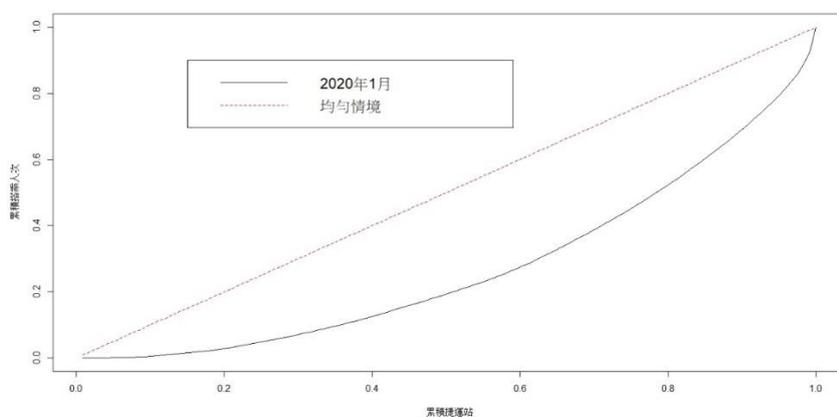


統計實務第四次作業

繳交日期 May 17, 2024

Spring 2024

- 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。
- 「基本資料分析」中有 4 行資料：
 - 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、範圍等。
 - 如果其中一筆資料是根據常態分配產生的亂數，判斷哪一筆資料的可能性最高，並說明採用的方法。
 - 假設某行資料在記錄時意外地將 900 筆來自 $N(\mu_1, \sigma^2)$ 分配、100 筆來自 $N(\mu_2, \sigma^2)$ 分配混在一起，其中 $\mu_1 \neq \mu_2$ ，請說明你/妳如何判斷 μ_1 或 μ_2 何者較大。（加分題：如何判斷 $\mu_1 - \mu_2$ 的數值。）
- 政府資料開放平台(data.gov.tw)有不少資料可做為大數據分析練習，請以「台北捷運人次」資料為依據：
 - 繪圖說明台北捷運開始營運以來，搭乘人數的時間趨勢，同時計算成長率以哪幾年最為不同，如有可能，嘗試找出造成這些結果的原因。
 - 以民國 98 年至今的資料，找出進出人數最多的前五名、最少的前五名。
 - Lorenz Curve 及 Gini Index 試測量不均度的常用工具之一，計算每年所有捷運站進出人次的 Gini Index，同時提供分析結果的說明。



4. 近年愈來愈多研究以房地產資料為題，套用迴歸分析估算房價。請各組同學至我的網站下載「波蘭華沙房價」的房地產資料（1000 筆模擬資料），其中包括一個目標變數、五個解釋變數。

(a) 請以 EDA 工具整理所有變數的基本特性。

(b) 以迴歸分析找出目標變數與解釋變數間的關係。

- *m2.price*, apartments price per meter-squared (in EUR), a numerical variable range 1607 – 6595;
- *construction.year*, the year of construction of the block of flats in which the apartment is located, a numerical variable range 1920 – 2010;
- *surface*, apartment's total surface in square meters, a numerical variable range 20 – 150;
- *floor*, the floor at which the apartment is located (ground floor taken to be the first floor), a numerical integer variable with values from 1 to 10;
- *no.rooms*, the total number of rooms, a numerical variable with values from 1 to 6;
- *district*, a factor with 10 levels indicating the district of Warsaw where the apartment is located.

5. 請以美國越戰抽籤入伍資料為例（「抽籤入伍資料」），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。

- 加分題：根據主計總處公布的「110 年家庭財富分配統計」，計算我國與南韓、澳洲等六個國家的吉尼係數。