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

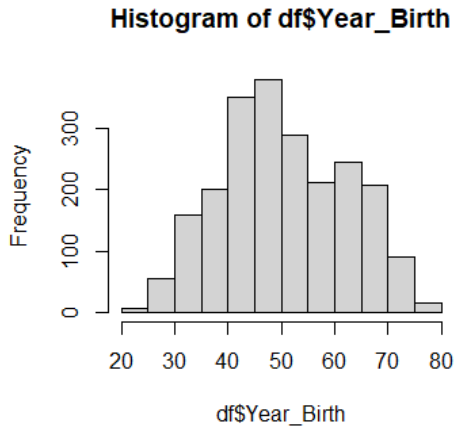
```
df <- read.csv("Marketing_campaign.csv")
df
```

(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
```

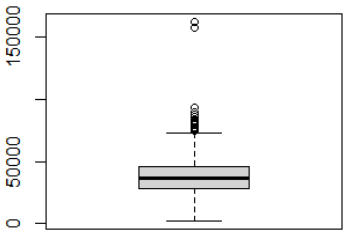
=>$do\ not\ reject\ \boldsymbol{H}0$

(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

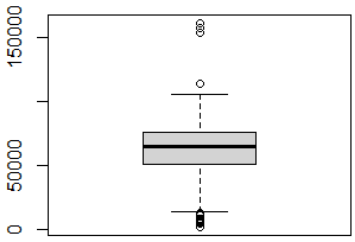
```
> mean(df$Income)
[1] 51959.18

> 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)

```
> nrow(child_accept1)
[1] 10
> nrow(child_not_accept1)
[1] 0
> nrow(without_child_accept1)
[1] 132
> nrow(without_child_not_accept1)
[1] 0
> x1 = c(10, 0)
> x2 = c(132, 0)
> x = cbind(x1, x2)
> chisq.test(x)

        Pearson's Chi-squared test

data:  x
X-squared = NaN, df = 1, p-value = NA

Warning message:
In chisq.test(x) : Chi-squared approximation may be incorrect
> x1 = c(10, 0)
> x2 = c(132, 0)
> x = cbind(x1, x2)
> chisq.test(x)

        Pearson's Chi-squared test

data:  x
X-squared = NaN, df = 1, p-value = NA

Warning message:
In chisq.test(x) : Chi-squared approximation may be incorrect
> x1 = c(10, 0)
> x2 = c(132, 0)
> x = cbind(x1, x2)
> chisq.test(x)

        Pearson's Chi-squared test
```

```
data:  x
X-squared = NaN, df = 1, p-value = NA

Warning message:
In chisq.test(x) : Chi-squared approximation may be incorrect
> # install.packages("dplyr")
> # library(tidyverse)
> # library(dslabs)
> install.packages('EnvStats')
WARNING: Rtools is required to build R packages but is not currently installed. Please do
wnload and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Paul/Documents/R/win-library/4.0'
(as 'lib' is unspecified)
嘗試 URL 'https://cran.rstudio.com/bin/windows/contrib/4.0/EnvStats_2.4.0.zip'
Content type 'application/zip' length 6057090 bytes (5.8 MB)
downloaded 5.8 MB

package 'EnvStats' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Paul\AppData\Local\Temp\RtmpQrYwby\downloaded_packages
> library(EnvStats)

Attaching package: 'EnvStats'

The following objects are masked from 'package:stats':

    predict, predict.lm

The following object is masked from 'package:base':

    print.default

Warning message:
package 'EnvStats' was built under R version 4.0.3
> # (一)
> 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 = 'smaller', sigma.squared = 625)
Error in match.arg(alternative, c("two.sided", "less", "greater")) :
  'arg' should be one of "two.sided", "less", "greater"
> # (一)
> 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 = 'less', sigma.squared = 625)

        Chi-Squared Test on Variance

data:  x
Chi-Squared = 0.56277, df = 9, p-value = 5.051e-05
alternative hypothesis: true variance is less than 625
95 percent confidence interval:
   0.0000 105.7796
sample estimates:
variance
  39.081

> 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

> # (一)
> 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)

        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

> age <- df$Year_Birth
> age
> x = age
> p = c(0.25, 0.25, 0.25, 0.25)
> chisq.test(x = x, correct = TRUE, p = p)
Error in chisq.test(x = x, correct = TRUE, p = p) :
  'x' and 'p' must have the same number of elements
> table(age)
age
24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
 2  5  3  5 13 15 18 29 29 27 41 32 38 41 44 38 39 53 76 51
44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63
89 83 69 72 78 86 75 70 51 44 50 74 41 44 44 35 49 50 52 41
64 65 66 67 68 69 70 71 72 73 74 75 76 77 79
55 48 49 35 52 42 29 30 21 16 16  8  7  6  1
> summary(age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  24.00   43.00   50.00   51.07   61.00   79.00
> 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)

        Chi-squared test for given probabilities

data:  x
X-squared = 3.2129, df = 3, p-value = 0.3599

> 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
```

```
> 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
```