慎選擇刻度

ABD 03-78. 下列哪些為迴歸分析中對於誤差項的假設？（複選）

（A）服從常態分配

（B）變異數齊質性

（C）線性關係

（D）無關性

ABC 03-79. 常用的統計值中，平均數的優點有以下哪幾項？（複選）

（A）資料無極端值或偏態時，具有代表性

（B）考慮所有觀察值，敏感度較高

（C）適合代數與統計推論計算

（D）不容易受極端值影響

C 03-80. 下列敘述何者有誤？

（A）變異數值恆大於或等於零

（B）變異數其值越大代表資料分散程度越大

（C）變異數可用於比較不同單位之資料間分散程度的大小

（D）變異數其值越小代表資料分散程度越小



### 第三章 試題彙編

- D      03-81.      計算各觀察值與平均數的距離總和，再取其算術平均數之統計值為以下何者？
- (A) 中位數 (median)
- (B) 四分位距 (Interquartile range, IQR)
- (C) 變異數 (variance)
- (D) 平均絕對離差 (mean absolute deviation, MAD)
- C      03-82.      請參閱附圖作答：
- 下列何種統計圖表可顯示數值資料最小值、第 1 四分位數 ( $Q_1$ )、中位數 ( $Q_2$ )、第 3 四分位數 ( $Q_3$ )、最大值的分散程度？
- (A) 圓餅圖 (Pie Chart)
- (B) 次數分配表 (Frequency Table)
- (C) 盒鬚圖 (Box plot)
- (D) 直方圖 (Histogram)
- B      03-83.      繪製直方圖前，將數值資料分組，計算各組發生次數整理而成的表格，稱為以下何者？
- (A) 索引 (Index Table)
- (B) 次數分配表 (Frequency Table)
- (C) 雜湊表 (Hash Table)
- (D) 分類表 (Class Table)
- D      03-84.      下列針對常態分配 (Normal distribution) 的敘述，何者不正確？
- (A) 常態分配又稱作高斯分配 (Gaussian distribution)
- (B) 常態分配其分配圖形為一個對稱鐘型
- (C) 在大樣本下，可以利用常態分配近似一些離散機率
- (D) 常態分配之平均數、中位數，和變異數相等
- D      03-85.      以下何者不是 Poisson 實驗的性質？
- (A) 在特定區間內，某事件發生的平均次數固定
- (B) 在特定區間內，某事件發生兩次以上的機率為 0
- (C) 在任何區間內，事件發生彼此不會互相影響
- (D) 在特定區間內，事件發生彼此會互相影響

## 企業電子化人才能力鑑定

2021/1/1

- D            03-86.    下列何者敘述不正確？
- (A) 抽樣是指利用具有代表性的樣本資料，藉此分析欲研究母體的一些特性
  - (B) 若對母體的每一個元素進行調查，稱之為普查 (census)
  - (C) 常用的抽樣方法可以分為：隨機抽樣、非隨機抽樣
  - (D) 非隨機抽樣中，母體中所有單位或元素皆有可能被選到
- B            03-87.    根據平均數和中位數的定義，下列有幾項說法是正確的？(1) 比較不同教育程度者的收入，平均數比中位數更合適、(2) 可以使用平均數比較兩個班級的總成績、(3) 一次射擊訓練中，甲的平均射擊環數比乙高，那麼甲一定比乙優秀、(4) 在資料偏斜程度較大時，中位數比平均數更能代表整體
- (A) 1
  - (B) 2
  - (C) 3
  - (D) 4
- B            03-88.    以下何者為理想上的迴歸分析運作流程？
- (A) 驗證資料符合模型前提假設 → 確認變數之間具有線性關係 → 執行迴歸分析 → 評估模型配適是否合適，適度校正模型
  - (B) 確認變數之間具有線性關係 → 驗證資料符合模型前提假設 → 執行迴歸分析 → 評估模型配適是否合適，適度校正模型
  - (C) 執行迴歸分析 → 確認變數之間具有線性關係 → 驗證資料符合模型前提假設 → 評估模型配適是否合適，適度校正模型
  - (D) 執行迴歸分析 → 驗證資料符合模型前提假設 → 確認變數之間具有線性關係 → 評估模型配適是否合適，適度校正模型
- C            03-89.    下列針對迴歸分析 (Regression Analysis) 的敘述何者不正確？
- (A) 迴歸分析是以自變數 (independent variable) 來描述依變數 (dependent variable) 的統計關係模式
  - (B) 在執行迴歸分析前，理想上必須要確認資料是否符合模型的基本統計假設：常態性、一致性與獨立性
  - (C) 簡單線性迴歸模型中，誤差平方和 (Sum of squared errors, SSE) 越大則代表模型的預測能力越好
  - (D) 當應變數 Y 是分類變數，無法使用一般線性迴歸模式來分析資料時，可用羅吉斯迴歸 (Logistic Regression) 來分析

### 第三章 試題彙編

- B      03-90.      進行單一樣本平均數  $t$  檢定時，當資料中樣本數較大時，通常以樣本筆數大於多少作為區分標準？
- (A) 20  
(B) 30  
(C) 40  
(D) 50

- A      03-91.      附圖中列出自 91 年度起至 100 年度，每年的廣告經費支出及年度銷售金額（單位：百萬元）。分析 " 廣告支出是否會影響銷售金額 " 可採用之分析方法為下列何者？（請參閱附圖作答）

| 年度   | 91   | 92   | 93  | 94  | 95  | 96   | 97  | 98  | 99  | 100  |
|------|------|------|-----|-----|-----|------|-----|-----|-----|------|
| 廣告支出 | 16.5 | 15.7 | 15  | 16  | 18  | 19.6 | 14  | 13  | 14  | 19.2 |
| 銷售金額 | 230  | 215  | 218 | 235 | 240 | 259  | 215 | 220 | 220 | 243  |

- (A) 簡單迴歸分析 (Simple regression analysis)  
(B) 卡方適合度檢定 (Chi-square test for goodness of fit)  
(C) 單一樣本平均數  $t$  檢定 (one-sample  $t$ -test)  
(D) (獨立) 多樣本平均數差異檢定 (或稱變異數分析) (ANalysis Of VAriance, ANOVA)
- A      03-92.      有關附圖最小平方迴歸方程式，請計算出  $y$  的最後一個數值，並精確到小數點後一位，下列何者正確？（請參閱附圖作答）

已知兩個變量的最小平方迴歸方程式為  $y = 2.33x - 0.29$

$x$  的值為 2.2、3.4、2.5、3.7、4.2

$y$  的數值對應為 4.7、8.1、5.5、8.2、\_\_\_\_（最後一個數值遺缺了）

- (A) 9.3  
(B) 8.9  
(C) 9.8  
(D) 8.7

## 企業電子化人才能力鑑定

2021/1/1

- D 03-93. 抽樣調查中，誤差來源可以分為兩大類：抽樣誤差和非抽樣誤差，判斷以下情況哪些屬於非抽樣誤差的來源？（a）受訪對象隱瞞曾用過毒品的事實、（b）記錄資料時打字錯誤、（c）用一個地區的電話號碼簿當作抽樣結構進行抽樣調查、（d）訪問員在某條街上隨機找人進行訪問、（e）打了 5 次電話都聯絡不到受訪者
- （A）a、b、c  
 （B）b、c、e  
 （C）b、c、d  
 （D）a、b、e

- C 03-94. 關於迴歸分析，附圖敘述何者有誤？（請參閱附圖作答）

|      |                                                                              |
|------|------------------------------------------------------------------------------|
| A 選項 | 最小平方法得到的迴歸直線，是使得所有資料點距離直線的垂直距離平方和為最小的直線                                      |
| B 選項 | 迴歸分析是一種統計方法，這種方法以資料進行模型配適，以便根據一個或多個解釋變量來預測反應變量的值                             |
| C 選項 | 一個研究中，將不同國家的電視機數目作為解釋變量，對人民的預期壽命進行迴歸分析，發現迴歸直線斜率顯著為正，這說明在很少電視機的國家，會使得人民預期壽命更長 |
| D 選項 | 每個迴歸分析得到的方程式都會有判定係數 $r^2$ ，它代表被解釋變量的變異當中，可以用所有解釋變量的最小平方迴歸線來解釋的比例             |

- （A）選項（A）  
 （B）選項（B）  
 （C）選項（C）  
 （D）選項（D）

- D 03-95. 變異數分析是一種常用的統計方法，下列關於變異數分析之敘述何者有誤？
- （A）要研究一個因子是否對糧食產量有影響，可以在這個因子的不同水準下做實驗，獲得產量資料，再進行單因子變異數分析來判斷
- （B）多因子變異數分析可以用來判斷多個因子對某一變量是否有影響
- （C）多因子變異數分析還可以判斷因子間的交互作用項，是否對某一變量有影響
- （D）現有 A 變量在 B 因子的 4 個水準下的實驗資料，對之進行單因子變異數分析，統計檢驗的 p 值是 0.02，這說明在 0.05 的顯著性水準下，這 4 個水準下 A 的平均值互不相同，因此 B 因子對 A 變量有顯著影響

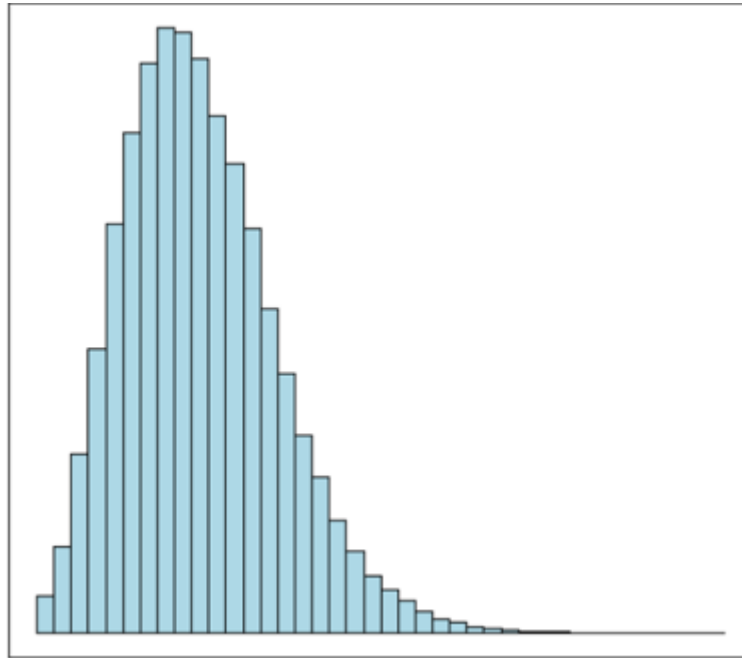
### 第三章 試題彙編

- D      03-96.      抽樣是指利用具有代表性的樣本資料，藉此分析欲研究母體的一些特性。下列何者非為隨機抽樣常用的方法？
- （A）隨機抽樣（Random Sampling）
- （B）系統性抽樣（Systematic Sampling）
- （C）分層抽樣（Stratified Sampling）
- （D）判斷抽樣（Judgmental Sampling）
- A      03-97.      描述或比較資料的離散程度常用三個指標：a 標準差，b 四分位距，c 變異係數。請選出以下情況最適合的指標：（1）有一個大公司和一個小公司，比較這兩個公司的業績，哪個波動更大；（2）用一周的訓練資料，比較兩個同一級別的射擊選手較能發揮穩定性；（3）現有兩個業績相近的公司給所有員工的薪資資料，比較哪個公司薪水整體差距更大？
- （A）（1）- c、（2）- a、（3）- b
- （B）（1）- a、（2）- b、（3）- c
- （C）（1）- c、（2）- b、（3）- a
- （D）（1）- a、（2）- c、（3）- b
- ACD      03-98.      在單因子變異數分析之中，考慮四大飲料品牌之平均銷售量分析，每一品牌隨機抽取 5 個樣本，下列敘述哪些正確？（複選）
- （A）組間變異 SSB <Sum of Squares of Between> 的自由度為 3
- （B）總變異的自由度為 20
- （C）隨機誤差的自由度為 16
- （D）如果 F 統計值大於 F 分配查表的臨界值時，則表示四大飲料品牌之平均銷售量不全相等
- D      03-99.      下列對於抽樣的敘述何者錯誤？
- （A）抽樣可以做為描繪資料特徵的工具之一
- （B）行政院主計處發布家庭收支調查資料為抽樣結果
- （C）系統抽樣是屬於隨機抽樣方法
- （D）分層抽樣是屬於非隨機抽樣方法

## 企業電子化人才能力鑑定

2021/1/1

C 03-100. 參考附圖，下列直方圖的敘述何者為正確？（請參閱附圖作答）



- (A) 資料呈現左偏偏斜
- (B) 眾數 > 平均數 > 中位數
- (C) 平均數 > 中位數 > 眾數
- (D) 中位數 > 平均數 > 眾數

### 第三章 試題彙編

DPA

「巨量資料處理與分析」學科試題及答案

#### 第四類：資料探勘與視覺化基礎

- ABCD      04-01.      購物籃分析所得出的關聯規則，下列選項哪些為真？(複選)
- (A) 可以提供有關產品間被同時購買的有用資訊
  - (B) 可用以調整商品擺設及交叉銷售的設計，藉以增加商店的銷售收入
  - (C) 可以進行相關產品推薦或者挑選相應的關聯產品進行精準行銷
  - (D) 可以尋找更多潛在的目標客戶
- D          04-02.      資料清理是指發現並糾正資料中的錯誤，下列何者非資料清理的方法？
- (A) 驗證資料的正確性
  - (B) 缺失值的處理
  - (C) 異常值的處理
  - (D) 迴歸係數的處理
- C          04-03.      把資料庫中的資料分類成群，是一種常見的統計分析技巧，其原理下列何者為非？
- (A) 群內差異小，群間差異大
  - (B) 依相似性將樣本分成幾群互相沒有交集的群組
  - (C) 群與群間有高相關性
  - (D) 同群組內的個體相似性高，不同群組的個體相似性低
- A          04-04.      下列哪一項不是用來評估關聯規則 (association rules) 的指標？
- (A) 可靠度 (reliability)
  - (B) 支持度 (support)
  - (C) 信賴度 (confidence)
  - (D) 增益 (lift)

## 企業電子化人才能力鑑定

2021/1/1

- A 04-05. 關於主成分分析，下列敘述何者有誤？
- (A) 主成分分析適用於分析變量的各種組合，包含線性和非線性的組合
  - (B) 主成分分析的優點是能夠實現資料的維度縮減
  - (C) 主成分分析之後，各主成分會失去原始變量的業務意義
  - (D) 主成分之間互不相關

- C 04-06. 在關聯分析中得到了如附圖之結果，下列選項何者正確？（請參閱附圖作答）

| lhs                   | rhs             | support     | confidence | lift     |
|-----------------------|-----------------|-------------|------------|----------|
| 1 {honey}             | => {whole milk} | 0.001118454 | 0.7333333  | 2.870009 |
| 2 {tidbits}           | => {rolls/buns} | 0.001220132 | 0.5217391  | 2.836542 |
| 3 {cocoa drinks}      | => {whole milk} | 0.001321810 | 0.5909091  | 2.312611 |
| 4 {pudding powder}    | => {whole milk} | 0.001321810 | 0.5652174  | 2.212062 |
| 5 {cooking chocolate} | => {whole milk} | 0.001321810 | 0.5200000  | 2.035097 |

- (A) 73.3% 的交易會同時購買 honey 和 whole milk
- (B) Whole Milk 是最受歡迎的商品
- (C) tidbits 和 rolls / buns 的消費會互相促進
- (D) 購買 pudding powder 的交易中有 0.13% 會購買 whole milk

- C 04-07. 如附圖的名稱為何，且一般用在何種分析上？（請參閱附圖作答）



- (A) 決策樹 (Decision Tree)，用在分類
- (B) K 平均法 (K-means)，用在分群
- (C) 文字雲 (Word Cloud)，用在文字探勘 (text mining)
- (D) 階層式分群 (hierarchical clustering)，用在分群



### 第三章 試題彙編

- B**      04-08.      下列哪一項不能用於比較決策樹的分類效果？
- (A) 通過對測試資料集的分類，得到分類準確率；然後透過統計計算，得到決策樹分類準確率的信賴區間 (confidence interval)，結合兩個一起判斷效果
  - (B) 用包含 10 個紀錄的測試資料集進行檢驗，準確率越高的效果越好
  - (C) 比較兩個決策樹分類效果時，檢驗兩樹準確率是否在統計上有顯著差別
  - (D) 用 k 折交叉驗證的方法比較分類效果
- C**      04-09.      關於決策樹的描述，下列何者錯誤？
- (A) 決策樹是功能強大的分類和預測工具
  - (B) 構建模型的成本低是決策樹的優點之一
  - (C) 決策樹可以輸出多個屬性
  - (D) 決策樹的結構是一種類似流程圖的樹結構
- C**      04-10.      請問購物籃分析 (market-basket analysis) 是運用哪一項資料探勘的技術？
- (A) 趨勢分析 (Trend Analysis)
  - (B) 迴歸分析 (Regression Analysis)
  - (C) 關聯分析 (Association Analysis)
  - (D) 集群分析 (Clustering Analysis)
- C**      04-11.      在 150 筆交易資料中，若想知道購買麵包的人會不會也購買牛奶，經由調查發現購買麵包的交易共有 80 筆，買麵包同時也會買牛奶的交易有 50 筆，下列關於支持度 (support) 和信賴度 (confidence) 的計算何者正確？
- (A) 支持度 53.3%、信賴度 33.3%
  - (B) 支持度 33.3%、信賴度 53.3%
  - (C) 支持度 33.3%、信賴度 62.5%
  - (D) 支持度 62.5%、信賴度 33.3%
- D**      04-12.      關於集群分析的描述，下列何者錯誤？
- (A) 同群組內的個體相似性高，不同群組的個體相似性低
  - (B) 集群分析為一種非監督式學習的過程
  - (C) 集群分析需要按照相似性來分組
  - (D) 可以利用集群分析，將學生姓名的首字母組成不同的組別

## 企業電子化人才能力鑑定

2021/1/1

- B      04-13.      視覺化能夠幫助讀者更好的理解資料，下列何種做法違背了視覺化的原則？
- (A) 進行視覺化之前，先去分析資料的內容，明確了解它的構造和每個變量的意義
  - (B) 在第一時間想好如何設計圖形，盡量將圖形設計得很複雜，看起來很美觀
  - (C) 視覺化需要能夠讓讀者在沒有資料的情況下，也能從圖中得到相應的資訊
  - (D) 在視覺化之前，先想好要表達的內容，和相應的資料，再去進行設計
- C      04-14.      在關聯規則方法中，同時購買產品 X 和產品 Y 的機率  $P(XUY)$ ，一般稱為？
- (A) 信賴度 (Confidence)
  - (B) 增益值 (Lift)
  - (C) 支持度 (Support)
  - (D) 閾值 (Threshold)
- D      04-15.      在關聯規則方法中，購買產品 X 的情況下，也會購買產品 Y 的機率  $P(Y|X)$ ，一般稱為？
- (A) 增益值 (Lift)
  - (B) 支持度 (Support)
  - (C) 閾值 (Threshold)
  - (D) 信賴度 (Confidence)
- ACD      04-16.      下列有關集群分析的敘述哪些正確？(複選)
- (A) 集群分析是一種非監督式學習 (Unsupervised learning)
  - (B) 群內差異大，群間差異小
  - (C) 依據個體間的相似性將樣本分成幾群互相沒有交集的群組
  - (D) 同群組內的個體相似性高，不同群組的個體相似性低
- C      04-17.      下列有關 K-means 集群的敘述何者有誤？
- (A) 每個類別都有自己的集群中心
  - (B) 類別中的點都是離自己類別的集群中心最近
  - (C) 無須先確定集群個數
  - (D) 資料需為標準形式，以避免因量度單位的不同而產生困擾

### 第三章 試題彙編

- C      04-18.      如附圖說明若丟擲 16 次銅板，出現了 12 個正面與 4 個反面，則期望訊息為何？（請參閱附圖作答）

決策樹分類模型中，使用信息增益(Information Gain)以及期望信息(Expected Information)之計算以決定最佳的分支屬性。資料集合 S 具有 c 個不同的類別，那麼資料集合 S 的期望信息(或稱熵值)的計算方式為：

$$\text{Entropy}(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

其中  $p_i$  為類別 i 在資料集合 S 出現的機率

- (A) 1
- (B) -1
- (C) 0.8112781
- (D) -0.8112781

- C      04-19.      以下哪一項不是處理缺失值的方法？

- (A) 直接刪除缺失值
- (B) 對缺失值進行手工填補
- (C) 將缺失值標記為異常值
- (D) 使用模型來預測缺失值

- ACD      04-20.      以下選項中屬於變數維度縮減方法的有哪些？(複選)

- (A) R 型集群分析
- (B) Q 型集群分析
- (C) 主成分分析
- (D) 相關分析

- A      04-21.      下列描述中哪一項不是決策樹的優點？

- (A) 可以輸出多個屬性
- (B) 分類非常迅速
- (C) 易於理解和解釋
- (D) 構建模型的成本低

## 企業電子化人才能力鑑定

2021/1/1

- B            04-22.    下列選項中哪一項不是資料視覺化的基本原則？  
                  ( A ) 明確視覺化的具體目標  
                  ( B ) 在圖形中盡可能呈現多樣化訊息  
                  ( C ) 考慮觀眾／聽眾的不同  
                  ( D ) 選擇合適的視覺編碼方法
- ABD        04-23.    關於 K-means 的描述中，哪些選項是正確的？(複選)  
                  ( A ) 需要事先確定集群數 K 的值  
                  ( B ) 每個類別都有自己的中心點  
                  ( C ) 只可採用歐式距離來測量接近的程度  
                  ( D ) 計算距離時默認所有變數都是同等重要
- AD           04-24.    集群與分類的不同在於，集群所要求劃分的類是未知的。 集群分析計算方法主要有：階層的方法 (hierarchical method)、分割方法 (partitioning method)、基於密度的方法 (density-based method)、基於網格的方法 (grid-based method)、基於模型的方法 (model-based method) 等。下列哪些演算法是利用統計學定義的距離進行度量？(複選)  
                  ( A ) 階層的方法 (hierarchical method)  
                  ( B ) 基於模型的方法 (model-based method)  
                  ( C ) 基於密度的方法 (density-based method)  
                  ( D ) 分割方法 (partitioning method)
- A            04-25.    將原始資料進行清理、變換、縮減…等，是在下列哪一步驟中的任務？  
                  ( A ) 資料前處理  
                  ( B ) 頻繁物項集探勘  
                  ( C ) 資料探勘流程  
                  ( D ) 分類和預測
- B            04-26.    下列何者非資料前處理該進行的動作？  
                  ( A ) 資料清理  
                  ( B ) 資料分析  
                  ( C ) 資料轉換  
                  ( D ) 資料縮減

### 第三章 試題彙編

- A      04-27.      有關資料縮減 (Data Reduction)，下列敘述何者錯誤？
- (A) 資料縮減的方法只有一種
  - (B) 可消除資料間的相互影響，增強模型的效果
  - (C) 資料縮減 (Data Reduction) 是指從原始資料中挑選出部分資料來進行分析
  - (D) 能夠高性能地進行運算
- B      04-28.      關於決策樹 (decision trees) 的基本概念，下列敘述何者有誤？
- (A) 決策樹是功能強大的預測工具
  - (B) 決策樹不是分類的工具
  - (C) 決策樹資料探勘 (Data Mining) 包含了分析者認為與決策有關的變數，以及這些變數對預測結果的能力
  - (D) 在資料探勘 (Data Mining) 的領域中，決策樹被認為是一種樹狀結構的規則
- C      04-29.      試計算出附圖表格內的吉尼指數 (Gini 指數)，下列何者正確？  
(請參閱附圖作答)

基尼指數 (Gini 指數)

|    |   |
|----|---|
| C1 | 0 |
| C2 | 6 |

(表格第二行表示節點 T 上的不同類別對應的數據個數)

- (A) 1
  - (B) 1.5
  - (C) 0
  - (D) 0.5
- A      04-30.      有關購物籃分析所得出的關聯規則，下列何者有誤？
- (A) 可以提升其他商品的買氣
  - (B) 可以提供有關產品間被同時購買的有用資訊
  - (C) 可以尋找更多潛在的目標客戶
  - (D) 可以進行相關產品推薦或挑選相應的關聯產品進行精準行銷

## 企業電子化人才能力鑑定

2021/1/1

- C      04-31.      下列何者不是集群分析的過程？  
    (A) 搜集資料  
    (B) 轉換成相似矩陣  
    (C) 簡單的分段  
    (D) 選擇一種集群方法進行分析
- C      04-32.      對於附圖中交易資料，計算 { whole milk , other vegetables }  
的支持度為何？（請參閱附圖作答）
- tropical fruit,yogurt,coffee  
whole milk  
pip fruit,yogurt,cream cheese ,meat spreads  
other vegetables,whole milk,condensed milk  
whole milk,butter,yogurt,rice,abrasive cleaner  
whole milk,cereals  
tropical fruit,other vegetables,white bread  
citrus fruit,tropical fruit,whole milk  
chicken,tropical fruit  
pork,berries,other vegetables,whole milk
- (A) 0.6  
    (B) 0.3  
    (C) 0.2  
    (D) 0.18
- D      04-33.      關於視覺化的原則，下列敘述何者錯誤？  
    (A) 明確視覺化的具體目標  
    (B) 選擇合適的視覺編碼方法  
    (C) 風格簡約、凸顯主題  
    (D) 只是用在數值資料

### 第三章 試題彙編

- AB      04-34.      資料探勘時在資料前處理階段，劃分測試集的常規方法有以下哪幾種方法？(複選)
- (A) 隨機抽樣
- (B) 交叉驗證 (Cross-validation)
- (C) 線性法
- (D) 排序法
- C      04-35.      資料清理 (Data Cleaning) 時，針對缺失值的處理不包含以下何種方式？
- (A) 使用迴歸模型填補缺失值
- (B) 直接刪除缺失值之資料
- (C) 隨機亂數填入
- (D) 使用平均值填補缺失值
- D      04-36.      下列哪一項集群 (clustering) 技術最適合處理任意形狀與大小的集群問題？
- (A) 階層式集群法
- (B) k-means
- (C) 決策樹
- (D) 以密度為基礎的集群法
- B      04-37.      請參閱附圖作答：

考慮某商場的一個時間段，記錄了如下顧客購買商品的交易集：

| <i>TID</i> | <i>Items</i>              |
|------------|---------------------------|
| 1          | Bread, Milk               |
| 2          | Bread, Diaper, Beer, Eggs |
| 3          | Milk, Diaper, Beer, Coke  |
| 4          | Bread, Milk, Diaper, Beer |
| 5          | Bread, Milk, Diaper, Coke |

依購物籃分析之方法，項集 {Milk, Bread, Diaper} 的支持度為以下何者？

- (A) 2
- (B) 2 / 5
- (C) 3
- (D) 3 / 5

## 企業電子化人才能力鑑定

2021/1/1

- D      04-38.      以下何者不是 K-means 演算法的特點？  
    (A) 每個類別都有自己的集群中心  
    (B) 類別中的點都是離自己類別的集群中心最近  
    (C) 必須先確定集群數 k  
    (D) 資料不需為標準形式
- C      04-39.      以下何者不是資料視覺化圖形？  
    (A) 散佈圖 (scatter plot)  
    (B) 儀表板 (dashboard)  
    (C) 甘特圖 (Gantt chart)  
    (D) 社會網路圖 (Social Network Graph, SNG)
- A      04-40.      異常值通常是指一個偶發的數值，與其他的測量值相差甚遠，我們可以使用何種統計量判別異常值？  
    (A) 標準差  
    (B) 相關係數  
    (C) 判定係數  
    (D) 變異係數
- D      04-41.      關於資料視覺化的描述，下列哪一項有誤？  
    (A) 資料視覺化主要將資料、文字、圖表，轉成各種圖表與影像，使得資料較容易瞭解  
    (B) 資料視覺化需考慮觀眾不同  
    (C) 熱繪圖、文字雲、網路圖等都屬於資料視覺化圖形  
    (D) 資料視覺化不可以追求絢麗
- AB      04-42.      資料探勘中將預測結果與真實值進行比較，下列敘述哪些正確？  
    **(複選)**  
    (A) 對感興趣的類別作了正確的分類稱為真陽性 True Positive (TP)  
    (B) 真陽性比率稱為特異性 (specificity)  
    (C) 對不感興趣的類別作了正確的分類稱為真陰性 True Negative (TN)  
    (D) 真陰性比率稱為靈敏度 (sensitivity)



### 第三章 試題彙編

- ABC      04-43.      下列有關資料視覺化圖形的敘述哪些正確？(複選)
- (A) 將企業營運所需之各項資訊，濃縮在儀表板 (Dashboard) 之中，以幫助經營決策者
  - (B) 用來說明多個變數 (例：四個變數) 間兩兩變數的相關性稱為散佈圖矩陣 (Scatterplot Matrix)
  - (C) 將二維數值矩陣以顏色表示稱為熱繪圖 (Heat Map)
  - (D) 由一個或多個特定類型的相互依存關係所構成稱為社會文字雲 (Word cloud)
- AB      04-44.      有關長寬表的敘述下列哪些正確？(複選)
- (A) 長表變數很少，紀錄很多
  - (B) Database 的結構通常是長表
  - (C) 長表指的是變數比較多紀錄比較少的資料
  - (D) 任何包含離散變數的資料都不可以在長表和寬表之間自由轉換
- D      04-45.      關於階層集群，下列何者不是常用的群間距離度量？
- (A) 組平均距離 (group average) 法
  - (B) 最短距離 (single linkage) 法
  - (C) 最長距離 (complete linkage) 法
  - (D) 最佳距離 (optimal linkage) 法
- C      04-46.      假定希望分析愛喝咖啡和愛喝茶的人之間的關係，收集一組人飲料偏好的資料如附圖，請計算關聯規則 {茶} → {咖啡} 的增益值 (lift) 為何？(請參閱附圖作答)

| 項目 | 咖啡  | 非咖啡 |
|----|-----|-----|
| 茶  | 150 | 50  |
| 非茶 | 650 | 150 |

- (A) 0.7525
- (B) 0.8255
- (C) 0.9375
- (D) 0.8535

## 企業電子化人才能力鑑定

2021/1/1

- D      04-47.      何者不是常用的群間距離衡量？  
                  ( A ) 最短距離法  
                  ( B ) 最長距離法  
                  ( C ) 中位數法  
                  ( D ) 圓心法
- D      04-48.      關於 K 平均法 (K-means)，下列敘述何者不正確？  
                  ( A ) 必須先確定群數 k  
                  ( B ) 各群都有自己的群中心  
                  ( C ) 基本算法非常簡單  
                  ( D ) 群中的樣本點都離自己的群中心最遠
- A      04-49.      關於 NoSQL 的敘述何者錯誤？  
                  ( A ) NoSQL 的意思是「不同於一般 SQL 指令的資料庫」〈No SQL〉  
                  ( B ) NoSQL 適用於不需要使用關聯模型的資料  
                  ( C ) NoSQL 資料庫主要並非以關聯式資料表建構而成，所以通常不以 SQL 指令做資料操作  
                  ( D ) NoSQL 資料庫一般不支援 ACID 特性
- D      04-50.      變數的維度縮減技術中，將多個變數通過線性組合的方法，以得到數量較少且互相獨立的新變數集合稱為什麼方法？  
                  ( A ) 決策樹 〈Decision tree〉  
                  ( B ) 集群分析 〈Clustering〉  
                  ( C ) 關聯規則 〈Association rule〉  
                  ( D ) 主成份分析 〈Principal Component Analysis, PCA〉