三軍總醫院北投分院統計及實驗設計課程之二

2021/6/26

ytai1123@gmail.com

使用方法:

1. 使用gmail帳號登入
2. 按"執行階段" -->"全部執行" 以執行全部內容, 若要個別執行可點選每格程式左方箭頭或按 Control + Enter 鍵執行。

```
##0-1
!git  clone  https://github.com/YuehMintTai/RPython.git
```

```
Cloning into 'RPython'...
remote: Enumerating objects: 47, done.
remote: Counting objects: 100% (47/47), done.
remote: Compressing objects: 100% (45/45), done.
remote: Total 47 (delta 14), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (47/47), done.
```

```
##0-2
!pip  install  rpy2
```

```
Requirement already satisfied: rpy2 in /usr/local/lib/python3.7/dist-packages (3.4.5)
Requirement already satisfied: pytz in /usr/local/lib/python3.7/dist-packages (from rpy2) (20
Requirement already satisfied: tzlocal in /usr/local/lib/python3.7/dist-packages (from rpy2)
Requirement already satisfied: cffi>=1.10.0 in /usr/local/lib/python3.7/dist-packages (from r
Requirement already satisfied: jinja2 in /usr/local/lib/python3.7/dist-packages (from rpy2) (
Requirement already satisfied: pycparser in /usr/local/lib/python3.7/dist-packages (from cffi
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-packages (fr
```

```
%load_ext  rpy2.ipython
```

```
##0-3
%%R
myData  <-  read.csv('RPython/samples.csv')
myData[1,]
```

```
  SID 性別      年齡 入伍前職業 教育程度 婚姻狀況 皆無過去病史01 早產兒01
1 137     1 21.33308       學生         4         1              0         0
  頭部曾受傷01 發展遲緩01 注意力不足過動症01 癲癇01 癲癇服藥治療 癲癇服藥期間
1            1          0                  0       0            0            0
  軍種 軍階 役別 入伍至今_年 聽過自殺課程_次 求助心輔_次 求助精神科_次
1   11    1    2         0.3               0           1             1
```

```
  使用1995_次 使用24h專線_次 特殊狀況 父母婚姻狀態 自殺意念_bsrs6 B型肝炎01
1           0             0        4           1              0          0
  C型肝炎01 氣喘史01 過敏史01 心臟病史01 高血壓01 醣尿病01 甲狀腺01 類風濕01
1         0        0        1          0        0        0        0        0
  重大意外01 自殺意念01 透露父母 透露手足 透露好友 透露同儕 透露長官 透露心輔
1          1          0        0        0        1        0        0        0
  透露醫師 拒告父母 拒告手足 拒告好友 拒告同儕 拒告長官 拒告心輔 拒告醫師
1        0        1        0        0        0        1        0        0
  BSRS總分 BSRSR總分 過動症總分 Inattention Impulsivity opposition depression
1        1         1         23          21           2          9         13
  anxiety burdensome belonging 家庭滿意度apgar 網路成癮症01 網路成癮分數YDQ
1      53         23        19              10            0               2
  existeness meaning control seeking death suicidea 睡眠困擾_bsrs1
1         18      18      11      18    15        0              0
  睡眠困擾_bsrsr1 睡眠困擾_bdi16 易怒_bsrs3 易怒_bsrsr3 depress impuls Internet
1               0              0          0           0      13      2        2
  ADHD
1   23
```

```
##2-1
%%R
formula='網路成癮分數YDQ  ~  家庭滿意度apgar'
model1<-  glm(formula,  myData,  family='gaussian')
summary(model1)
```

```
Call:
glm(formula = formula, family = "gaussian", data = myData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5125  -1.8546  -0.8546   1.8138   6.1454

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.51255    0.45740   7.679 8.82e-13 ***
家庭滿意度apgar -0.16580    0.05836  -2.841    0.005 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 5.856059)

    Null deviance: 1136.5  on 187  degrees of freedom
Residual deviance: 1089.2  on 186  degrees of freedom
AIC: 869.8

Number of Fisher Scoring iterations: 2
```

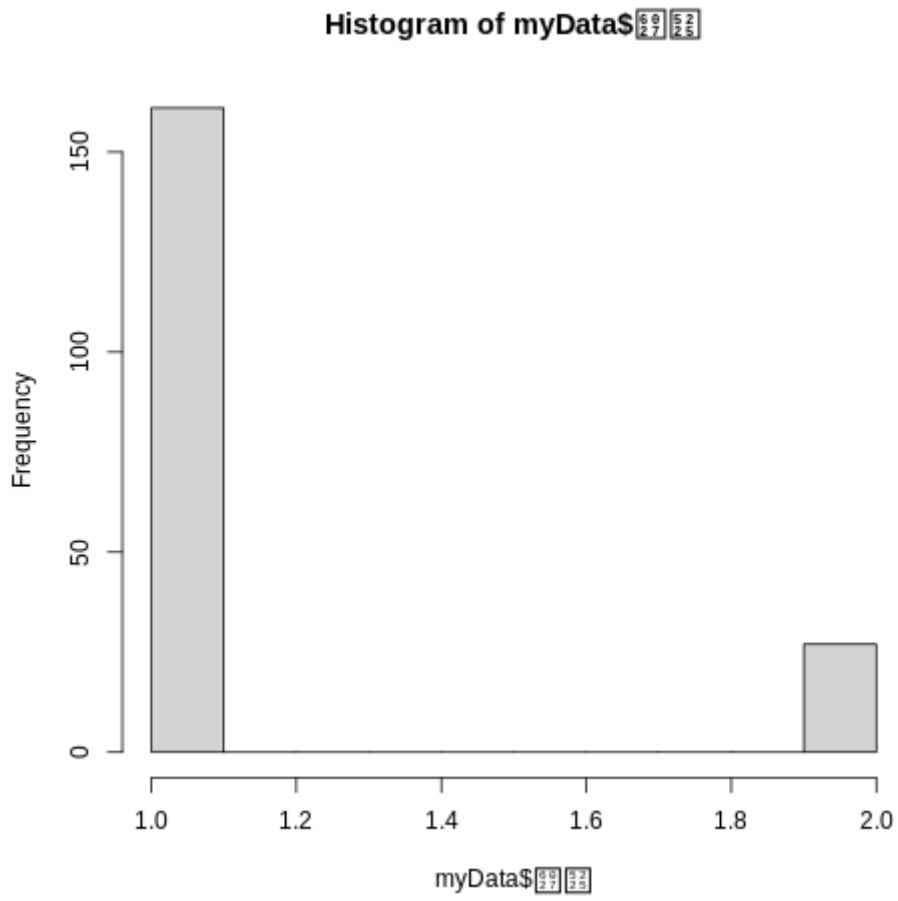
```
##2-2
%%R
formula='網路成癮分數YDQ  ~  性別'
model2  <-  glm(formula,  myData,  family='gaussian')
summary(model2)
```

```
Call:
glm(formula = formula, family = "gaussian", data = myData)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4783  -2.4783  -0.4783   1.5217   5.6667

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.6232     0.6066   5.972 1.16e-08 ***
性別         -1.1449     0.5072  -2.258   0.0251 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 5.947172)

    Null deviance: 1136.5  on 187  degrees of freedom
Residual deviance: 1106.2  on 186  degrees of freedom
AIC: 872.7

Number of Fisher Scoring iterations: 2
```

##2-3

```
%%R
hist(myData$性別)
```



##2-4

```
%%R
##myData$sex.f<-factor(myData$性別)
##formula='網路成癮分數YDQ  ~  sex.f'
formula='網路成癮分數YDQ  ~  as.factor(性別)'
```

```
formula = 縣鄉成應分數TDS ~ as.factor(性別)
model2 <- glm(formula, myData, family='gaussian')
summary(model2)
```

```
Call:
glm(formula = formula, family = "gaussian", data = myData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4783  -2.4783  -0.4783   1.5217   5.6667

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       2.4783     0.1922  12.895   <2e-16 ***
as.factor(性別)2  -1.1449     0.5072  -2.258   0.0251 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 5.947172)

    Null deviance: 1136.5  on 187  degrees of freedom
Residual deviance: 1106.2  on 186  degrees of freedom
AIC: 872.7

Number of Fisher Scoring iterations: 2
```

##2-5

```
%%R
myData$性別<-relevel(factor(myData$性別),ref='2')
model3 <- glm(formula, myData, family='gaussian')
summary(model3)
```

```
Call:
glm(formula = formula, family = "gaussian", data = myData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4783  -2.4783  -0.4783   1.5217   5.6667

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.3333     0.4693   2.841   0.0050 **
as.factor(性別)1  1.1449     0.5072   2.258   0.0251 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 5.947172)

    Null deviance: 1136.5  on 187  degrees of freedom
Residual deviance: 1106.2  on 186  degrees of freedom
AIC: 872.7

Number of Fisher Scoring iterations: 2
```

##2-6

```
%%R
t.test(myData$網路成癮分數YDQ  ~  myData$性別)
```

```
	Welch Two Sample t-test

data:  myData$網路成癮分數YDQ by myData$性別
t = -2.9195, df = 46.432, p-value = 0.005393
alternative hypothesis: true difference in means between group 2 and group 1 is not equal to
95 percent confidence interval:
 -1.9341234 -0.3557317
sample estimates:
mean in group 2 mean in group 1
       1.333333        2.478261
```

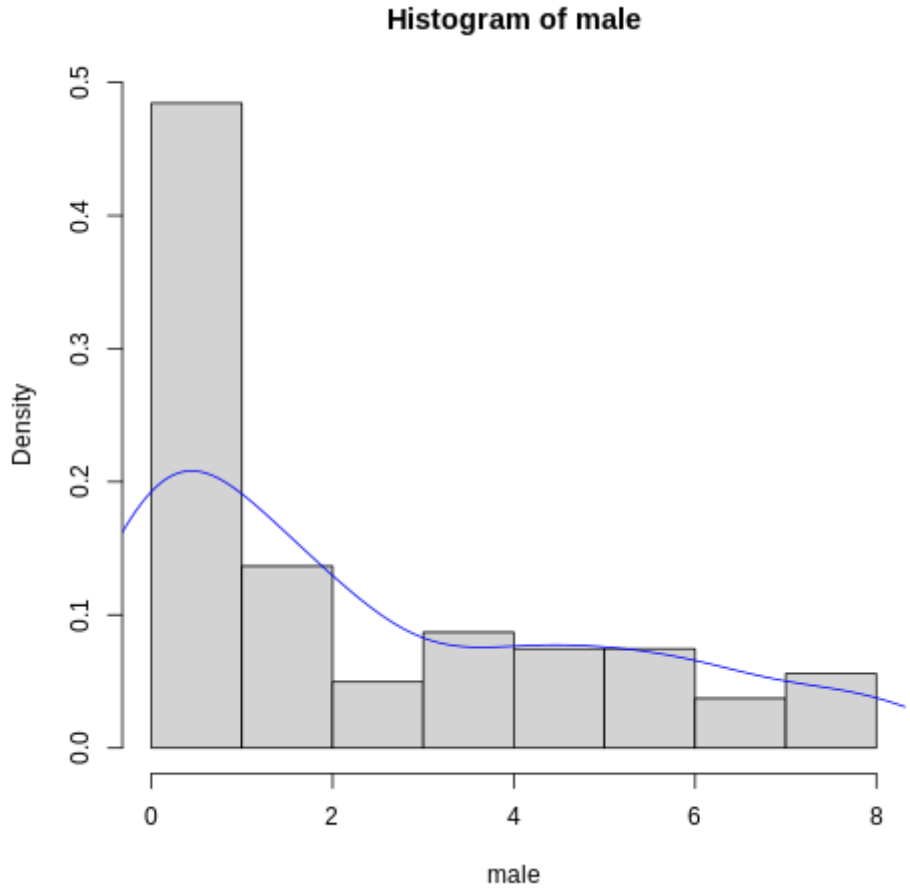
```
##2-7
%%R
male=subset(myData,  性別==1)$網路成癮分數YDQ
female=subset(myData,  性別==2)$網路成癮分數YDQ
t.test(male,female)
```

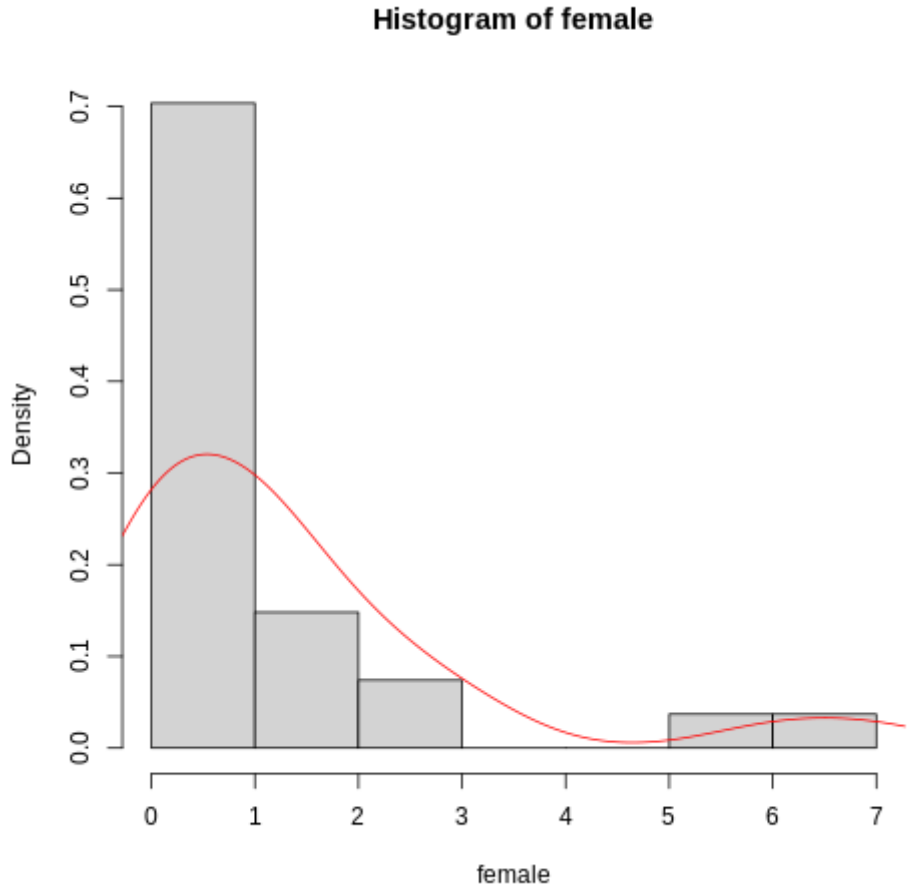
```
	Welch Two Sample t-test

data:  male and female
t = 2.9195, df = 46.432, p-value = 0.005393
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3557317 1.9341234
sample estimates:
mean of x mean of y
 2.478261  1.333333
```

```
##2-7-1使用python  stasmodels作t-test
import  pandas  as  pd
from  statsmodels.stats.weightstats  import  ttest_ind
df=pd.read_csv('RPython/samples.csv')
male=df['網路成癮分數YDQ'][df['性別']==1]
female=df['網路成癮分數YDQ'][df['性別']==2]
result=ttest_ind(male,female,alternative='two-sided',usevar='unequal')
result
```

```
(2.9194792899491007, 0.00539309754637806, 46.43217543281508)
```

```
##2-8
%%R
hist(male,  prob=TRUE)
lines(density(male),col='blue')
hist(female,  prob=TRUE)
lines(density(female),col='red')
```

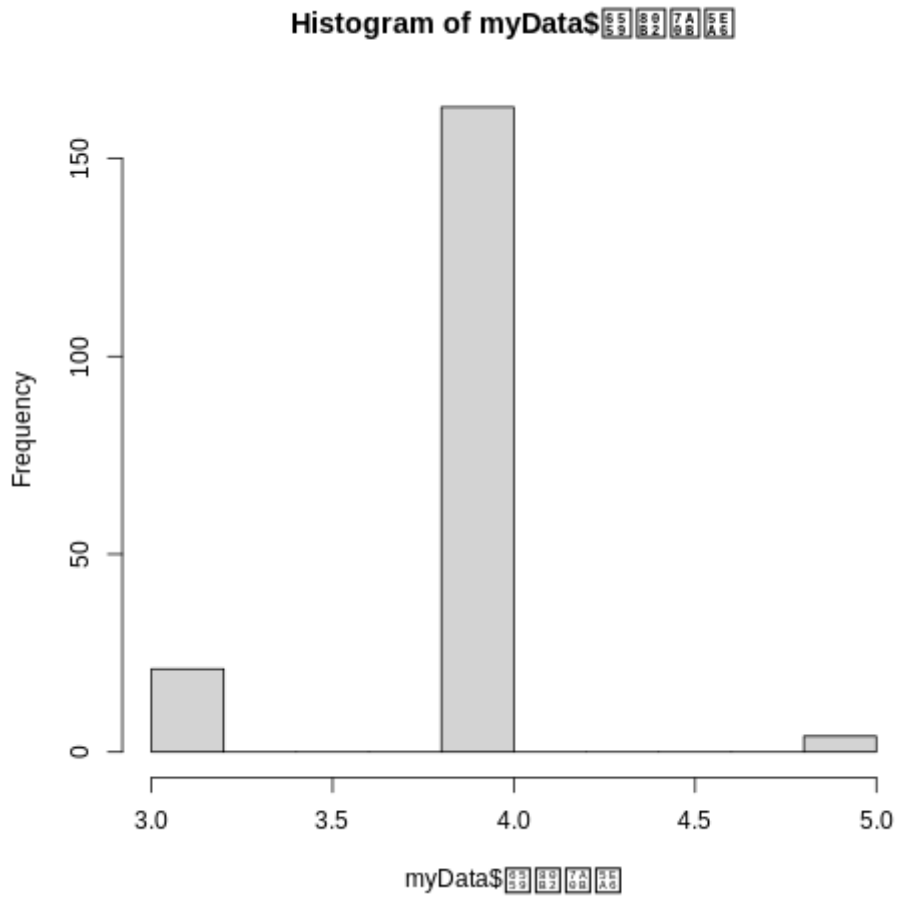




##2-9

%%R

hist(myData$教育程度)




```
##2-10
%%R
formula='網路成癮分數YDQ  ~  as.factor(教育程度)'
model2  <-  glm(formula,  myData,  family='gaussian')
summary(model2)
```

```
Call:
glm(formula = formula, family = "gaussian", data = myData)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-3.667  -2.178  -1.178   1.822   5.822

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)           3.6667     0.5285   6.937  6.5e-11 ***
as.factor(教育程度)4  -1.4888     0.5615  -2.651  0.00872 **
as.factor(教育程度)5  -2.9167     1.3213  -2.207  0.02852 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 5.866255)

    Null deviance: 1136.5  on 187  degrees of freedom
Residual deviance: 1085.3  on 185  degrees of freedom
AIC: 871.11

Number of Fisher Scoring iterations: 2
```

```
##2-11
%%R
myData$教育程度<-relevel(factor(myData$教育程度),ref='4')
formula='網路成癮分數YDQ  ~  as.factor(教育程度)'
model4  <-  glm(formula,  myData,  family='gaussian')
summary(model4)
```

```
Call:
glm(formula = formula, family = "gaussian", data = myData)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-3.667  -2.178  -1.178   1.822   5.822

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            2.1779     0.1897  11.480  < 2e-16 ***
as.factor(教育程度)3    1.4888     0.5615   2.651  0.00872 **
as.factor(教育程度)5   -1.4279     1.2258  -1.165  0.24556
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 5.866255)

    Null deviance: 1136.5  on 187  degrees of freedom
Residual deviance: 1085.3  on 185  degrees of freedom
AIC: 871.11

Number of Fisher Scoring iterations: 2
```

```
##2-12  ##python  GLM  test
import  statsmodels.api  as  sm
import  statsmodels.formula.api  as  smf
formula='網路成癮分數YDQ  ~  C(教育程度)'
model5=smf.glm(formula,df).fit()
model5.summary()
```

```
##2-13 Simple One-way ANOVA in R
%%R
model6 <- aov(網路成癮分數YDQ ~ as.factor(教育程度),data=myData)
print(summary(model6))
TukeyHSD(model6)
```

```
                      Df Sum Sq Mean Sq F value Pr(>F)
as.factor(教育程度)    2   51.2  25.613   4.366  0.014 *
Residuals           185 1085.3   5.866
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = 網路成癮分數YDQ ~ as.factor(教育程度), data = myData)

$`as.factor(教育程度)`
         diff        lwr       upr     p adj
3-4  1.488753  0.1619682 2.8155369 0.0235826
5-4 -1.427914 -4.3241193 1.4682910 0.4756670
5-3 -2.916667 -6.0386116 0.2052783 0.0725632
```

```
!pip3 install bioinfokit
```

```
##2-14 python anova test
import statsmodels.api as sm
from statsmodels.formula.api import ols
formula='網路成癮分數YDQ ~ C(教育程度)'
model7=ols(formula, df).fit()
anova_table=sm.stats.anova_lm(model7, type=2)
print(anova_table)
from bioinfokit.analys import stat
res=stat()
res.tukey_hsd(df=df,res_var='網路成癮分數YDQ',
                          xfac_var='教育程度',
                          anova_model=formula)
res.tukey_summary
```

```
                df        sum_sq    mean_sq         F    PR(>F)
C(教育程度)      2.0     51.226885  25.613443  4.366234  0.014033
Residual  185.0  1085.257157   5.866255       NaN       NaN
```

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	4	3	1.488753	0.161875	2.815630	3.749312	0.023579
1	4	5	1.427914	-1.468494	4.324322	1.647412	0.477314
2	3	5	2.916667	-0.205497	6.038830	3.121699	0.072656