

中央存款保險公司 108 年進用正式職員甄試試題

甄試類別【代碼】：數據分析人員-風險管理類(七職等)【P2404】

專業科目 2：統計實務與數據分析：包括統計實務、數據資料架構處理及分析，以實例命題，含程式語言(如 R、Python)撰寫，不包括統計法規

*入場通知書編號：_____

注意：①作答前先檢查答案卷，測驗入場通知書編號、座位標籤、應試科目等是否相符，如有不同應立即請監試人員處理。使用非本人答案卷作答者，不予計分。
 ②本試卷為一張單面，非選擇題共 4 大題，每題各 25 分，共 100 分。
 ③非選擇題限以藍、黑色鋼筆或原子筆於答案卷上採橫式作答，並請依標題指示之題號於各題指定作答區內作答。
 ④請勿於答案卷上書寫姓名、入場通知書編號或與答案無關之任何文字或符號。
 ⑤本項測驗僅得使用簡易型電子計算器(不具任何財務函數、工程函數、儲存程式、文數字編輯、內建程式、外接插卡、攝(錄)影音、資料傳輸、通訊或類似功能)，且不得發出聲響。應考人如有下列情事扣該節成績 10 分，如再犯者該節不予計分。1.電子計算器發出聲響，經制止仍執意續犯者。2.將不符規定之電子計算器置於桌面或使用，經制止仍執意續犯者。
 ⑥答案卷務必繳回，違反者該節成績以零分計算。

第一題：

為了解什麼樣的網頁設計可以讓瀏覽者在網頁停留久一點，網頁工程師設計了三個網頁，瀏覽者點選特定的 icon 後將隨機進入三個網頁其中之一，瀏覽器會記錄該瀏覽者在該網頁的停留時間，以 Y_{ij} 代表第 j 個瀏覽者在第 i 個網頁停留的時間，其中 $i = 1, 2, 3, j = 1, 2, \dots, n$ 。前 120 個瀏覽者的瀏覽時間資料的敘述性統計、處理平方和及總平方和，如下：

webpage	個數	平均數	標準差	處理平方和	43128.965
1	40	21.5071	10.94382	總平方和	60994.371
2	40	44.4813	12.23800		
3	40	67.9438	13.73140		

- (一) 假設 μ_i 代表瀏覽者在第 i 個網頁停留的平均時間。欲檢定三個網頁的平均停留時間是否有差異，請先寫出資料分析的統計模式，並寫出適當的檢定假說。【10 分】
- (二) 請完成變異數分析表，並根據表中的數據，在 5% 的顯著水準下，檢定的結論為何？【15 分】

附註：假設 $F \sim F_{m,n}$ ，令 $F_{m,n,\alpha}$ 表示在 α 顯著水準下的臨界點，即 $P(F \geq F_{m,n,\alpha}) = \alpha$
 $F_{1,100,0.025} = 5.1786, F_{2,100,0.025} = 3.8284, F_{3,100,0.025} = 3.2496,$
 $F_{1,120,0.025} = 5.1523, F_{2,120,0.025} = 3.8046, F_{3,120,0.025} = 3.2269,$
 $F_{1,100,0.05} = 3.9361, F_{2,100,0.05} = 3.0873, F_{3,100,0.05} = 2.6955,$
 $F_{1,120,0.05} = 3.9201, F_{2,120,0.05} = 3.0718, F_{3,120,0.05} = 2.6802.$

第二題：

某軟體開發公司為了徵才，舉辦程式設計競賽。60 位參賽選手平均分成兩組，分別以商業套件及免費軟體，在相同問題之下，撰寫程式並執行解決方案。兩組選手的程式執行時間的敘述性統計如下：

敘述統計

	個數	最小值	最大值	平均數	標準差
商業套件組	30	8.42	51.64	26.0521	9.69749
免費軟體組	30	15.49	50.67	33.3491	9.47678

假設程式執行時間服從常態分配。

- (一) 請分別算出兩組執行時間的 95% 信賴區間。【15 分】
- (二) 請算出兩組執行時間差的 95% 信賴區間。【10 分】

附註：假設 $Z \sim$ 標準常態，令 Z_α 表示在 α 顯著水準下的臨界點，即 $P(Z \geq z_\alpha) = \alpha$
 $z_{0.10} = 1.285, z_{0.05} = 1.645, z_{0.025} = 1.96.$

第三題：

集群分析(clustering analysis)是一種可用來將資料分群的實用統計分析方法，請回答下列問題：

- (一) 何謂集群分析？請詳述集群分析中常見 K-mean 的分析方法。【10 分】
 ※答案須包含 K-mean 方法的運作方式，否則不予計分。
- (二) 請以 Python 或 R 語言，撰寫出具備集群分析功能與圖形之程式碼。【15 分】
 ※程式碼設計及資料分析情境需與任何風險管理議題相關，否則不予計分。

第四題：

在資料分析中，相關分析(correlation analysis)是一種常使用到的統計分析方式，同時它亦是許多統計分析方法的重要基礎，請問：

- (一) 何謂相關分析？請舉例說明變數之間所存在的相關性不等於變數之間的因果性。【10 分】
- (二) 下表是名為 test.csv 簡易資料表，請以 Python 或 R 語言撰寫出具備相關分析之程式碼。【15 分】
 ※資料欄位或變數可自行擇定，答案程式碼必須能夠執行出散佈圖。

Variable 1	Variable 2	Variable 3	Variable 4
1	2	3	5
1	3	4	5
3	4	5	1
5	4	2	1
5	3	1	4