Fall 2017 第四次作業 統計諮詢

12/11/2016繳交

說明：本次作業的目的在於探索性分析(EDA)，原則上以敘述性統計量與統計圖表為主，請盡量不使用較複雜的模型或方法。

1. 菸酒公賣局的紹興酒問卷，在樣本代表性的分析中有瑕疵，請以「地區」、「年齡」、「性別」三者為例，示範樣本代表性的分析操作；根據你/妳們的分析結果，和原報告中的結果比較（例如：表二及表三），並提供各組認為比較合適的詮釋。
2. 「基本資料分析」中有4行資料：
3. 請以敘述性統計量及圖表，說明這些資料有哪些基本特性（例如：對稱）；另外，以圖形比較資料平均數及變異數的大小。
4. 如果這些資料是常態、t分配、指數分配三者的亂數，判斷這四筆資料分屬於那一種分配。
5. 假設某行資料在記錄時意外地將900筆來自*N()*分配、100筆來自*N()*分配混在一起，其中。請說明你/妳會如何判斷哪一行資料比較可能符合這個條件，另外，也請說明如何判斷的大小。
6. (a) Boxplot可用於判斷離群值，試以電腦模擬驗證Boxplot顯示的離群值，在資料來自於標準常態分配時，離群值代表的發生機率大約為何？（註：建議重複模擬1000次，每次抽取500、5000、或更多亂數。）

(b)另外，如果資料本身不是常態分配，以Boxplot判斷離群值是否會產生問題？（註：如果資料來自t分配，自由度為1、5、10、15、20有多少差異？）

1. 下表為臺灣樂透彩前20期中獎號碼，請以圖像化方式說明它們是否為亂數。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Date | Numbers | **Power****Ball** | Date | Numbers | **Power Ball** |
| 0329 | 22 | 31 | 34 | 25 | 21 | 19 | **13** | 0222 | 32 | 10 | 15 | 02 | 30 | 23 | **36** |
| 0326 | 05 | 18 | 25 | 26 | 35 | 42 | **29** | 0219 | 24 | 20 | 36 | 19 | 07 | 12 | **26** |
| 0321 | 32 | 21 | 09 | 27 | 31 | 06 | **2** | 0215 | 01 | 06 | 07 | 12 | 42 | 20 | **35** |
| 0319 | 05 | 25 | 02 | 16 | 32 | 09 | **7** | 0212 | 25 | 39 | 20 | 38 | 29 | 37 | **28** |
| 0315 | 15 | 29 | 05 | 36 | 13 | 10 | **1** | 0208 | 26 | 02 | 15 | 29 | 04 | 33 | **39** |
| 0312 | 36 | 16 | 12 | 26 | 08 | 34 | **5** | 0205 | 17 | 39 | 03 | 15 | 11 | 01 | **34** |
| 0308 | 04 | 40 | 27 | 21 | 14 | 05 | **12** | 0201 | 13 | 39 | 28 | 30 | 25 | 29 | **21** |
| 0305 | 29 | 04 | 10 | 23 | 39 | 14 | **36** | 0129 | 07 | 09 | 29 | 34 | 39 | 36 | **16** |
| 0301 | 30 | 12 | 40 | 32 | 35 | 20 | **34** | 0125 | 28 | 31 | 16 | 35 | 06 | 30 | **2** |
| 0226 | 40 | 06 | 20 | 29 | 38 | 35 | **41** | 0122 | 10 | 32 | 13 | 04 | 09 | 33 | **37** |

1. 請以美國越戰抽籤入伍資料為例（「Draft Lottery Datafile.txt」），示範如何以圖形探討各月份中籤順序的差異性；另外，提議統計分析及檢定方法，檢證一至十二月份的中籤機率是否相同。