**IMBS 1** **ISCID-CO Dunkerque/ULCO**

**Mathematics applied to economics and management**
Foundations of Descriptive and Inferential Statistics

December 2015 - Final assessment - Session 1 - Semester 1
Time allowed : 1h30 - All documents allowed

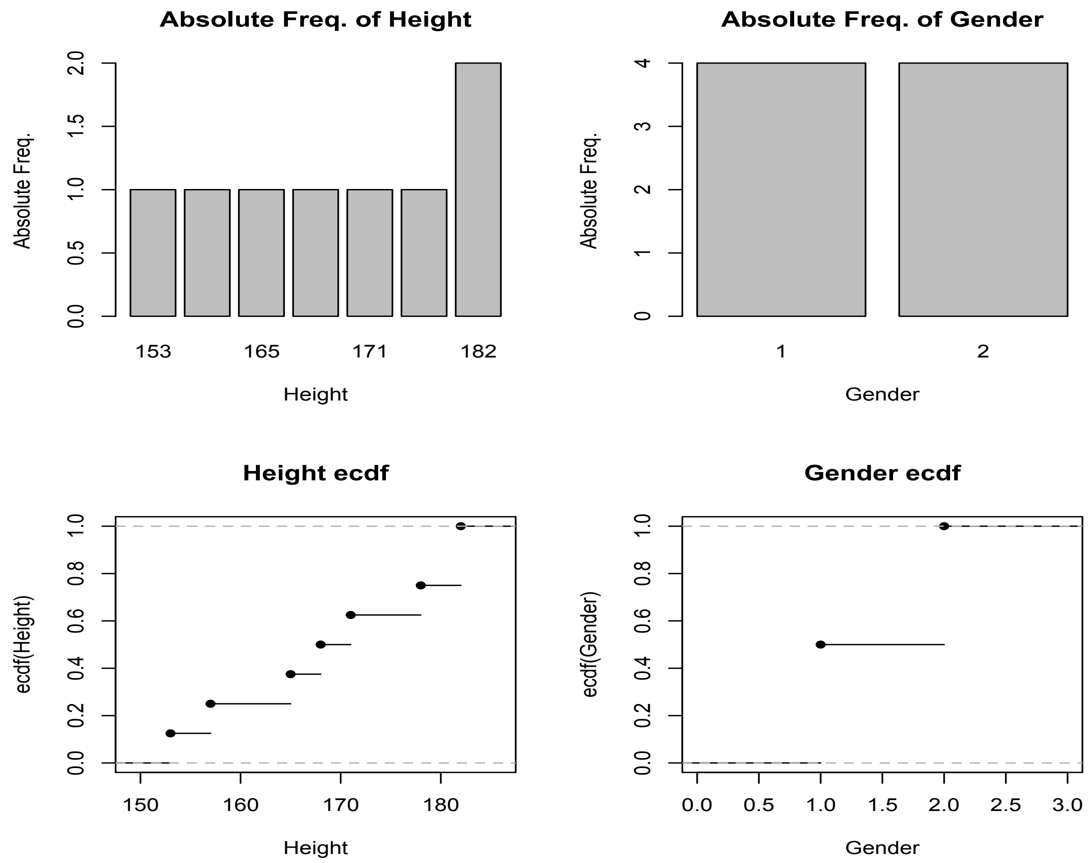
**Exercice 1** *Answer* : 7pts

- 2pts Frequency tables

```
> Height<-c(165,157,168,178,171,182,182,153)
> Gender<-c("1","2","2","1","2","1","1","2")
> d1<-data.frame(table(Height))
> d11<-data.frame(d1,"Percent.Freq."=d1$Freq/length(Height)*100,
+ "Cumulative.Freq."=cumsum(d1$Freq))
> # At last, we may modify the label of the second column
> names(d11)[2]<-"Absolute.Freq."
> d11
   Height  Absolute.Freq.  Percent.Freq.  Cumulative.Freq.
 1   153          1             12.5              1
 2   157          1             12.5              2
 3   165          1             12.5              3
 4   168          1             12.5              4
 5   171          1             12.5              5
 6   178          1             12.5              6
 7   182          2             25.0              8
> d2<-data.frame(table(Gender))
> d22<-data.frame(d2,"Percent.Freq."=d2$Freq/length(Height)*100,
+ "Cumulative.Freq."=cumsum(d2$Freq))
> # At last, we may modify the label of the second column
> names(d22)[2]<-"Absolute.Freq."
> d22
   Gender  Absolute.Freq.  Percent.Freq.  Cumulative.Freq.
 1    1           4              50               4
 2    2           4              50               8
```

- 2pts Graphs

```
> par(mfrow=c(2,2))
> barplot(table(Height),xlab="Height",ylab="Absolute Freq.",main="Absolute
+ Freq. of Height")
> barplot(table(Gender),xlab="Gender",ylab="Absolute Freq.",main="Absolute
+ Freq. of Gender")
> plot(ecdf(Height),xlab="Height",ylab="ecdf(Height)",main="Height ecdf")
> plot(ecdf(Gender),xlab="Gender",ylab="ecdf(Gender)",main="Gender ecdf")
```




- $\boxed{\text{1pt}}$ Contingency table

```
> table(Gender,Height)
          Height
 Gender     153    157  165  168  171  178  182
    1        0      0    1    0    0    1    2
    2        1      1    0    1    1    0    0
```

- $\boxed{\text{2pts}}$ The two variables we want to compare do not belong to the same scale levels. The variable "Height" is metrically scaled while the variable "Gender" is nominally scaled. In order to describe the relation between Gender and Height, we may refer to *http ://pbil.univ-lyon1.fr/R/pdf/tdr208.pdf* and use the following commands :

```
> vartot <- function(x) {res <- sum((x-mean(x))∧2) ; return(res)}
> varinter <- function(x,gpe){
moyennes <- tapply(x,gpe,mean)
effectifs <- tapply(x,gpe,length)
res <- (sum(effectifs*(moyennes-mean(x))∧2))
return(res)
}
> eta2 <- function(x,gpe){res <- varinter(x,gpe)/vartot(x) ; return(res)}
> eta2(Height,Gender)
[1]  0.50179
```

The result is not significant enough to state or not a link between the two variables. Nevertheless, we may provide some descriptive measures for each distribution with respect to the gender :

```
> split(Height,Gender)
```

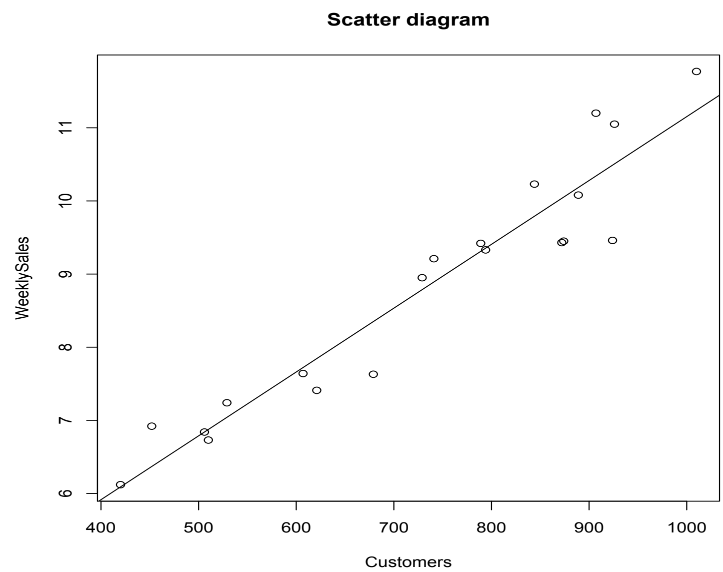
```
$‘1‘
[1]  165 178 182 182

$‘2‘
[1]  157 168 171 153

> summary(sp$‘1‘)
  Min.   1st Qu.  Median   Mean  3rd Qu.   Max.
 165.0    174.8    180.0  176.8   182.0   182.0
> summary(sp$‘2‘)
  Min.   1st Qu.  Median   Mean  3rd Qu.   Max.
 153.0    156.0    162.5  162.2   168.8   171.0
```

Since the measures are quite different, we may conclude that relatively to our distribution, the gender affects strongly the height.

**Exercice 2** *Answer* : 4pts

1. 1pt

```
> Customers<-c(907,926,506,741,789,889,874,510,529,1010,
+ 420,679,872,924,607,452,729,794,844,621)
> WeeklySales<-c(11.20,11.05,6.84,9.21,9.42,10.08,9.45,6.73,7.24,11.77,
+ 6.12,7.63,9.43,9.46,7.64,6.92,8.95,9.33,10.23,7.41)
> plot(WeeklySales~Customers,main="Scatter diagram")
```




2. 1pt

```
> model<-lm(WeeklySales~Customers)
> model

Call :
lm(formula = WeeklySales~Customers)

Coefficients :
 (Intercept)  Customers
   2.423044    0.008729
```

We deduce from the previous results that $b_0 = 2.423044$ and $b_1 = 0.008729$. We may plot the trend line whose equation is $Y = 0.008729X + 2.423044$ help to the command :

```
> abline(model)
```

3. $\boxed{\text{1pt}}$ The slope $b_1$ may be interpreted as the marginal value : when Customers increase to $t\%$, the Weekly Sales increase to $0.008729 \times t\%$.
4. $\boxed{\text{1pt}}$ The average weekly Sales (in thousands in dollars) for stores that have 600 customers may be estimated help to the following command :

```
> predict(model,list(Customers=c(600)))

1
7.660647
```

**Exercice 3** *Answer* : $\boxed{\text{4pts}}$

1. $\boxed{\text{1pt}}$

```
> dim(iris)
[1]150 5
> str(iris)
> str(iris)
 'data.frame' :    150 obs. of 5 variables :
 $ Sepal.Length :  num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width :   num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length :  num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width :   num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species :       Factor w/ 3 levels "setosa","versicolor",.. : 1 1 1 1 1 1 1 1 1 1 ...
```

2. $\boxed{\text{1pt}}$

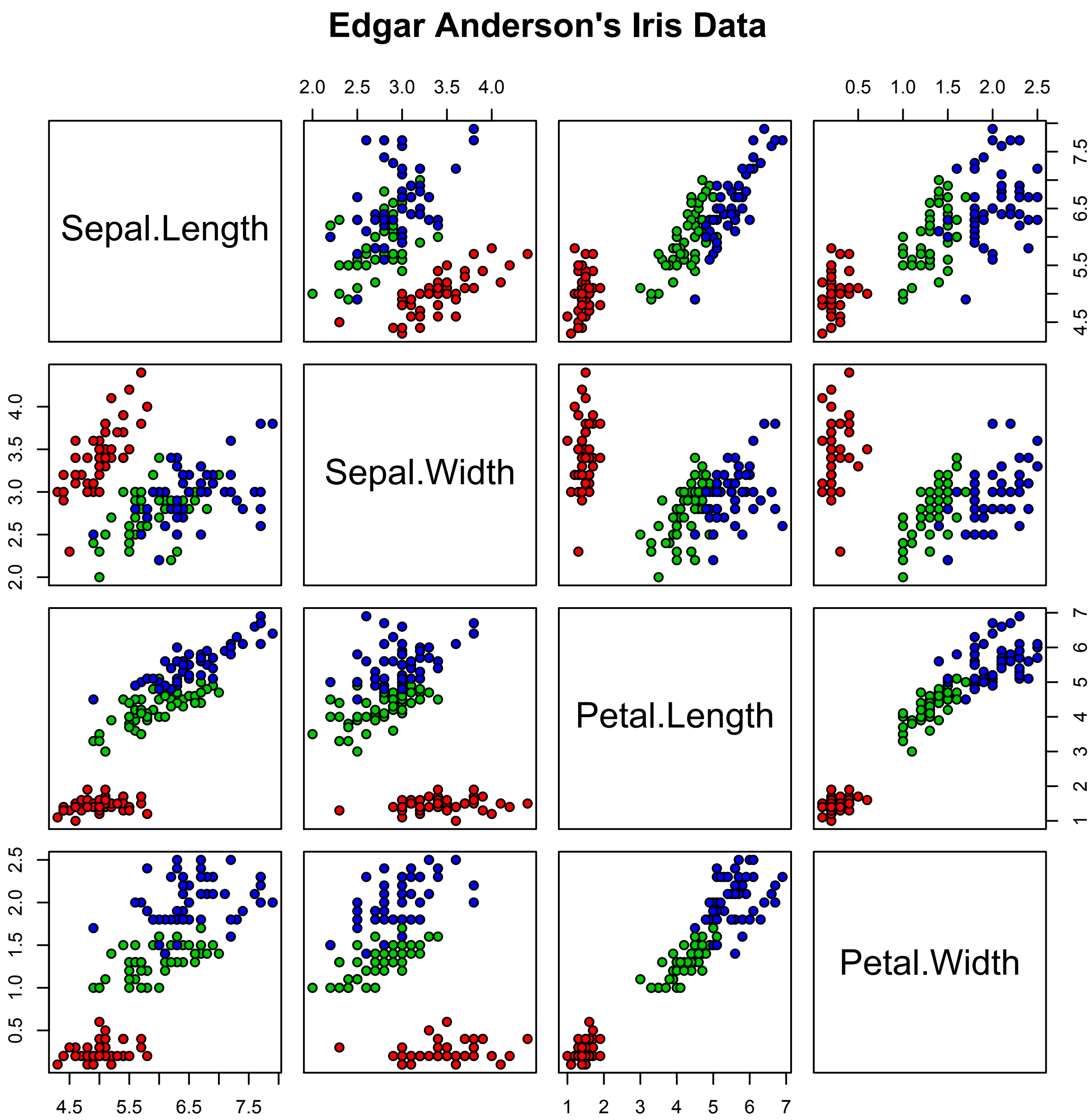
```
> summary(iris)
 Sepal.Length    Sepal.Width     Petal.Length    Petal.Width     Species
 Min. :4.300     Min. :2.000     Min. :1.000     Min. :0.100     setosa :50
 1st Qu. :5.100  1st Qu. :2.800  1st Qu. :1.600  1st Qu. :0.300  versicolor :50
 Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50
 Mean :5.843     Mean :3.057     Mean :3.758     Mean :1.199
 3rd Qu. :6.400  3rd Qu. :3.300  3rd Qu. :5.100  3rd Qu. :1.800
 Max. :7.900     Max. :4.400     Max. :6.900     Max. :2.500
```

3. $\boxed{\text{1pt}}$ The answer was available on internet, see

*http ://www2.warwick.ac.uk/fac/sci/moac/people/students/peter_cock/r/iris_plots/*

for example. We may use the command :

```
> pairs(iris[1 :4], main = "Edgar Anderson's Iris Data", pch = 21, bg = c("red",
+ "green3","blue")[unclass(iris$Species)])
```




4. 1pt `> cor(iris[1 :4])`

```
             Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
Sepal.Length    1.0000000   -0.1175698     0.8717538    0.8179411
Sepal.Width    -0.1175698    1.0000000    -0.4284401   -0.3661259
Petal.Length    0.8717538   -0.4284401     1.0000000    0.9628654
Petal.Width     0.8179411   -0.3661259     0.9628654    1.0000000
```

**Exercice 4** *Answer* : 5pts

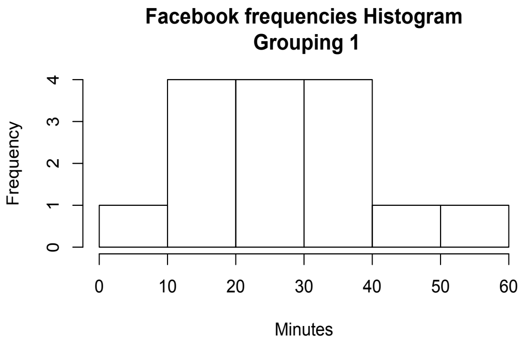
- 2,5pts Frequency distributions

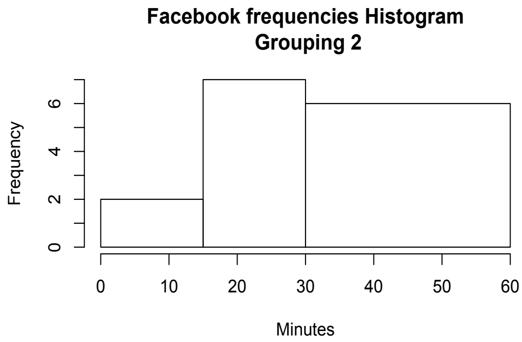
```
> Facebook<-c(0,11,17,16,22,17,25,30,27,27,31,35,30,45,60)
> Statistics<-c(25,22,28,30,22,27,26,21,27,28,23,29,30,24,27)
> class1X<-c(0,10,20,30,40,50,60)
> class2X<-c(0,15,30,60)
> class1Y<-c(18,22,26,30)
> class2Y<-c(18,21,24,27,30)
```

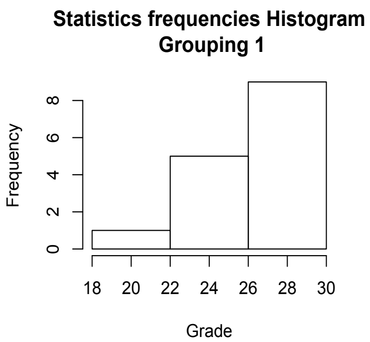
```
> class3Y<-c(18,20,22,24,26,28,30)
> table(cut(Facebook,class1X,right=F,include.lowest=T))

 [0,10)  [10,20)  [20,30)  [30,40)  [40,50)  [50,60]
      1        4        4        4        1        1

> table(cut(Facebook,class2X,right=F,include.lowest=T))

 [0,15)  [15,30)  [30,60]
      2        7        6

> table(cut(Statistics,class1Y,right=F,include.lowest=T))

 [18,22)  [22,26)  [26,30]
       1        5        9

> table(cut(Statistics,class2Y,right=F,include.lowest=T))

 [18,21)  [21,24)  [24,27)  [27,30]
       0        4        3        8

> table(cut(Statistics,class3Y,right=F,include.lowest=T))

 [18,20)  [20,22)  [22,24)  [24,26)  [26,28)  [28,30]
       0        1        3        2        4        5
```

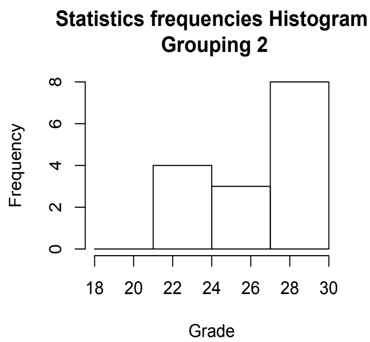
- $\boxed{\text{2,5pts}}$ Histograms

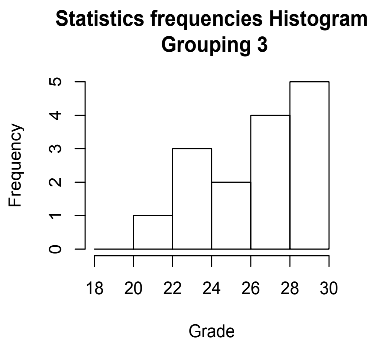
```
> split.screen( figs = c( 2, 1 ) )
[1]  1 2
> split.screen( figs = c( 1, 2 ), screen = 1 )
[1]  3 4
> split.screen( figs = c( 1, 3 ), screen = 2 )
[1]  5 6 7
> screen(3)
> hist(Facebook,breaks=class1X,right=F,main="Facebook frequencies Histogram\n
+ Grouping 1",xlab="Minutes",freq=T,include.lowest=T)
> screen(4)
> hist(Facebook,breaks=class2X,right=F,main="Facebook frequencies Histogram\n
+ Grouping 2",xlab="Minutes",freq=T,include.lowest=T)
> screen(5)
> hist(Statistics,breaks=class1Y,right=F,main="Statistics frequencies Histogram\n
+ Grouping 1",xlab="Grade",freq=T,include.lowest=T)
> screen(6)
> hist(Statistics,breaks=class2Y,right=F,main="Statistics frequencies Histogram\n
+ Grouping 2",xlab="Grade",freq=T,include.lowest=T)
> screen(7)
> hist(Statistics,breaks=class3Y,right=F,main="Statistics frequencies Histogram\n
+ Grouping 3",xlab="Grade",freq=T,include.lowest=T)
```


Facebook frequencies Histogram
Grouping 1
Frequency
Minutes


Facebook frequencies Histogram
Grouping 2
Frequency
Minutes


Statistics frequencies Histogram
Grouping 1
Frequency
Grade


Statistics frequencies Histogram
Grouping 2
Frequency
Grade


Statistics frequencies Histogram
Grouping 3
Frequency
Grade