**R語言上手隨堂練習**

名字:何OO

單位:元智大學工業工程與管理學系

以下報告內容: 本會第五十六屆年會暨 2020國際品質管理研討會 (ISQM 2020) 歡迎國內外與品質管理、品質工程、品質改善、六標準差、全面品質管理、ISO 9000、IATF16949、AS9100、可靠度分析與工程、智慧品質、智能製造之品管模式、智慧工廠之品質系統、智慧醫療、醫療品質、防疫品質、紡織業品質、數位化、數位轉型有關之論文(中英文)。

**1. 模擬分析與結果 (題目14字)**

若有子標題 (12 號粗體字)，子標題的前面應要有阿拉伯數字之序號，並向左對齊 (凸排 2.5 字元)，例如：

**1.1 模擬分析**

**1.1.1 模擬設定**

8. 數學式：所有稿件中引用到的公式及方程式都必須以阿拉伯數字編號，編號置於括號內 (如： (1), (2),….) 且是向右靠齊。方程式內容以斜體繕打為原則，例如：

 (1)

#基本使用

**#查看資料集設定**

?cars

?iris

library(ggplot2)

**#ggplot的邏輯是用疊的**

ggplot(cars, aes(x=speed, y=dist)) #引入資料與畫表格底

ggplot(cars, aes(x=speed, y=dist)) +geom\_point() #表現形式為點點圖

**#盒鬚圖 後方參數皆為輸入資料集中的資料**

ggplot(data=CO2) + geom\_boxplot(data=CO2, aes(x=conc, y=uptake, colour=Plant))

ggplot(CO2, aes(x=conc, y=uptake, colour=Plant)) +geom\_boxplot()

# 以上兩種作法皆可 ggplot(資料來源,aes(x = 一種數值資料,y = 一種數值資料,colour=類別資料))

**#用boxplot 畫盒鬚圖 差異是 這是以x軸分類 所以少一個參數**

boxplot(cars$speed ~ cars$dist)

**#相關分析**

#𝑐𝑜𝑣(𝑋, 𝑌)=𝐸[(𝑋−𝐸(𝑋))(𝑌−𝐸(𝑌))] 相關係數

#𝑉𝑎𝑟(𝑋)=[(〖𝑋−𝐸(𝑋))〗^2] 變異數

cov (1:5, 2:6)

cor(iris$Sepal.Width, iris$Sepal.Length)

cor(iris[, 1:4])

**#相關性視覺化**

install.packages("corrgram")

library(corrgram)

corrgram(iris, upper.panel=panel.conf) #參數為右上角是甚麼

**#分析範例**

x<-1:10

y<-x^3

cor(x, y)

**#檢定**

#每一個命題都有對應的假設→然後運用適合的統計方法進行檢定

#相關係數檢定(用假設檢定 假設相關性成立 超過百分之95%就拒絕)

cor.test(c(1,2,3,4,5), c(1,0,3,4,5), method = "pearson")

#推論檢定

#透過t.test計算信賴水準為95%的信賴區間 和其平均值 所以可以知道抽取合理性(平均值合理)

x <- rnorm(30) #從平均數為0、變異數為1的常態分布中抽取30個樣本

t.test(x)#平均在區間內 可信賴 基於母體與抽取變異數相等 可推論信賴區間 只要平均數在裡面就代表抽取成功(平均值合理)

x <- rnorm(30, mean=10)#從平均數為10、變異數為1的常態分布中抽取30個樣本

t.test(x, mu=10)

# f 檢定 了解兩資料變異數有沒有一樣(拒絕表示 變異數有顯著差異)

tapply(sleep$extra,sleep$group,mean) #看兩筆數據的某特質

var.test (extra~group, sleep) #extra~group 表示一種group的extra當一種數據

# 程序 先F檢定 發現變異數近似 再做t檢定 就可以理解兩組數據之關聯

#ex: (要因為變異數一樣,才可以使用t檢定)

var.test (extra~group, sleep)

#得到變異數差異不顯著 0.79 要小於 0.05 ,所以進行t test ,var設為相等

t.test(extra~group, data=sleep, paired=FALSE, var.equal=TRUE) #paired 兩組數據獨立?

#p-value 0.079 不夠大 要小於 0.05才可以拒絕 得到兩組數據近似(平均數差異不明顯) 即使平均值差幾倍

#配對樣本檢定 看兩組數據是否相依(成對) 要用with在前面包著 表示成對

#t test檢定

with(sleep,t.test(extra[group == 1],extra[group == 2],paired = TRUE))

#p value < 0.05 表示超過信賴區間 拒絕 --> 完全不相通 --> 平均數差異顯著

#變異數統計(功能像F 檢定 ,要通過變異數相等才可以往下算平均)

with(iris, var.test(Sepal.Width, Sepal.Length))

#p value < 0.05 表示超過信賴區間 拒絕 --> 完全不相通 --> 變異數差異顯著

#回歸分析

data(cars)

head(cars)

#獲得回歸模型

#lm:線性回歸 {依變數~自變數}

m <- lm(dist~speed,data = cars) #m <- lm(dist~speed,cars)也行

summary(m)#右邊星星越多表示 p value越小 表示越準

#p value==0.01 就表示只有0.01的可能性會失敗, p value == 0.05 就達到了95%信心標準

#用變異數分析評估模型

anova(m)

#畫圖

plot(m)

#1.誤差服從平均數為0的常態分布理想狀態：一條斜率為0的直線

#2.查看殘差是否符合常態分布理想狀態：一條斜率為1的直線

#3.x軸為Y值，Y軸為標準化殘差

#4.右上與右下位置查找對迴歸直線形狀影響很大的點

plot(cars$speed,cars$dist)

abline(coef(m))

#coef函數用於提取線性迴歸模型的截距與斜率，abline()函數用於根據給定截距與斜率繪製直線

#多元分析

model <- lm(formula= Sepal.Length ~ Sepal.Width + Petal.Length + Petal.Width,data=iris) #(formula = 目標值 ~ 參數1+參數2+參數3+....)

summary(model)# y = 1.85 + 0.65\*Sepal.Width + 0.709\*Petal.Length - 0.556\*Petal.Width

#Residuals 殘差 其下是殘差分布 表示點和線的差異值 越大的區段表示越不准

#R-squared 表示 相關係數的平方 r^2

#殘差基本假設

model$residuals

#常態性假設驗證 --> 隨機分布 p value > 0.05 表示無法拒絕 具有常態性

shapiro.test(resid(model))

#獨立性假設驗證 --> 殘差彼此無關

library(car)

durbinWatsonTest(model)

#變異數同質性假設 --> 殘差屬於常態分布

ncvTest(model) #在此項 < 0.05 表示被拒絕 此模型不具有"變異數同質性"

#試用該公式

new.iris <- data.frame(Sepal.Width=3.456, Petal.Length=1.535, Petal.Width=0.341)

new.iris

predict(model, new.iris)#得到答案 = 5.004048

#變異數分析 #對品種不同的寬度進行變異數分析

#發現:pValue<0.5-->拒絕-->表示不同品種平均數有大差異 變異數分析的基本假設(各類別平均相等)

a.lm <- lm(Petal.Width~Species,data = iris)

anova(a.lm)