SCI03 - Analyse de données expérimentales
Statistique descriptive

Thierry Denœux[1]

[1]Université de Technologie de Compiègne
tél : 44 96
tdenoeux@hds.utc.fr

Automne 2014

Exemples : résultats SY02

```
> sy02
median correcteur1 final correcteur2 moyenne resultat

1 8.5 SK 14.0 FA 11.8 C
2 9.5 SK 11.5 FA 10.7 C
3 1.5 SK 4.5 FA 3.3 F
4 14.5 SK 9.5 FA 11.5 C
5 5.5 SK 11.0 FA 8.8 E
6 9.5 SK 6.0 FA 7.4 FX
7 5.0 SK 0.0 FA 2.0 F
8 5.5 SK 13.0 FA 10.0 D
9 3.5 SK 14.0 FA 9.8 D
10 6.0 SK 13.5 FA 10.5 C
11 6.0 SK 8.0 FA 7.2 FX
12 7.0 SK 13.5 FA 10.9 C
13 8.0 SK 7.0 FA 7.4 FX
...
```

Tableau de fréquences

```
> sy02$resultat
  [1] C C F C E FX F D D C FX C FX F C D E A A B F A B C A B E C F C E F D E B D
 [37] E C D D D D F B F D C C B B F C C A D B B FX E FX B A B C B B A E D B B
 [73] B E B B C F E A D C B C C B B B FX FX E E D FX E D C D E B C D A D D B C D
[109] F FX E D F A D C D C E FX D C F D C F C A C B B E B A A A A E C F A F C
[145] A D E A C B B A F A A D E B A B E C B B C C F F A D FX C FX E D D FX A E D
[181] E E FX FX C E E B FX D C A FX A F FX E C B C B B D D F F D D E FX C C FX E C E
[217] E B FX C B A C F A E B E D E D F C FX A C E C F F B F C C F D A C B C D A
[253] FX FX D B E A FX E A A C E B E C F FX C B D C A D D A D E A F F FX C E A C B
[289] E B A A E B D E
Levels: A B C D E F FX

> ects=factor(sy02$resultat,levels=c("A","B","C","D","E","FX","F"),ordered=TRUE);

> T=table(ects)
> T
ects
 A  B  C  D  E FX  F
41 48 57 47 45 27 31

> f=100*T/sum(T)
> print(f,digits=2)
 A   B   C   D   E  FX   F
13.9 16.2 19.3 15.9 15.2 9.1 10.5

> F=cumsum(f)
> print(F,digits=3)
 A    B    C    D    E   FX    F
13.9 30.1 49.3 65.2 80.4 89.5 100.0
```
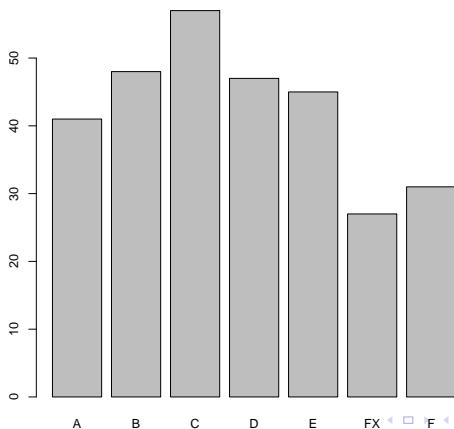
Description de la distribution d'une variable
Relation entre deux variables
variable qualitative
Variable quantitative

Diagramme à bandes (bar graph)

```
> barplot(T)
```

Exemple : note moyenne

```
> sy02$moyenne
  [1] 11.8 10.7  3.3 11.5  8.8  7.4  2.0 10.0  9.8 10.5  7.2 10.9  7.4  0.3 11.6 10.2  9.1
 [18] 14.3 14.6 13.6  4.6 14.5 12.4 10.7 16.2 13.3  8.2 11.8  6.3 10.5  9.0  6.7  9.8  9.2
 [35] 12.1 10.1  9.0 11.5 10.4 10.1 10.2  9.9 10.2  5.4 13.4  1.2 10.2 10.7 11.3 12.2 12.2
 [52]  5.3 11.1 11.9 17.2 10.2 12.5 12.4  7.4  9.1  7.5 12.7 16.9 12.3 10.8 13.1 12.6 14.1
 [69]  8.2  9.9 12.2 13.5 12.9  9.2 12.1 12.0 11.2  6.4  8.2 18.0 10.1 11.0 13.2 11.7 11.8
 [86] 13.2 13.4 12.3  7.2  7.1  9.3  8.8  9.7  7.8  8.7  9.6 11.0  9.5  9.2 12.7 10.5 10.3
[103] 15.6  9.7  9.6 13.8 11.6 10.4  0.8  7.2  8.1 10.3  6.7 15.5  9.8 11.8  9.6 11.9  9.0
[120]  7.9 10.0 11.6  2.9 10.1 11.1  5.9 11.5 15.4 10.9 12.6 12.8  9.3 13.8 14.3 17.3 19.4
[137] 16.9 16.1  8.2 10.9  6.4 15.4  4.6 10.8 14.4  9.6  8.2 14.1 11.4 12.1 12.3 15.1  2.4
[154] 17.1 15.0  9.9  9.0 13.7 14.1 13.1  8.6 11.9 12.1 13.5 10.9 11.0  6.0  6.8 14.0 10.4
[171]  7.9 11.0  7.5  8.7 10.3  9.8  7.2 18.6  8.4  9.6  9.4  9.3  7.8  7.8 11.4  9.0  9.4
[188] 13.4  7.4  9.6 10.6 15.2  7.8 18.0  6.4  7.2  9.2 11.2 12.6 11.6 12.1 12.0  9.8  9.8
[205]  5.3  4.6 10.0 10.4  9.0  7.8 11.3 11.9  7.8  9.2 10.6  8.5  8.6 12.0  8.2 11.0 13.4
[222] 18.4 11.6  5.8 15.4  8.6 12.0  8.6 10.0  8.7  9.5  4.7 10.7  7.9 16.4 11.9  8.8 11.7
[239]  4.8  0.4 12.6  3.4 10.6 10.9  5.8  9.6 16.7 10.9 12.0 11.7 10.3 17.0  7.5  7.4 10.2
[256] 12.9  8.6 15.6  7.8  8.8 15.1 14.2 11.5  8.1 12.1  8.4 11.5  5.7  7.8 10.8 12.2 10.2
[273] 11.2 17.0  9.8 10.3 17.4  9.9  9.2 18.1  3.6  1.5  7.3 10.7  9.3 17.1 10.7 13.7  8.5
[290] 12.6 15.2 17.2  8.2 13.2  9.9  8.5
```

Diagramme de dispersion (stripchart, dotplot)
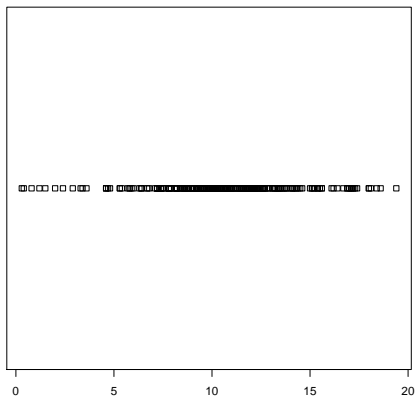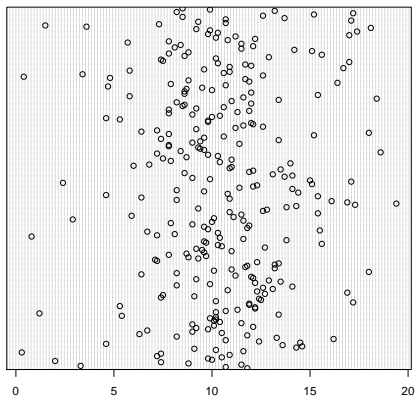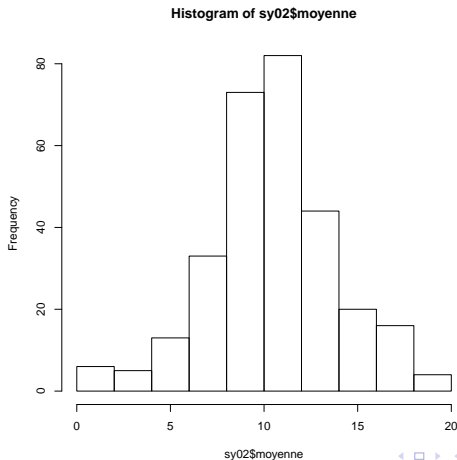
```
> stripchart(sy02$moyenne)
```

Diagramme de dispersion (stripchart, dotplot)

```
> dotchart(sy02$moyenne)
```

Histogramme

```
> hist(sy02$moyenne)
```


Histogram of sy02$moyenne

Histogramme (suite)

```
> brk=c(0,5,7,9,11,13,15,18,20)
> hist(sy02$moyenne,breaks=brk)
```

Histogram of sy02$moyenne

Description de la distribution d'une variable
Relation entre deux variables
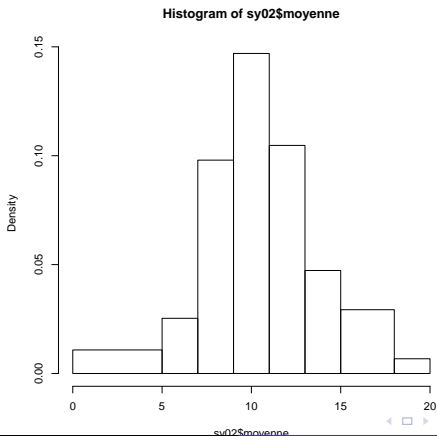variable qualitative
Variable quantitative

Diagramme en tige et feuilles (stem-and-leaf plot)

```
> stem(sy02$moyenne)
The decimal point is at the |

 0 | 348
 1 | 25
 2 | 049
 3 | 346
 4 | 66678
 5 | 3347889
 6 | 03444778
 7 | 1222234444455588888888999
 8 | 1122222224455566667778888
 9 | 00000011222222333344556666666677888888899999
10 | 0000111122222223333344455566677777788899999
11 | 00000122334455555666667778888999999
12 | 00000111111222233344456666677899
13 | 11222344445567788
14 | 0111233456
15 | 01122444566
16 | 124799
17 | 00112234
18 | 00146
19 | 4
```

Fonction de répartition empirique

```
> plot.ecdf(sy02$moyenne,breaks=brk)
```


ecdf(x)
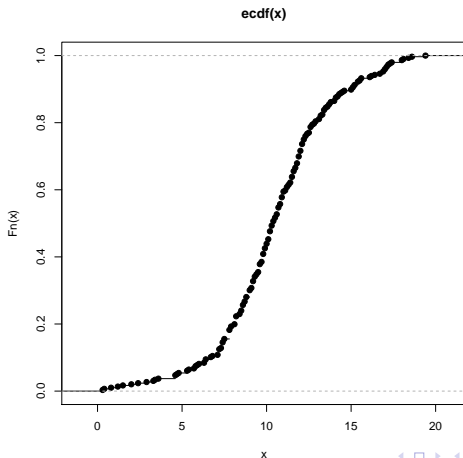
Résumés numériques

```
> summary(sy02$moyenne)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.30 8.60 10.40 10.52 12.22 19.40
```

Boîte à moustaches (boxplot)

```
> boxplot(sy02$moyenne)
```

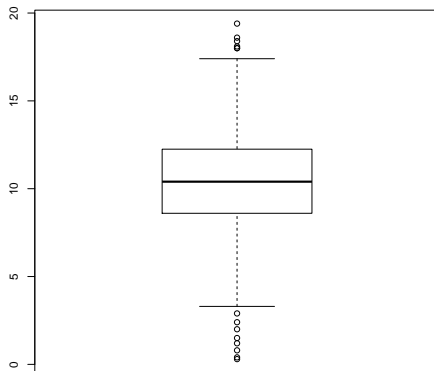Description de la distribution d'une variable
Relation entre deux variables

Deux variables quantitatives
Une variable qualitative et une variable quantitative
Deux variables qualitatives

Diagramme de corrélation

```
plot(sy02$median,sy02$final)
```

Description de la distribution d'une variable
Relation entre deux variables

Deux variables quantitatives
Une variable qualitative et une variable quantitative
Deux variables qualitatives
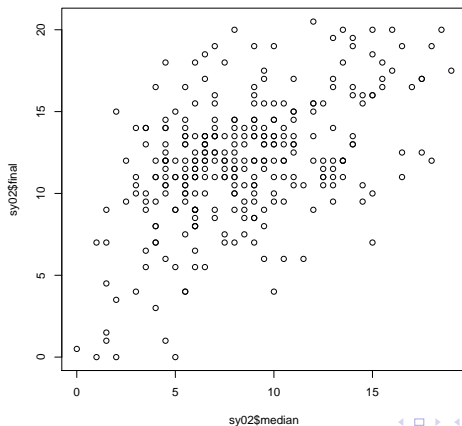
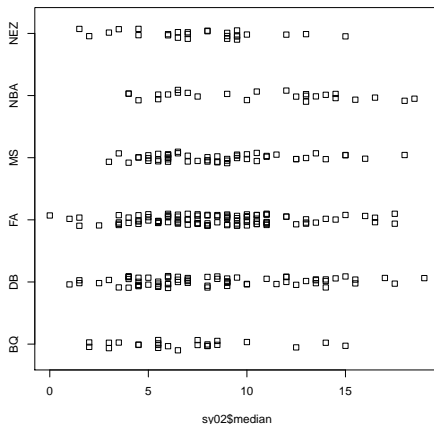# Coefficients de corrélation

```
> cor(sy02$median,sy02$final,method="pearson")
[1] 0.4933762
> cor(sy02$median,sy02$final,method="kendall")
[1] 0.31153
> cor(sy02$median,sy02$final,method="spearman")
[1] 0.4269235
```

Description de la distribution d'une variable
Relation entre deux variables

Deux variables quantitatives
Une variable qualitative et une variable quantitative
Deux variables qualitatives

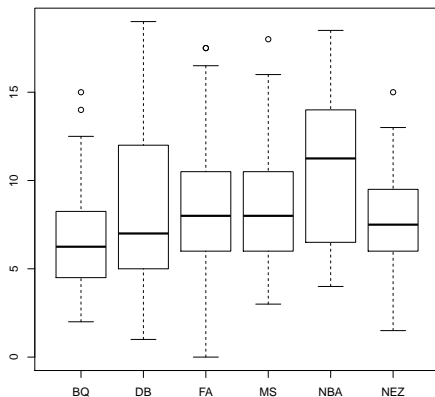# Diagrammes de dispersion parallèles

```
stripchart(sy02$median~sy02$correcteur2,method="jitter",jitter=0.1)
```

Description de la distribution d'une variable
Relation entre deux variables

Deux variables quantitatives
Une variable qualitative et une variable quantitative
Deux variables qualitatives

Boîtes à moustaches parallèles

```
boxplot(sy02$median~sy02$correcteur2)
```

Description de la distribution d'une variable
Relation entre deux variables

Deux variables quantitatives
Une variable qualitative et une variable quantitative
Deux variables qualitatives

Boîtes à moustaches parallèles

```
boxplot(sy02$median~ects)
```

Description de la distribution d'une variable
Relation entre deux variables

Deux variables quantitatives
Une variable qualitative et une variable quantitative
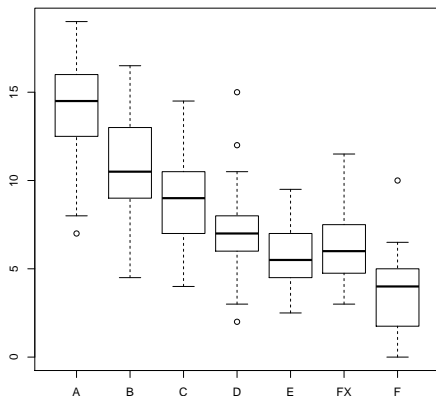Deux variables qualitatives

Tableau de contingence

```
> T=table(sy02$correcteur2,sy02$resultat)
> T

    A  B  C  D  E FX  F
BQ  2  1  5  9  2  2  3
DB 12  8 13 10 12  6  8
FA  8 23 20 17 14  8 10
MS  5  7  9  6 11  9  4
NBA 9  4  6  1  3  0  3
NEZ 5  5  4  4  3  2  3
```

Description de la distribution d'une variable
Relation entre deux variables

Deux variables quantitatives
Une variable qualitative et une variable quantitative
Deux variables qualitatives

Diagramme à bandes superposées

```
barplot(t(T),legend.text=levels(ects),args.legend=list(x = "topleft"))
```