109-1 統計學實習作業 07

商研一 A09741303 鄭守開

一、匙上米廠為了控制生產的每包米重量都差不多，希望將米袋的重量變方控 制在 25 以下，因此每週會從生產線上隨機抽樣 10 包米，若是抽樣結果顯示重 量變方可能>25，則會重新校正裝米的機器。這週抽樣的結果如下，請問這週是 否有需要重新校正機器呢。(hint:檢定變方大小)

x = c(252.1, 244.6, 254.9, 253.4, 239.3, 246.4, 249.2, 241.7, 252.3, 237.2)

varTest(x, alternative = 'greater', sigma.squared = 625)

> varTest(x, alternative = 'greater', sigma.squared = 625)

Chi-Squared Test on Variance

data: x

Chi-Squared = 0.56277, df = 9, p-value = 0.9999

alternative hypothesis: true variance is greater than 625

95 percent confidence interval:

20.78902 Inf

sample estimates:

variance

39.081

=>P-value > α => 不拒絕H0

二、請根據’Marketing\_Campaign.csv’這筆資料進行以下分析，如果題目需要進 行假說檢定，請寫出該檢定的虛無假說、對立假說、檢定統計量為何並陳 述檢定結果

df <- read.csv("Marketing\_campaign.csv")

df

(1) 請問該資料的顧客的平均年齡、年齡標準差分別是多少呢

> mean(df$Year\_Birth)

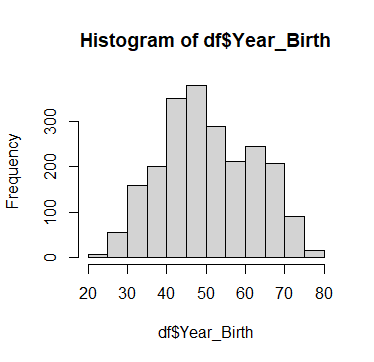
[1] 51.07327

> sd(df$Year\_Birth)

[1] 11.68807

(2) 請畫出平均年齡的 histogram

hist(df$Year\_Birth)



(3) 請問這家店顧客年齡的分布是否服從𝑁𝑜𝑟𝑚𝑎𝑙(50, 122 )的分布呢。請將顧客 資料依照𝑁𝑜𝑟𝑚𝑎𝑙(50, 122 )的 25%, 50%, 75%百分位數分成四組並進行適合度 檢定(α = 0.05)

age <- df$Year\_Birth

age

summary(age)

x = c(43, 50, 51.07, 61)

p = c(0.25, 0.25, 0.25, 0.25)

chisq.test(x = x, correct = TRUE, p = p)

> chisq.test(x = x, correct = TRUE, p = p)

Chi-squared test for given probabilities

data: x

X-squared = 3.2129, df = 3, p-value = 0.3599

=>𝑑𝑜 𝑛𝑜𝑡 𝑟𝑒𝑗𝑒𝑐𝑡 𝑯0

(4) 請問該資料的顧客的年收入平均、年收入標準差分別是多少呢

> mean(df$Income)

[1] 51959.18

> sd(df$Income)

[1] 21532.14

(5) 將顧客依照有沒有小孩區分成兩組，請畫出這兩組顧客的年收入的 boxplot

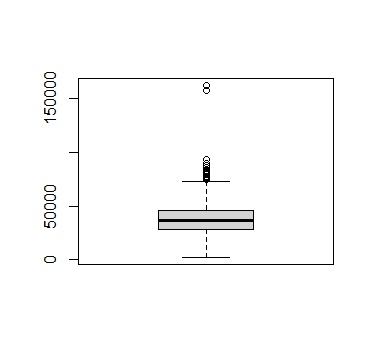
data\_with\_child <- subset(df, df$Kidhome !=0)

data\_without\_child <- subset(df, df$Kidhome == 0)

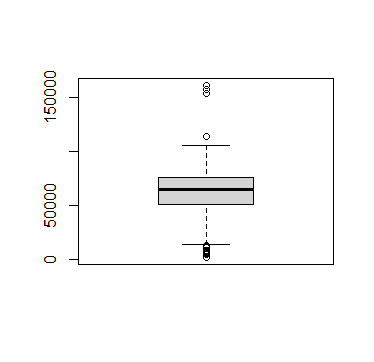
income\_with\_child <- data\_with\_child$Income

income\_without\_child <- data\_without\_child$Income

boxplot(income\_with\_child)



boxplot(income\_without\_child)



(6) 請問有小孩的客戶以及沒有小孩這兩組客戶的平均年收入是否相等呢(hint: 請進行兩樣本的均值檢定，α = 0.05)

t.test(x = income\_with\_child, y = income\_without\_child, alternative = 'two.sided', mu = 0, var.equal = T)

> t.test(x = income\_with\_child, y = income\_without\_child, alternative = 'two.sided', mu = 0, var.equal = T)

Two Sample t-test

data: income\_with\_child and income\_without\_child

t = -29.724, df = 2209, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-24842.68 -21767.59

sample estimates:

mean of x mean of y

38467.29 61772.42

* Reject H0 => 不相等

(7) 請問接受/沒有接受第一次促銷活動的人數分別是多少呢

customer\_accept1 <- subset(df, df$AcceptedCmp1 == 1)

customer\_not\_accept1 <- subset(df, df$AcceptedCmp1 == 2)

customer\_accept1 <- customer\_accept1$AcceptedCmp1

customer\_not\_accept1 <- customer\_not\_accept1$AcceptedCmp1

> table(customer\_accept1)

customer\_accept1

1

142

> table(customer\_not\_accept1)

< table of extent 0 >

(8) 請問有小孩的客戶以及沒有小孩的客戶對於第一次促銷活動的接受與否，反 應是否相同，請先根據有小孩/沒有小孩劃分兩組樣本，依照接受活動/沒有 接受活動這兩類，檢定兩樣本的同質性(α = 0.05)

child\_accept1 <- subset(data\_with\_child, data\_with\_child$AcceptedCmp1 == 1)

child\_accept1

child\_not\_accept1 <- subset(data\_with\_child, data\_with\_child$AcceptedCmp1 == 2)

child\_not\_accept1

nrow(child\_accept1)

nrow(child\_not\_accept1)

without\_child\_accept1 <- subset(data\_without\_child, data\_without\_child$AcceptedCmp1 == 1)

without\_child\_accept1

without\_child\_not\_accept1 <- subset(data\_without\_child, data\_without\_child$AcceptedCmp1 == 2)

without\_child\_not\_accept1

nrow(without\_child\_accept1)

nrow(without\_child\_not\_accept1)

x1 = c(10, 0)

x2 = c(132, 0)

x = cbind(x1, x2)

chisq.test(x)

> chisq.test(x)

Pearson's Chi-squared test

data: x

X-squared = NaN, df = 1, p-value = NA