

TD7

October 23, 2017

Luc Adjengue

Contents

1	Exercice 1	1
1.1	histogramme	3
1.2	BoxPlot	4
1.3	Diagramme quantile normal	5
1.4	Calculs de statistiques	6
2	Exercice 2	7
2.1	2-a) Calcul de la moyenne, écart type, 5eme et 95eme percentile	9
2.2	2-b) Vérification de la normalité	9
2.3	2-c) Calcul de la moyenne selon 'sexe'	10
3	Exercice 3	11
3.1	3-a)	13
3.2	3-b)	13
3.3	3-d)	15
3.4	3-e)	16

1 Exercice 1

```
In [1]: coton <- read.csv2("Coton.csv") # les données sont lues et placées dans une base de données
# si cette commande ne fonctionne pas, essayer celle-ci
# coton <- read.csv("Coton.csv", header = TRUE, sep = ";", dec = ".")
```

```
In [2]: coton # on peut visualiser l'ensemble des données en tapant le nom de la base de données
```

Pourcentage.de.coton

34.2
33.1
34.5
35.6
34.3
35.1
34.7
33.6
33.6
34.7
35.0
35.4
36.2
36.8
35.1
35.3
33.8
34.2
33.4
34.7
34.6
35.2
35.0
34.9
34.7
33.6
32.5
34.1
35.1
36.8

35.4
34.6
33.8
37.1
34.0
34.1
32.6
33.1
34.6
35.9
34.7
33.6
32.9
33.5
35.8
37.6
37.3
34.6
35.5
32.8
32.1

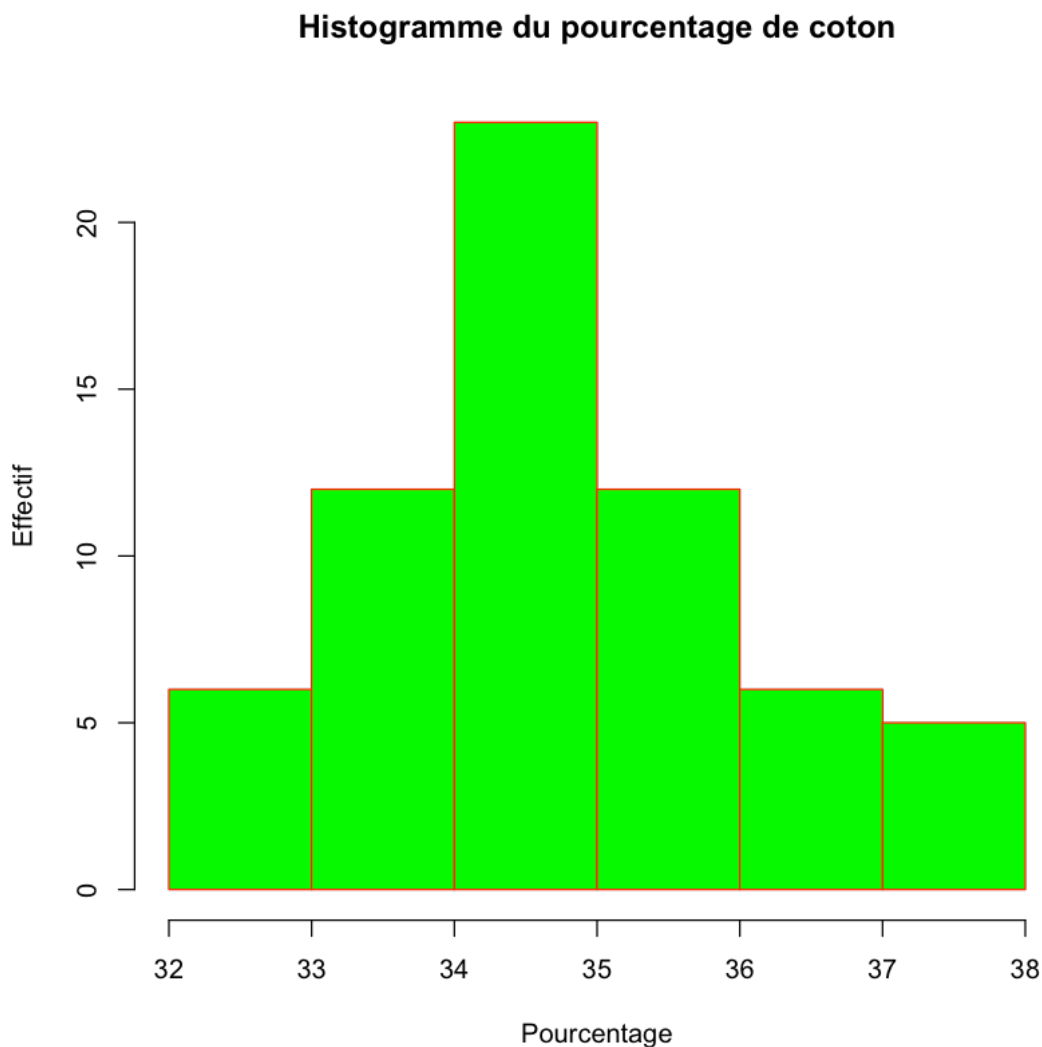
```
In [3]: x <- coton$Pourcentage.de.coton # on peut choisir un nom de variable plus court, e.g. ic
```

```
In [4]: x # on peut voir les données d'une variable en tapant le nom de la variable
```

```
1. 34.2 2. 33.1 3. 34.5 4. 35.6 5. 34.3 6. 35.1 7. 34.7 8. 33.6 9. 33.6 10. 34.7 11. 35 12. 35.4 13. 36.2
14. 36.8 15. 35.1 16. 35.3 17. 33.8 18. 34.2 19. 33.4 20. 34.7 21. 34.6 22. 35.2 23. 35 24. 34.9 25. 34.7
26. 33.6 27. 32.5 28. 34.1 29. 35.1 30. 36.8 31. 37.9 32. 36.4 33. 37.8 34. 36.6 35. 35.4 36. 34.6 37. 33.8
38. 37.1 39. 34 40. 34.1 41. 32.6 42. 33.1 43. 34.6 44. 35.9 45. 34.7 46. 33.6 47. 32.9 48. 33.5 49. 35.8
50. 37.6 51. 37.3 52. 34.6 53. 35.5 54. 32.8 55. 32.1 56. 34.5 57. 34.6 58. 33.6 59. 34.1 60. 34.7 61. 35.7
62. 36.8 63. 34.3 64. 32.7
```

1.1 histogramme

```
In [5]: hist(x,col="green",main="Histogramme du pourcentage de coton",
            border="red", xlab="Pourcentage",ylab="Effectif")
```



R est programmé pour produire un histogramme avec un nombre de classes optimal. Le tableau de disribution des fréquences ne présente donc pas un grand intérêt. Si on veut un tableau des fréquences, il faut créer des limites de classes

```
In [6]: breaks <- seq(32,38,by=1) # by donne la longueur de l'intervalle (essayer différentes val
```

```
In [7]: breaks
```

```
1. 32 2. 33 3. 34 4. 35 5. 36 6. 37 7. 38
```

```
In [8]: classx <- factor(cut(x, breaks))
```

```
In [9]: xout <- as.data.frame(table(classx))
      xout
```

classx	Freq
(32,33]	6
(33,34]	12
(34,35]	23
(35,36]	12
(36,37]	6
(37,38]	5

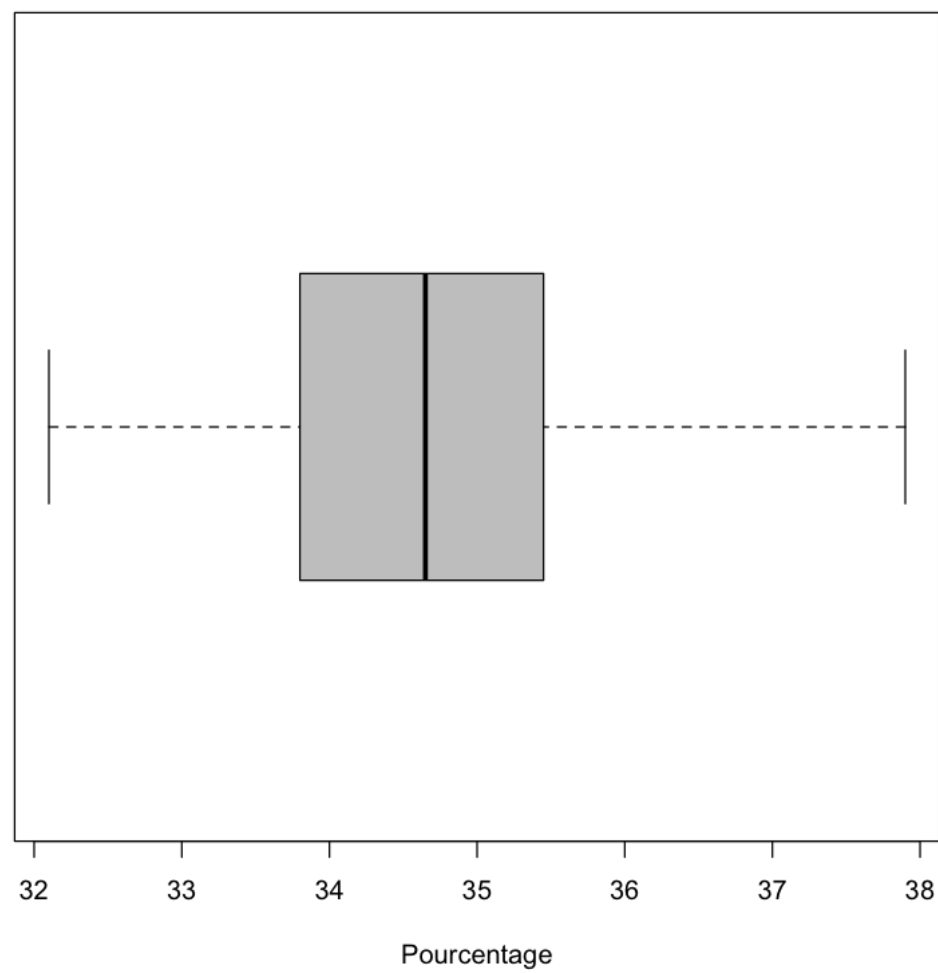
On peut ajouter des effectifs cumulés et des fréquences relatives

```
In [10]: xout <- transform(xout, Freq_Cum = cumsum(Freq), freq_relative = prop.table(Freq))
      xout
```

classx	Freq	Freq_Cum	freq_relative
(32,33]	6	6	0.093750
(33,34]	12	18	0.187500
(34,35]	23	41	0.359375
(35,36]	12	53	0.187500
(36,37]	6	59	0.093750
(37,38]	5	64	0.078125

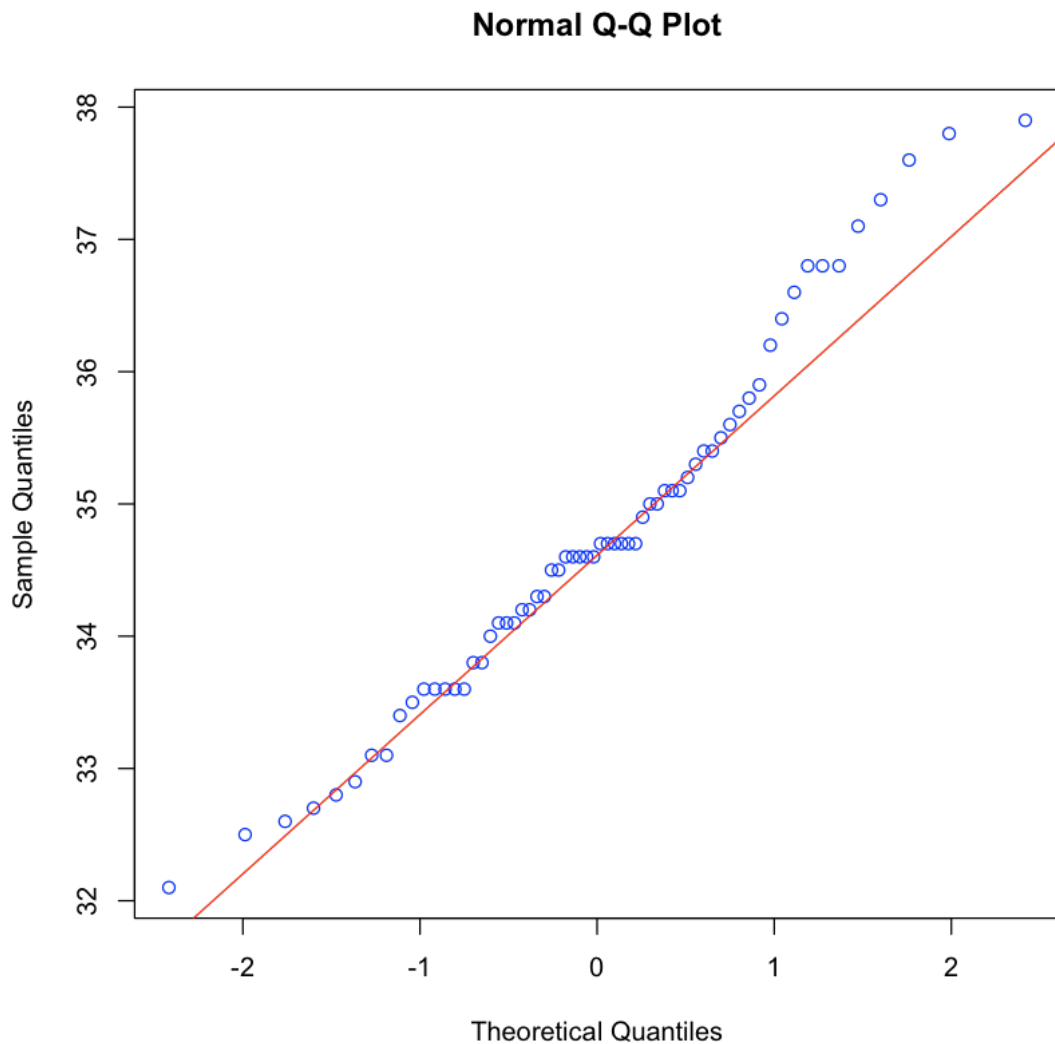
1.2 BoxPlot

```
In [11]: boxplot(x, horizontal=T, col="grey", xlab="Pourcentage")
```



1.3 Diagramme quantile normal

```
In [12]: qqnorm(x,col="blue")  
         qqline(x,col="red")
```



1.4 Calculs de statistiques

```
In [13]: m=mean(x)
          md=median(x)
          v=var(x)
          s=sd(x)
          cv=s/m
          cat('moyenne =',m,'médiane =',md,', écart-type = ', s,', variance = ', v ,'\n',
              'coefficient de variation = ', cv )
```

```
moyenne = 34.76719 médiane = 34.65 , écart-type =  1.352018 , variance =  1.827954
coefficient de variation =  0.03888777
```

La commande 'summary()' donne certaines statistiques : le minimum, le 1er quartile, la médiane, la moyenne, le 3eme quartile et le maximum.

```
In [14]: summary(x)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
32.10   33.80   34.65   34.77   35.42   37.90
```

Les percentiles de tout ordre peuvent être calculés avec la commande quantile(x, ordres)

```
In [15]: quantile(x, .5) # la médiane
```

```
50\%: 34.65
```

```
In [16]: quantile(x, c(.25,.5,.75)) # pour les 3 quartiles
```

```
25\%          33.8 50\%          34.65 75\%          35.425
L'écart inter quartile (commande IQR)
```

```
In [17]: IQR(x)
```

```
1.625
```

2 Exercice 2

```
In [18]: mem <-read.csv2("Memoire.csv") # les données sont lues du fichier et placées dans la base de données
# si la commande ne fonctionne pas, essayer celle-ci
# mem <- read.csv("Memoire.csv", header = TRUE, sep = ";",dec = ".")
```

```
In [19]: str(mem) # cette commande permet d'avoir un aperçu du nombre d'observations, du nombre
# des variables d'une base de données, e.g. ici, 'mem',
# avec leur type, 'factor' pour catégoriel, 'int' pour entier, 'num' pour continue,
```

```
'data.frame':      48 obs. of  8 variables:
 $ groupe   : Factor w/ 2 levels "control","experimental": 1 1 1 2 1 2 2 2 2 1 ...
 $ sexe     : Factor w/ 2 levels "femme","homme": 1 2 2 1 2 1 1 2 2 2 ...
 $ temps    : Factor w/ 3 levels "apres-1","apres-2",...: 2 2 1 2 1 1 2 1 3 3 ...
 $ pay_     : Factor w/ 2 levels "non","oui": 2 2 1 2 2 1 2 2 2 1 ...
 $ stress   : int   14 10 11 11 13 14 10 14 1 5 ...
 $ correct.1: int   12 15 4 8 5 5 6 11 7 10 ...
 $ correct.2: int    2 2 7 5 6 1 6 4 6 8 ...
 $ correct.3: int    5 7 4 6 2 1 9 6 0 1 ...
```

```
In [20]: mem # on peut voir les données elles-mêmes avec le nom de la base de données
```

groupe	sexe	temps	pay_	stress	correct.1	correct.2	correct.3
control	femme	apres-2	oui	14	12	2	5
control	homme	apres-2	oui	10	15	2	7
control	homme	apres-1	non	11	4	7	4
experimental	femme	apres-2	oui	11	8	5	6
control	homme	apres-1	oui	13	5	6	2
experimental	femme	apres-1	non	14	5	1	1
experimental	femme	apres-2	oui	10	6	6	9
experimental	homme	apres-1	oui	14	11	4	6
experimental	homme	avant	oui	1	7	6	0
control	homme	avant	non	5	10	8	1
control	homme	apres-2	oui	8	12	7	6
experimental	homme	avant	oui	2	11	7	3
control	femme	apres-1	oui	12	10	9	3
control	homme	avant	non	5	11	7	5
control	homme	apres-1	non	13	5	5	5
control	femme	apres-2	non	7	8	7	4
experimental	homme	avant	non	2	12	4	6
experimental	homme	apres-2	oui	14	9	6	5
experimental	homme	apres-2	non	14	19	5	4
control	femme	avant	non	4	16	5	7
experimental	femme	apres-1	oui	13	7	6	0
control	femme	avant	non	7	14	4	8
control	homme	apres-2	non	5	4	5	4
control	femme	avant	oui	6	12	2	4
control	femme	apres-2	oui	15	7	1	6
experimental	homme	avant	non	2	3	3	7
experimental	femme	avant	non	13	6	5	5
control	femme	apres-1	non	13	16	1	1
control	femme	avant	oui	5	13	4	5
control	femme	apres-1	oui	11	9	8	2
experimental	femme	avant	non	15	8	2	2
control	homme	apres-2	non	5	8	8	5
experimental	homme	apres-1	oui	3	15	3	3
control	femme	apres-2	non	6	6	8	5
experimental	femme	apres-1	non	14	6	3	2
experimental	femme	apres-1	oui	15	6	6	8
experimental	homme	apres-1	non	13	8	2	7
experimental	homme	apres-2	non	14	11	6	7
experimental	femme	apres-2	non	11	5	4	7
experimental	femme	apres-2	non	14	7	6	8
control	homme	avant	oui	7	9	9	5
experimental	femme	avant	oui	14	8	4	4
control	homme	avant	oui	10	8	8	6
control	homme	apres-1	oui	11	6	8	1
experimental	homme	apres-2	oui	14	10	4	8
experimental	femme	apres-1	oui	17	7	5	7
control	femme	apres-1	non	13	14	8	2
experimental	homme	apres-1	non	3	15	1	7


```
In [21]: y <- mem$stress # on peut utiliser un nom de variable plus court
```

```
In [22]: y
```

```
1. 14 2. 10 3. 11 4. 11 5. 13 6. 14 7. 10 8. 14 9. 1 10. 5 11. 8 12. 2 13. 12 14. 5 15. 13 16. 7 17. 2 18. 14
19. 14 20. 4 21. 13 22. 7 23. 5 24. 6 25. 15 26. 2 27. 13 28. 13 29. 5 30. 11 31. 15 32. 5 33. 3 34. 6 35. 14
36. 15 37. 13 38. 14 39. 11 40. 14 41. 7 42. 14 43. 10 44. 11 45. 14 46. 17 47. 13 48. 3
```

2.1 2-a) Calcul de la moyenne, écart type, 5eme et 95eme percentile

```
In [23]: summary(y)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	5.750	11.000	9.854	14.000	17.000

```
In [24]: mean(y)
```

```
9.85416666666667
```

```
In [25]: sd(y)
```

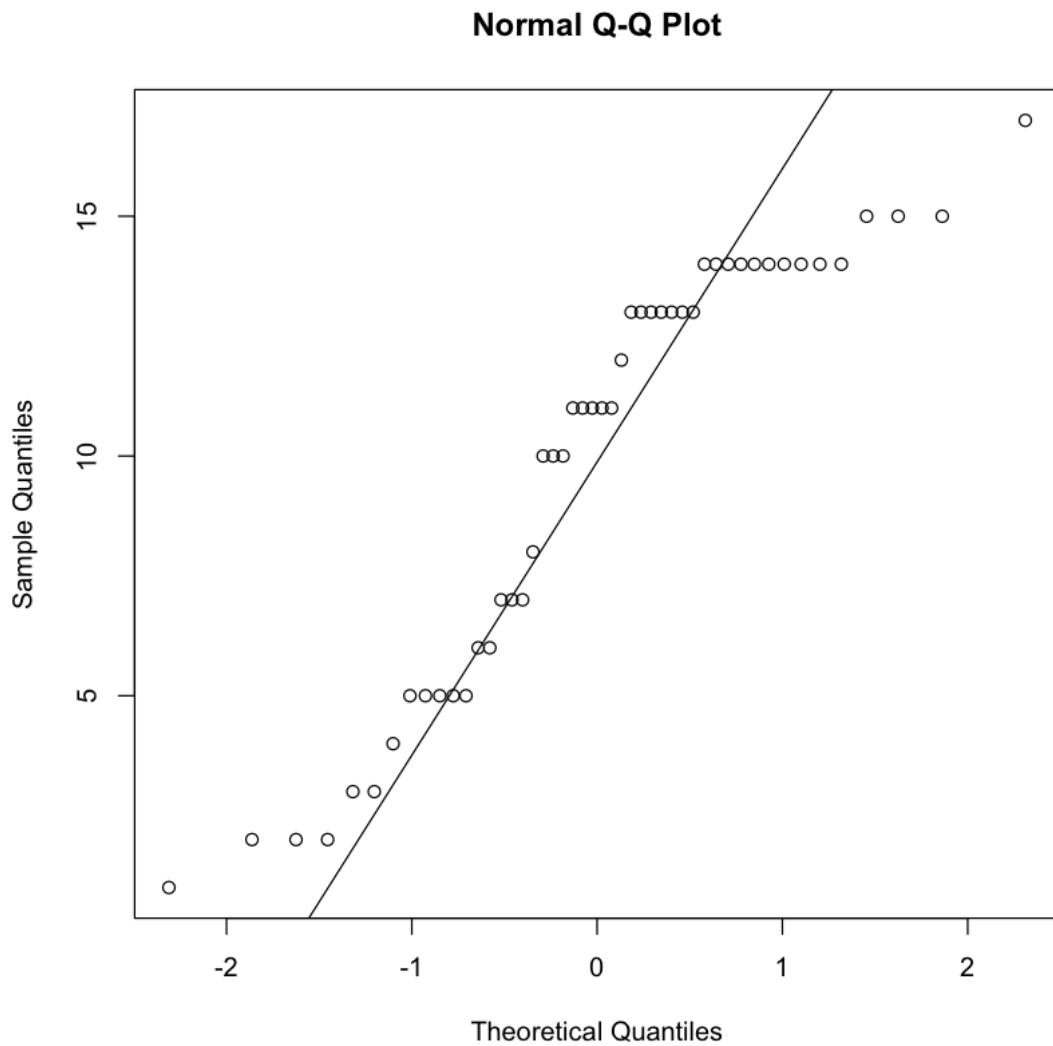
```
4.49581168367131
```

```
In [26]: quantile(y, c(.05, .95))
```

5\%	2 95\%	15
-----	--------	----

2.2 2-b) Vérification de la normalité

```
In [27]: qqnorm(y)
          qqline(y)
```



L'hypothèse d'une distribution normale n'est pas plausible

2.3 2-c) Calcul de la moyenne selon 'sexe'

```
In [28]: by(mem$stress, mem$sexe, function(x) mean(x)) # calcul de la moyenne selon 'sexe'
```

```
mem$sexe: femme
```

```
[1] 11.41667
```

```
-----
```

```
mem$sexe: homme
```

```
[1] 8.291667
```

```
In [29]: by(y, mem$sexe, function(x) {mean(x)}) # autre façon de calculer la moyenne selon 'sexe'
```

```
mem$sexe: femme  
[1] 11.41667
```

```
-----  
mem$sexe: homme  
[1] 8.291667
```

```
In [30]: by(y, mem$sexe, function(x) var(x)) # calcul de la variance selon 'sexe'
```

```
mem$sexe: femme  
[1] 13.47101
```

```
-----  
mem$sexe: homme  
[1] 22.73732
```

3 Exercice 3

```
In [31]: web <- read.csv2("Sites_web.csv")  
          # si la commande ne fonctionne pas, essayer celle-ci  
          # web <- read.csv(Sites_web.csv", header = TRUE, sep = ";", dec = ".")
```

```
In [32]: str(web)
```

```
'data.frame':      73 obs. of  8 variables:  
 $ no                : int  1 2 3 4 5 6 7 8 9 10 ...  
 $ nbre.de.sites.livres : int  1 2 7 2 1 10 10 1 1 6 ...  
 $ carnet.de.commandes : int 12 18 26 28 36 45 36 18 25 28 ...  
 $ no.equipe         : int  1 1 1 1 1 1 1 2 2 2 ...  
 $ experience.de.l.equipe: int  3 6 9 12 15 18 21 3 6 9 ...  
 $ processus.modifie  : int  0 0 0 0 0 1 1 0 0 0 ...  
 $ annee              : int 2001 2001 2001 2001 2002 2002 2002 2001 2001 2001 ...  
 $ trimestre          : int  1 2 3 4 1 2 3 1 2 3 ...
```

```
In [33]: web
```

no	nbre.de.sites.livres	carnet.de.commandes	no.equipe	experience.de.l.equipe	processus.modifie
1	1	12	1	3	0
2	2	18	1	6	0
3	7	26	1	9	0
4	2	28	1	12	0
5	1	36	1	15	0
6	10	45	1	18	1
7	10	36	1	21	1
8	1	18	2	3	0
9	1	25	2	6	0
10	6	28	2	9	0
11	5	28	2	12	0
12	11	38	2	15	0
13	15	38	2	18	1
14	13	34	2	21	1
15	3	18	3	3	0
16	3	23	3	6	0
17	13	29	3	9	0
18	6	24	3	12	0
19	2	32	3	15	0
20	11	41	3	18	1
21	3	33	3	21	1
22	7	21	4	3	0
23	12	22	4	6	0
24	7	25	4	9	0
25	1	27	4	12	0
26	10	37	4	15	0
27	12	38	4	18	1
28	8	34	4	21	1
29	1	19	5	3	0
30	8	25	5	6	0
44	11	34	7	7	0
45	16	31	7	10	1
46	27	26	7	13	1
47	5	18	8	5	0
48	6	35	8	8	0
49	22	38	8	11	1
50	30	29	8	14	1
51	6	12	9	3	0
52	13	23	9	6	0
53	6	21	9	9	0
54	5	33	9	12	0
55	19	37	9	15	1
56	22	27	9	18	1
57	7	18	10	4	0
58	9	20	10	7	0
59	8	33	10	10	0
60	7	33	10	13	1
61	6	33	10	16	1
62	3	23	11	5	0
63	13	34	11	8	1
64	21	29	11	11	1

```
In [34]: # ceci n'est pas nécessaire :
# attribuons de noms plus courts aux variables de la base de données
v1 <- web$no
v2 <- web$nbre.de.sites.livres
v3 <- web$carnet.de.commandes
v4 <- web$no.equipe
v5 <- web$experience.de.l.equipe
v6 <- web$processus.modifie
v7 <- web$annee
v8 <- web$trimestre
```

3.1 3-a)

```
In [35]: summary(v2)
s2<-sd(v2)
cv2<-sd(v2)/mean(v2)
cat('moyenne =',mean(v2),', médiane =',median(v2),', écart-type = ', s2,
    ', coefficient de variation = ', cv2 )
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	3.000	7.000	9.041	13.000	30.000

moyenne = 9.041096 , médiane = 7 , écart-type = 7.083703 , coefficient de variation = 0.783500

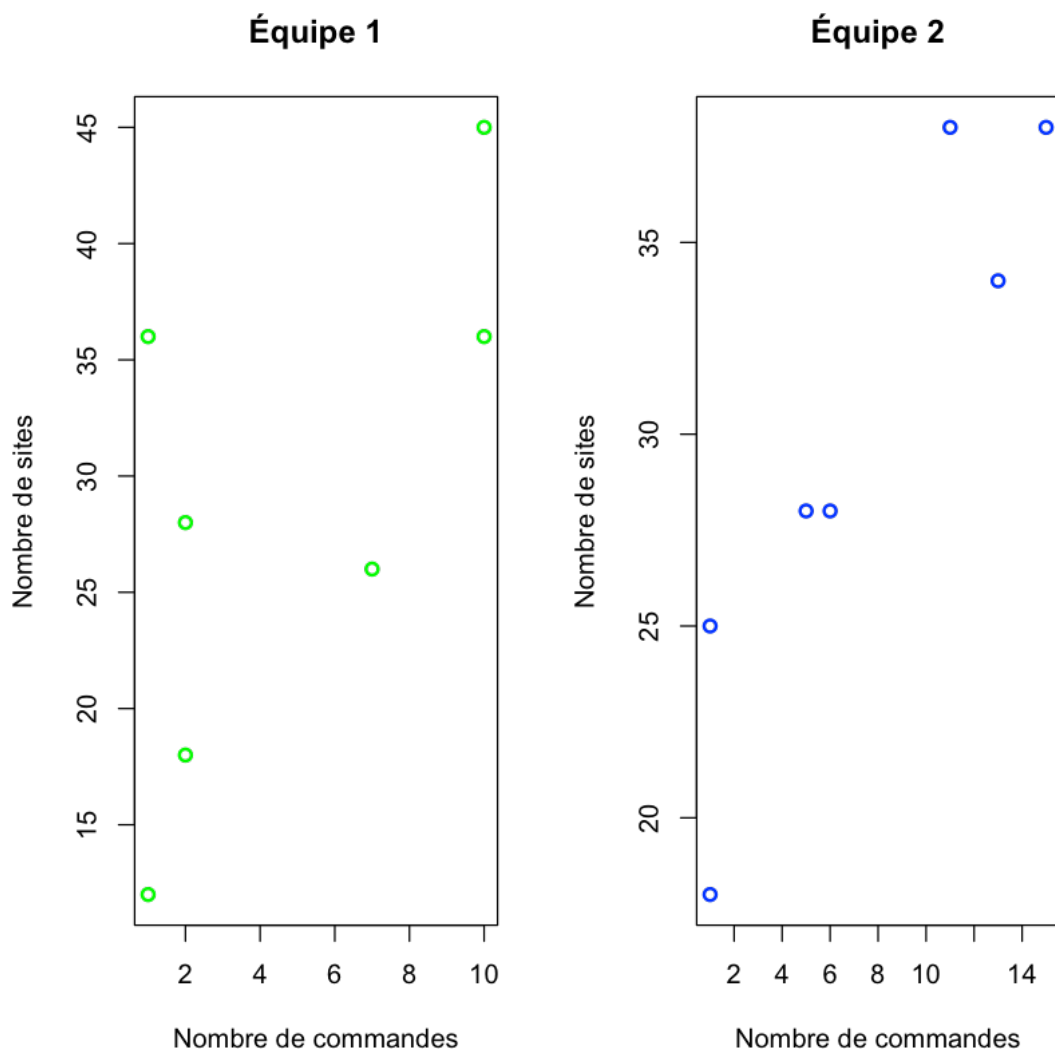
```
In [36]: summary(v3)
s3<-sd(v3)
cv3<-sd(v3)/mean(v3)
cat('moyenne =',mean(v3),', médiane =',median(v3),', écart-type = ', s3,
    ', coefficient de variation = ', cv3 )
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.00	23.00	28.00	27.82	34.00	45.00

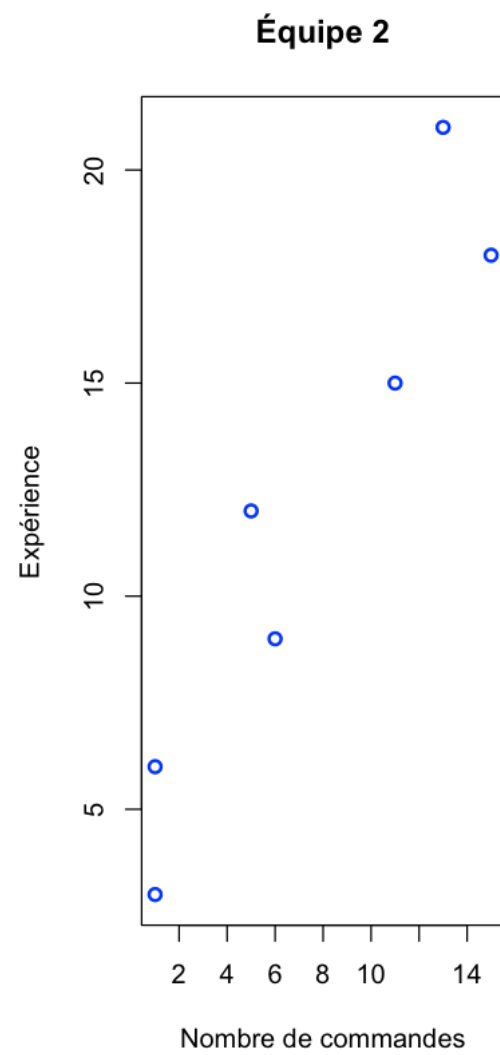
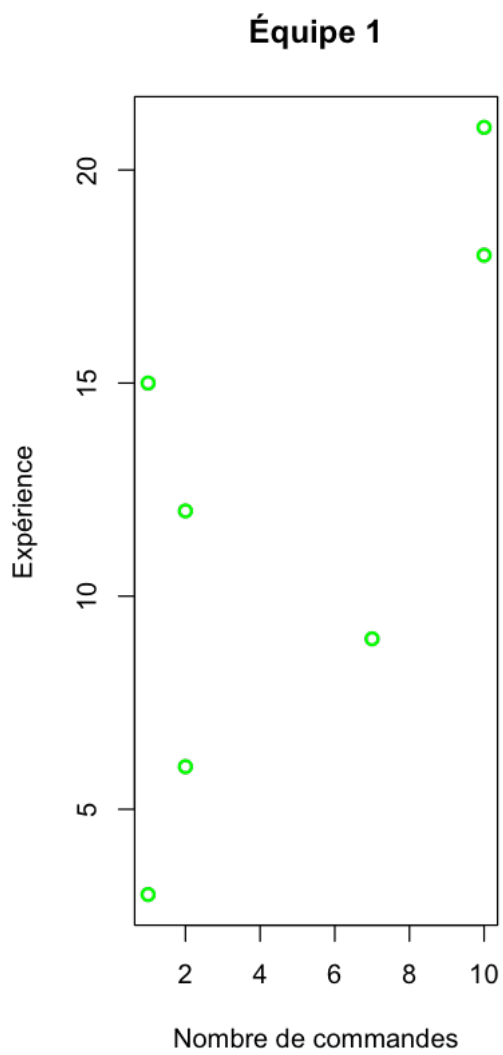
moyenne = 27.82192 , médiane = 28 , écart-type = 7.976254 , coefficient de variation = 0.28668

3.2 3-b)

```
In [37]: layout(matrix(1:2,1,2)) # ceci permet de diviser la sortie graphique en 2
plot(v2[v4==1],v3[v4==1],main="Équipe 1",col="green",lwd=2,
     xlab="Nombre de commandes",ylab="Nombre de sites")
plot(v2[v4==2],v3[v4==2],main="Équipe 2",col="blue", lwd=2,
     xlab="Nombre de commandes",ylab="Nombre de sites")
```

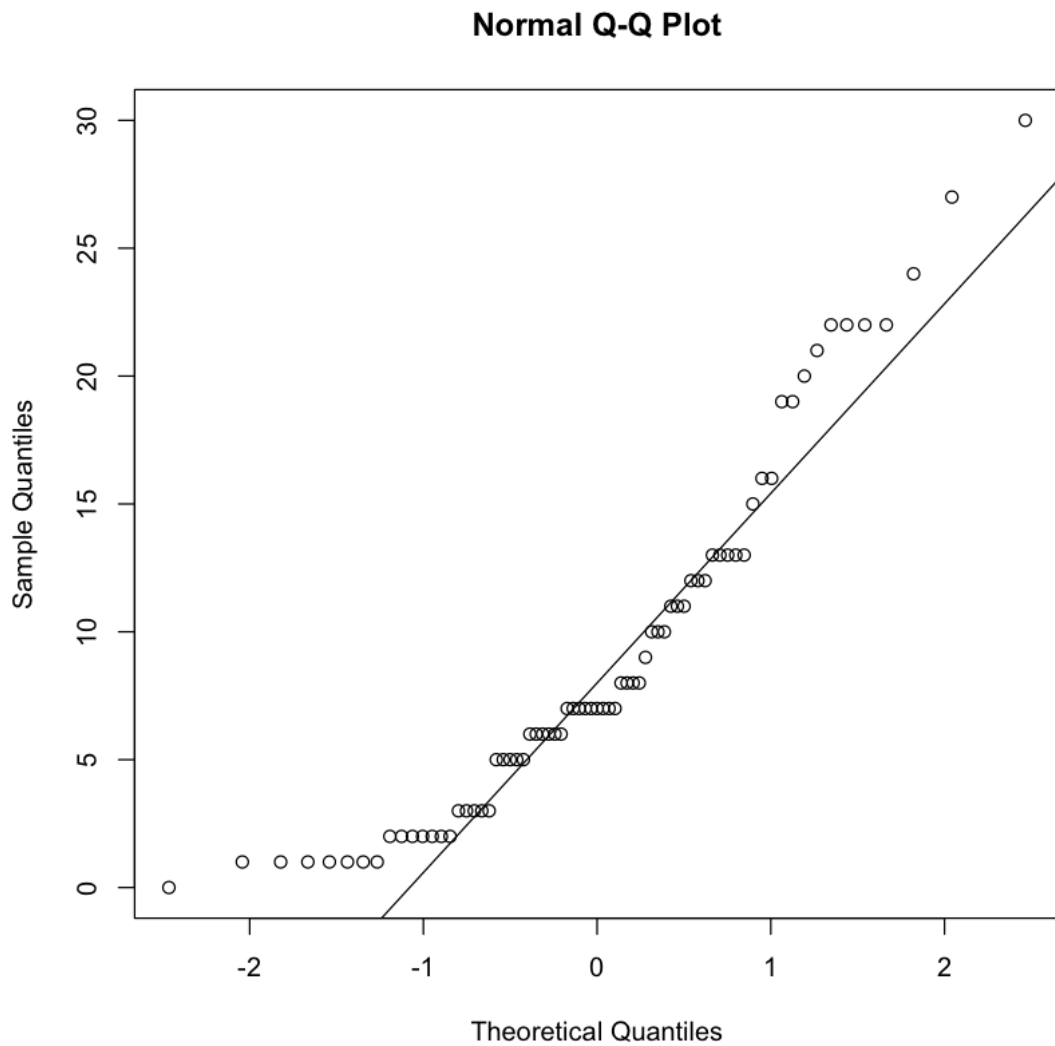


```
In [38]: layout(matrix(1:2,1,2)) # ceci permet de diviser la sortie graphique en 2
plot(v2[v4==1],v5[v4==1],main="Équipe 1",col="green",lwd=2,
      xlab="Nombre de commandes",ylab="Expérience")
plot(v2[v4==2],v5[v4==2],main="Équipe 2",col="blue", lwd=2,
      xlab="Nombre de commandes",ylab="Expérience")
```



3.3 3-d)

```
In [39]: qqnorm(v2)  
         qqline(v2)
```



Non, la normalité n'est pas plausible.

3.4 3-e)

Avant

```
In [40]: summary(v2[v6==0])
          sd(v2[v6==0])
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	2.000	5.000	5.447	7.000	16.000

3.83807305953922

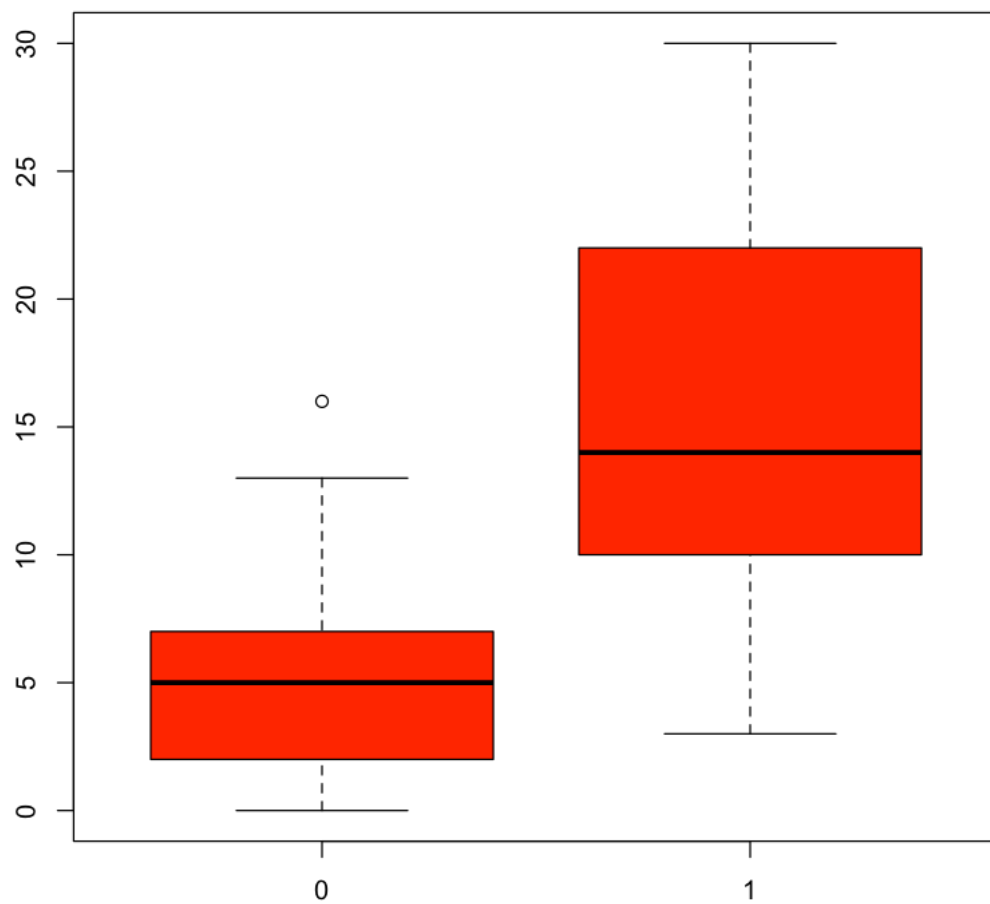
Après


```
In [41]: summary(v2[v6==1])  
sd(v2[v6==1])
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.00	10.25	14.00	15.54	21.75	30.00

7.01558704161395

```
In [42]: boxplot(v2~v6,col="red")
```



```
In [43]: boxplot(v3~v6,col="yellow")
```

