

# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 1 頁，共 11 頁

單選題 50 題 (佔 100%)

C	1. 下列何者屬於量化屬性？ (A) 商品支付的方式 (B) 學生參加的社團 (C) 一批產品的不良數 (D) 員工隸屬的公司部門
B	2. 參考附圖，關於 Python 語言對於 requests 套件讀取 HTML 網頁資料，下列敘述何者「不正確」？  <pre>In [1]: import requests</pre> <pre>In [2]: url = 'https://data.gov.tw/'</pre> <pre>In [3]: r = requests.get(url)</pre> <pre>In [4]: r</pre> <pre>Out[4]: &lt;Response [200]&gt;</pre> (A) r.encoding 可以設定 HTML 資料的編碼 (B) r.value 的結果會顯示資料讀取內容 (C) r.status_code 的結果會顯示資料讀取的狀態碼 (D) <Response [200]>的結果表示 HTML 資料讀取正常
C	3. 在 Python 資料處理中，使用 sorted()，執行附圖程式後，下列何者為輸出結果？ <pre>&gt;&gt;&gt; my_list = [1, 3, 5, 7, 9, 2]</pre> <pre>&gt;&gt;&gt; sorted(my_list, reverse=True)</pre> <pre>&gt;&gt;&gt; print(my_list)</pre> (A) [9, 7, 5, 3, 2, 1] (B) [1, 2, 3, 5, 7, 9] (C) [1, 3, 5, 7, 9, 2] (D) [4, 3, 2, 1, 5, 0]
C	4. 參考附圖，關於 Python 語言對於 numpy 套件，下列敘述何者為正確？

# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 2 頁，共 11 頁

	<pre>import numpy as np  a = np.array([[1, 2, 3], [4, 5, 6]])  b = a.reshape((3, 2))  b[1, 0] = 999  c = b.reshape(-1, 3)</pre> <p>(A) a.shape 的結果為(3, 3) (B) a.ndim 的結果為 1 (C) a[0, 2]的結果為 999 (D) c[1, 0]的結果為 2</p>
C	<p>5. 關於文字資料處理，下列敘述何者「不正確」？</p> <p>(A) 字詞（word or term）是語言中最小的單位 (B) 詞素（morpheme）則是最小可區辨的單元 (C) 詞幹提取（stemming）是將詞綴移除以獲得可在字典中查閱得到的字詞基本形式 (D) 詞組提取（chunk extraction）是依詞性標記（part-of-speech tagging）提取出詞組</p>
B	<p>6. Python 的資料型態中，有可變（mutable）與不可變（immutable），以容器（container）為例，下列何者是『不可變物件』？</p> <p>(A) list（串列） (B) tuple（值組） (C) dict（字典） (D) set（集合）</p>
A	<p>7. 關於 Python 之資料結構字典（dictionary）與例外處理（exception），執行附圖程式後，下列何者為輸出結果？</p> <pre>d = {"M":1, "A":1, "R":2, "X":1} for c in "HARPO":     print(d[c] + '-')</pre> <p>(A) KeyError: 'H' (B) TypeError: list indices must be integers, not str (C) SyntaxError: invalid syntax (D) TypeError: unsupported operand type(s) for +: 'int' and 'str'</p>
A	<p>8. R 語言中，下列何者為 class(list(1)[1])之回傳值？（即 list(1)[1]之物件類型為何？）</p> <p>(A) list</p>

# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 3 頁，共 11 頁

	(B) numeric (C) 1 (D) vector
C	9. 關於 MySQL，下列敘述何者「不正確」？ (A) 可以在許多作業平台上執行，包括 Windows、Unix 及 Linux (B) 預設的最高管理員帳號為 root (C) 使用 FORTRAN 編寫 (D) MariaDB 是由 MySQL 原始開發者領導開發
D	10. 下列何者「不是」NoSQL 資料庫的特色？ (A) 可以處理大量的資料 (B) 有較高的可用性 (availability) (C) 有較高的縮放性 (scalability) (D) 適合存放結構性較好的資料
D	11. 下列何者資料庫的建立及運作方式與其它有明顯的區別？ (A) MariaDB (B) MySQL (C) MS-SQL (D) MongoDB
D	12. 關於資料庫系統，若一個關聯表 (relation) 有五個屬性 (attributes)，請問有多少種主鍵 (primary key) 的可能組合？(如：A、B、C、D 與 E 五個屬性，BDE 是一種可能的主鍵) (A) 25 (B) 5 (C) 120 (D) 31
C	13. 資料庫管理系統 (Database Management System, DBMS) 在寫入或更新資料的過程中，允許多個交易 (transaction) 同時對其數據資料進行讀寫和修改，但可以防止多個交易同時執行時數據的不一致，請問符合下列何者特性？ (A) Atomicity (原子性) (B) Consistency (一致性) (C) Isolation (隔離性) (D) Durability (持久性)
A	14. 對於關聯式資料庫，使用者可以使用資料操作語言 (Data Manipulation Language, DML)，對於資料表記錄進行一些操作，下列何者「不是」資料操作語言？

# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 4 頁，共 11 頁

	(A) CREATE (建立) (B) INSERT (新增) (C) UPDATE (修改) (D) DELETE (刪除)
A	15. 關於 SQL，下列敘述何者正確？ (A) 資料定義語言 (Data Definition Language, DDL)，用來「定義」資料庫的結構、欄位型態及長度 (B) 資料操作語言 (Data Manipulation Language, DML)，用來「操作」使用者對「資料庫內容」的存取權利 (C) 資料控制語言 (Data Control Language, DCL)，用來「控制」資料庫的新增資料、修改資料、刪除資料、查詢資料等功能 (D) CREATE、ALTER、DROP 屬於資料定義語言
B	16. Python 語言中可用 open() 函數開啟檔案，若僅要讀取檔案，mode 引數需使用下列何項設定值？ (A) a (B) r (C) w (D) wb
C	17. 在 R 語言 {base} 與 Python pandas 中，檢視資料物件 data 前幾筆觀測值的函數均記為 head()，但其語法有所差異，請問下列何者正確？ (A) R：data.head(n=10) (B) Python：head.data(n=10) (C) R：head(data,n=10L) (D) Python：head(data,n=10L)
D	18. 下列何者資料之結構性較高？ (A) 影像 (B) 論文 (C) 電子郵件 (D) 人口統計資料
B	19. 關於 JSON 格式，下列敘述何者「不正確」？ (A) 易於人們閱讀與編寫 (B) 可用大括弧{ }建立陣列，以寫入資料 (C) 是源自於 JavaScript 的一種資料交換格式 (D) 選與其他程式語言相容性高
C	20. 關於讀取.csv 逗號分隔檔案，下列敘述何者「不正確」？ (A) 在 R 語言中，可使用 read.csv() 函數

# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 5 頁，共 11 頁

	<p>(B) 在 Python 語言中，可使用 pandas 套件的 read_csv() 方法</p> <p>(C) 在 R 語言中，可使用 read() 函數</p> <p>(D) 在 Python 語言中，可使用 open() 函數</p>
C	<p>21. 關於 R 語言與 Python 語言，下列敘述何者正確？</p> <p>(A) R 語言中，無法使用 write.table() 函數匯出 .csv 逗號分隔檔案</p> <p>(B) Python 語言中，可使用 numpy 的 to_csv() 方法匯出 .csv 逗號分隔檔案</p> <p>(C) R 語言中，可使用 write.csv() 函數匯出 .csv 逗號分隔檔案</p> <p>(D) Python 語言中，可使用 open('file.csv', 'r') 讀寫 .csv 逗號分隔檔案</p>
A	<p>22. 關於 Python 語言讀取檔案，下列敘述何者「不正確」？</p> <p>(A) 使用 open('file', 'r')，若 file 不存在會創建 file</p> <p>(B) 使用 open('file', 'w')，若 file 不存在會創建 file</p> <p>(C) 使用 open('file', 'w+') 為可讀寫的檔案模式</p> <p>(D) 使用 open('file', 'a+') 為附加讀寫檔案模式，寫入時添加於後，亦可以讀檔</p>
C	<p>23. 下列何者「不是」中文文字常用的編碼系統？</p> <p>(A) CP950</p> <p>(B) BIG5</p> <p>(C) Shift-JIS</p> <p>(D) UTF-8</p>
D	<p>24. 執行下方程式後，下列何者為輸出結果？</p> <pre>b = 30 def test_func():     global b     b = 50     print('b: ', b)  test_func() print('b: ', b)</pre> <p>(A) b: 30 b: 50</p> <p>(B) b: 50 b: 30</p> <p>(C) b: 30 b: 30</p> <p>(D) b: 50</p>

# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 6 頁，共 11 頁

	b: 50
A	25. 關於 Python 語言，下列敘述何者正確？ (A) Python 有豐富的函式庫，應用範圍廣泛 (B) Python 毫無記憶體方面的（Leaks）洩漏問題 (C) Python 程式碼執行速度比 C++快速 (D) Python 是第三代程式語言，C++是第四代程式語言
B	26. 關於統計計算語言 R 與 SPSS 的特性，下列敘述何者「不正確」？ (A) R 是一個免費的自由軟體，它有 Linux、MacOS 和 Windows 的版本 (B) SPSS 是世界上最早的統計分析軟體 (C) 兩者皆有完整的數據處理、計算和繪圖功能 (D) R 不僅提供統計計算的函數，還提供各種數學計算與最佳化函數，進行數據分析
A	27. 在 Python 3 中，下列何者「不能」作為變數名稱？ (A) lambda (B) 變數名稱 (C) false (D) foobar
C	28. R 語言中，儲存字串會使用何種基本資料型態？ (A) string (B) numeric (C) character (D) logical
B	29. Python 語言中，下列何者為執行 5//2 的輸出結果？ (A) 1 (B) 2 (C) 3 (D) 2.5
C	30. 在 Python 3 中，根據下列程式碼，下列敘述何者正確？ mylist=[1,2,3] mylist+= [4,5,6]  mytuple=(1,2,3) mytuple+= (4,5,6) (A) mylist 的值變成[1,2,3,[4,5,6]] (B) mylist+= [4,5,6]此行創建新物件存放[1,2,3,4,5,6]到 mylist 中

# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 7 頁，共 11 頁

	<p>(C) mytuple 的值變成(1, 2, 3, 4, 5, 6)</p> <p>(D) mytuple+=(4,5,6)此行出現錯誤訊息</p>
B	<p>31. R 語言中，下列何種運算子可建立全域變數？</p> <p>(A) &lt;&lt;</p> <p>(B) &lt;&lt;-</p> <p>(C) ==</p> <p>(D) &lt;-</p>
D	<p>32. 下列何者「不是」R 語言中常見的迴圈控制敘述？</p> <p>(A) while</p> <p>(B) for</p> <p>(C) repeat</p> <p>(D) do</p>
B	<p>33. 關於 R 語言中的控制敘述，下列何者「不正確」？</p> <p>(A) while 判斷後面小括號內的條件，若成立就做迴圈</p> <p>(B) for 後面小括號內需先提供判斷條件</p> <p>(C) if 後面小括號內為邏輯值或判斷式</p> <p>(D) repeat 若沒有 break 則會繼續執行下去</p>
B	<p>34. 屬性轉換為資料前處理的重要步驟，下列敘述何者「不正確」？</p> <p>(A) 可以降低資料尺度</p> <p>(B) 會破壞模型的預測能力</p> <p>(C) 可以降低資料偏斜性</p> <p>(D) 可以降低離群資料對於模型的不良影響</p>
A	<p>35. Python 語言中，執行下方程式後，輸出結果為下列何者？</p> <pre>def mysum(a,b,c=6):     return a+b+c def mysum(a,b):     return a+b print(mysum(1,2,3))</pre> <p>(A) 程式出現錯誤訊息</p> <p>(B) 3</p> <p>(C) 6</p> <p>(D) 9</p>
D	<p>36. 下列 R 語言中的控制敘述語法，何者編譯時不會產生問題且傳出的 y 為 5？（假設 x=1）</p> <p>(A) while(x &gt; 0){y = 5}</p> <p>(B) for(x &lt; 10)[y = 5]</p>



# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 8 頁，共 11 頁

	<p>(C) repeat(x &gt; 10, y = 5)</p> <p>(D) if(x &lt; 10) y = 5</p>
C	<p>37. 欲計算一資料表中各縱行的算術平均數，可以使用附圖之方法。在其他條件相同的情況下，附圖三種方法所耗費的時間，下列何者為由慢到快的排序？</p> <p>1. 撰寫巢狀的外顯迴圈 (explicit looping)，逐行取出各個元素加以計算算術平均數</p> <p>2. 利用 R 語言的向量化計算函數 colMeans()</p> <p>3. 使用 R 語言或 Python 語言的 apply() 隱式迴圈函數</p> <p>(A) 1 &gt; 2 &gt; 3</p> <p>(B) 3 &gt; 2 &gt; 1</p> <p>(C) 1 &gt; 3 &gt; 2</p> <p>(D) 3 &gt; 1 &gt; 2</p>
B	<p>38. R 語言中，執行下方程式後，下列何者為 y 值的輸出結果？</p> <pre> x=10 y=0 while(x &gt; 0) {   x=x-2   if (x &lt; 5) break   y=x*2 } </pre> <p>(A) 8</p> <p>(B) 12</p> <p>(C) 16</p> <p>(D) 20</p>
A	<p>39. 關於進行資料導向程式設計，下列敘述何者「不正確」？</p> <p>(A) 需多撰寫迴圈，避免使用向量化函數</p> <p>(B) 多採用邏輯值索引，避免運用條件式語法</p> <p>(C) 善用自定義函數模組化工作流程</p> <p>(D) 注意引用函數其預設的引數值</p>
A	<p>40. R 語言中，若無其它敘述，下列何者「不能」使用 break 指令？</p> <p>(A) if ... else ...</p> <p>(B) for ...</p> <p>(C) while ...</p> <p>(D) repeat ...</p>
B	<p>41. 當 x 為 0 時，執行下列 R 語言程式後，x 值為何？</p>



# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 9 頁，共 11 頁

	<pre>for (i in 1:3) {   for (j in 2:4) {     if (i &gt; j) x = x + 1 else x = x - 1   } }</pre> <p>(A) -9 (B) -7 (C) 7 (D) 9</p>
A	<p>42. R 語言中，關於程式碼執行時產生的訊息，下列敘述何者「不正確」？</p> <p>(A) 產生錯誤（errors）訊息時，將繼續執行無錯誤的程式 (B) 警告（warnings）訊息會說明潛在的問題 (C) 一般傳回的訊息在於說明代碼輸出的結果 (D) 產生警告（warnings）訊息時，可繼續執行程式</p>
B	<p>43. 例外狀況處理函數允許程式設計師在例外狀況發生時，採取某些行動，例如結合 stop( ) 函數，以讓程式設計師在必要的時候傳回訊息，瞭解程式碼執行的狀況。下列何者「不是」R 語言的例外狀況處理函數？</p> <p>(A) withCallingHandlers( ) (B) suppressMessages( ) (C) tryCatch( ) (D) try( )</p>
C	<p>44. 關於 traceback( ) 函數，下列敘述何者「不正確」？</p> <p>(A) 可以協助獲得錯誤訊息的額外資訊 (B) 會顯示錯誤發生於何處 (C) 會顯示為何發生錯誤 (D) 未顯示錯誤訊息的程式碼，不一定代表結果無誤</p>
A	<p>45. 關於程式除錯與效率監測 R 函數，下列敘述何者「不正確」？</p> <p>(A) debug( ) 可關閉函數的偵錯模式 (B) suppressMessages( ) 可用來不顯示提示訊息 (C) traceback( ) 可以協助獲得錯誤訊息的額外資訊 (D) system.time( ) 可以衡量程式碼的執行時間</p>
C	<p>46. 關於資料導向程式設計，下列敘述何者「不正確」？</p> <p>(A) 精進記憶體管理以提升程式效能 (B) 經常注意經函數處理後輸出的物件其類別型態 (C) 多採用條件式語法，避免運用邏輯值索引</p>

# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 10 頁，共 11 頁

	(D) 加強撰寫平行化程式的技巧可有助提升效能
B	<p>47. 關於程式碼錯誤處理，下列敘述何者「不正確」？</p> <p>(A) 撰寫程式時讓錯誤可以重製，可便於偵錯</p> <p>(B) 可執行的程式即為無誤的程式</p> <p>(C) 可先要求程式碼的邏輯正確，再設法提升執行效率</p> <p>(D) 可藉由函數輔助處理例外狀況與瞭解錯誤訊息</p>
B	<p>48. 一般而言，程式設計師透過下列程序偵錯與解決代碼的錯誤，下列何者為流程順序？1.找出錯誤的原因、2.讓錯誤可以重製、3.修正錯誤並測試、4.意識到錯誤的存在</p> <p>(A) 1&gt;2&gt;3&gt;4</p> <p>(B) 4&gt;2&gt;1&gt;3</p> <p>(C) 4&gt;3&gt;2&gt;1</p> <p>(D) 2&gt;4&gt;1&gt;3</p>
D	<p>49. 參考附圖 R 語言對於正規表示式 (regular expression)，下列敘述何者正確？</p> <pre>mydata &lt;- c("BigData", "Geodata", "BigDataAnalysis")</pre> <p>(A) <code>grep(pattern = "data", x = mydata)</code>，執行結果為 1</p> <p>(B) <code>grep(pattern = "data", x = mydata, value = TRUE)</code>，執行結果為 "BigData"</p> <p>(C) <code>grepl(pattern = "Data", x = mydata)</code>，執行結果為 TRUE TRUE TRUE</p> <p>(D) <code>grepl(pattern = "Data\$", x = mydata)</code>，執行結果為 TRUE FALSE FALSE</p>
C	<p>50. 執行附圖程式碼後，下列何者為輸出結果？</p> <pre>x = 5 try:   y = z   print('y 值為: {}'.format(y))   x = 6 except:   print('except: 錯誤產生') finally:   print('finally: 結束') print('x 值為: {}'.format(x))</pre> <p>(A) y 值為: error</p>

# 109 年度初級巨量資料分析師能力鑑定試題

科目 1：資料導向程式設計

考試日期：109 年 5 月 30 日

第 11 頁，共 11 頁

	<p>finally: 結束 x 值為: 5</p> <p>(B) finally: 結束 x 值為: 6</p> <p>(C) except: 錯誤產生 finally: 結束 x 值為: 5</p> <p>(D) except: 錯誤產生 finally: 結束 x 值為: 6</p>
--	---