

SI-M1-TD5

Guillaume Metzler

11/23/2021

Exercice 1

On va commencer par recopier notre tableau de données

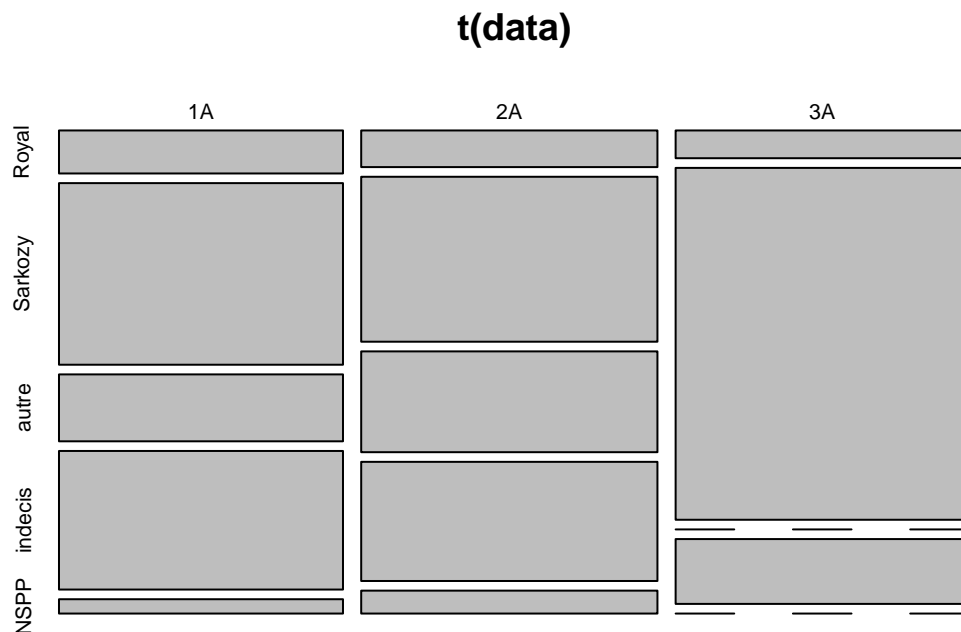
```
data=rbind(c(9,8,6),c(38,36,76),c(14,22,0),c(29,26,14),c(3,5,0))
colnames(data)=c("1A","2A","3A")
rownames(data)=c("Royal","Sarkozy","autre","indecis","NSPP")
data=as.table(data)
```

On souhaite savoir si le profil du candidat pour lequel les étudiants votent à tendance à évoluer au cours des données. On va donc formuler les hypothèses suivantes :

- H_0 : les deux facteurs sont indépendants
- H_1 : les deux facteurs sont liés

Comme nous étudions deux variables qualitatives, nous allons donc procéder à un test du Khi-deux. Regardons déjà cela graphiquement.

```
plot(t(data))
```



On remarque que la proportion de votant pour un candidat évolue bien avec les années. Regardons maintenant si toutes les conditions sont réunies. ici ce n'est pas le cas car certains effectifs sont inférieurs à 5. Ainsi pour faire notre test du Khi-deux, nous devons ajouter une petite option afin d'améliorer la précision de notre test.

```
chisq.test(data, simulate.p.value = TRUE)
```

```
##  
## Pearson's Chi-squared test with simulated p-value (based on 2000  
## replicates)  
##  
## data: data  
## X-squared = 51.383, df = NA, p-value = 0.0004998
```

On peut donc bien affirmer que les opinions politiques ou intentions de votes évoluent avec le temps.

Exercice 2

On charge le jeu de données

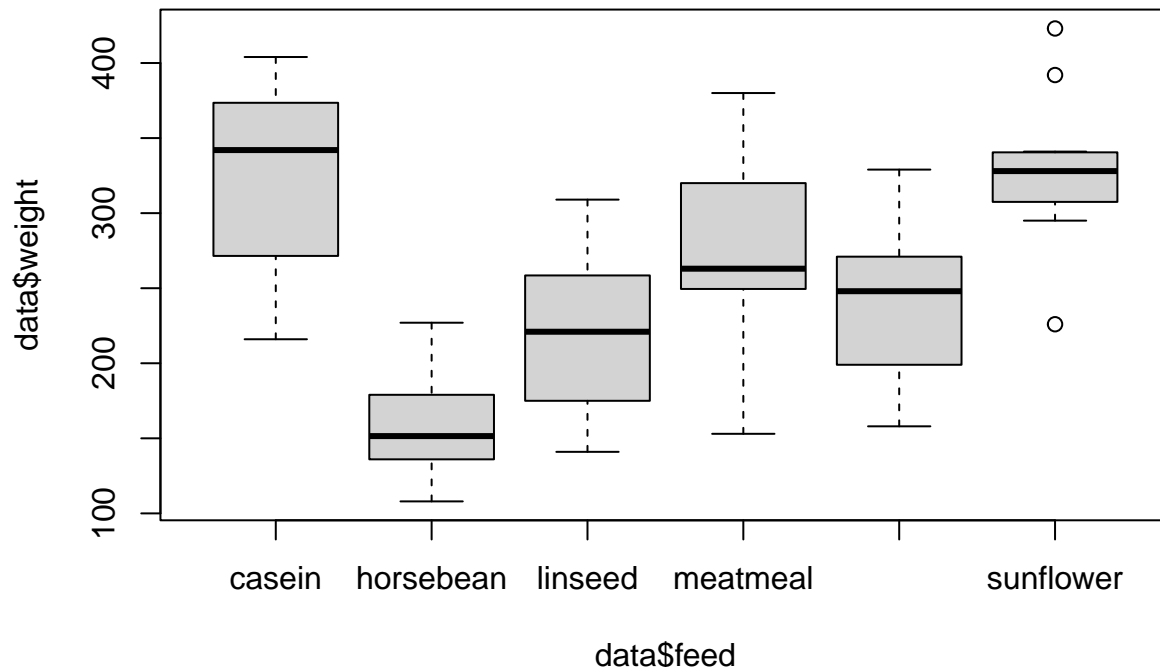
```
data <- chickwts
```

On souhaite savoir si l'alimentation du poulet a un impact sur sa masse, on va donc formuler le test d'hypothèses suivant :

- H_0 : l'alimentation n'a pas d'impact sur la masse des poulets, i.e. les moyennes sont identiques entre les différents groupes.
- H_1 l'alimentation a un bien un impact ! Il existe donc un groupe pour lequel la moyenne est différente de celle des autres groupes.

Représentons cela graphiquement

```
boxplot(data$weight~data$feed)
```



Les moyennes sont bien différentes d'un groupe à l'autre. Etant donné le nombre de modalités de la variable qualitative, on va donc faire une Analyse de Variance pour effectuer ce test. Commençons par regarder les hypothèses de normalité et d'homoscédasticité.

```
table(data$feed)
```

```
##  
##   casein horsebean  linseed  meatmeal  soybean sunflower  
##     12         10      12        11      14         12
```

Vu les effectifs des différents groupes, nous n'aurons pas d'autre choix que de faire un test de Kruskal Wallis. En effet, les échantillons des différents groupes sont trop petits pour que le test de Shapiro soit significatif.

```
kruskal.test(data$weight~data$feed)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data:  data$weight by data$feed  
## Kruskal-Wallis chi-squared = 37.343, df = 5, p-value = 5.113e-07
```

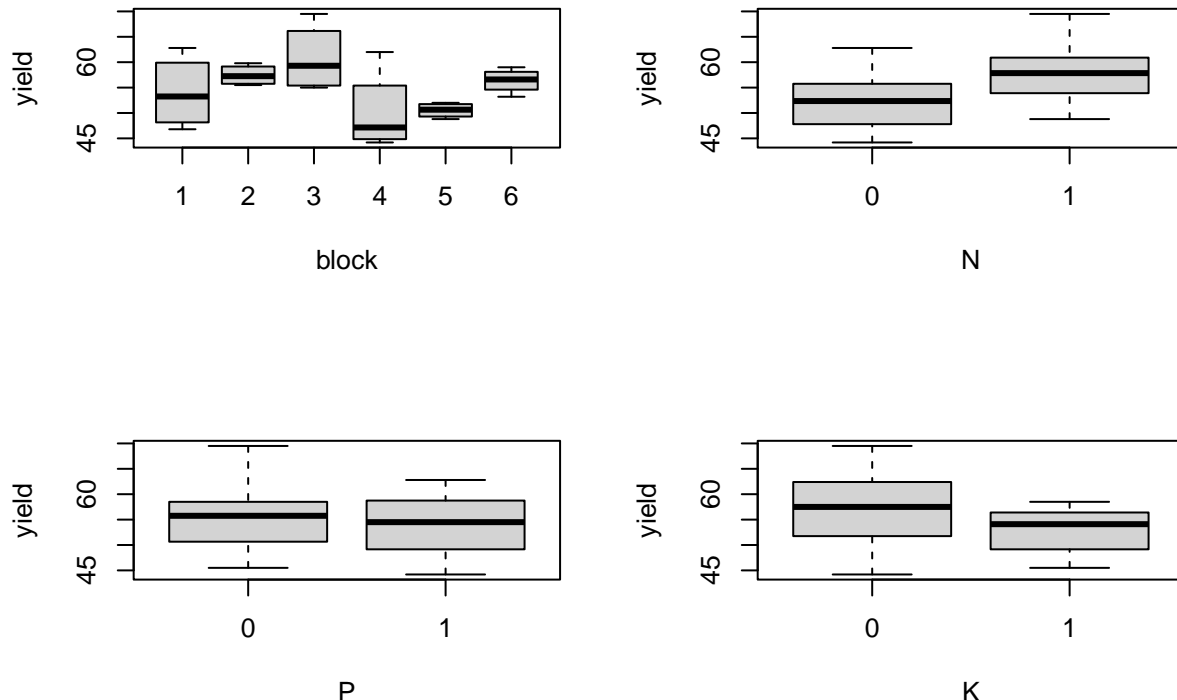
On remarque que l'alimentation a donc bien un impact sur la masse du poulet.

Exercice 3

On souhaite regarder l'impact de la présence/absence d'éléments dans la terre sur rendement d'une parcelle. Regardons nos données d'un peu plus près dans un premier temps.

```
data <- npk
```

```
par(mfrow=c(2,2))  
plot(yield ~ block+N+P+K, data=data)
```



On remarque que certains éléments présents dans les parcelles semblent avoir un impact sur le rendement de cette dernière. On va supposer, dans le cadre cet exercice que les rendements affichés sont des rendements moyens obtenus sur un ensemble de plusieurs parcelles.

On va donc pouvoir effectuer une Analyse de Variance pour faire notre étude. Il faudrait bien sûr vérifier l'homogénéité des variances pour les différents groupes étudiés (pas nécessairement les croisements !)

Pour le facteur block

```
bartlett.test(data$yield~data$block)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: data$yield by data$block  
## Bartlett's K-squared = 11.508, df = 5, p-value = 0.04219
```

Pour le facteur N

```
bartlett.test(data$yield~data$N)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: data$yield by data$N  
## Bartlett's K-squared = 0.057652, df = 1, p-value = 0.8102
```

Pour le facteur P

```
bartlett.test(data$yield~data$P)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: data$yield by data$P  
## Bartlett's K-squared = 0.1555, df = 1, p-value = 0.6933
```

Pour le facteur K

```
bartlett.test(data$yield~data$K)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: data$yield by data$K  
## Bartlett's K-squared = 3.0059, df = 1, p-value = 0.08296
```

Excepté pour le facteur “block”, toutes les variances sont homogènes (et encore c’est limite pour le facteur block) on va donc pouvoir faire notre Analyse de Variance.

On va cependant la faire d’une façon un petit peu particulière, on voit l’apparition d’un facteur block dans notre Analyse de Variance, on va donc l’exclure de cette analyse en extrayant la part de variance expliquée par ce facteur, puis en incluant les autres facteurs de façon “traditionnelle”

```
res <- aov(yield~block+N*P*K, data = data)  
summary(res)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)  
## block      5  343.3   68.66   4.447 0.01594 *  
## N          1  189.3  189.28  12.259 0.00437 **  
## P          1    8.4    8.40   0.544 0.47490  
## K          1   95.2   95.20   6.166 0.02880 *  
## N:P        1   21.3   21.28   1.378 0.26317  
## N:K        1   33.1   33.13   2.146 0.16865  
## P:K        1    0.5    0.48   0.031 0.86275  
## Residuals 12  185.3   15.44
```

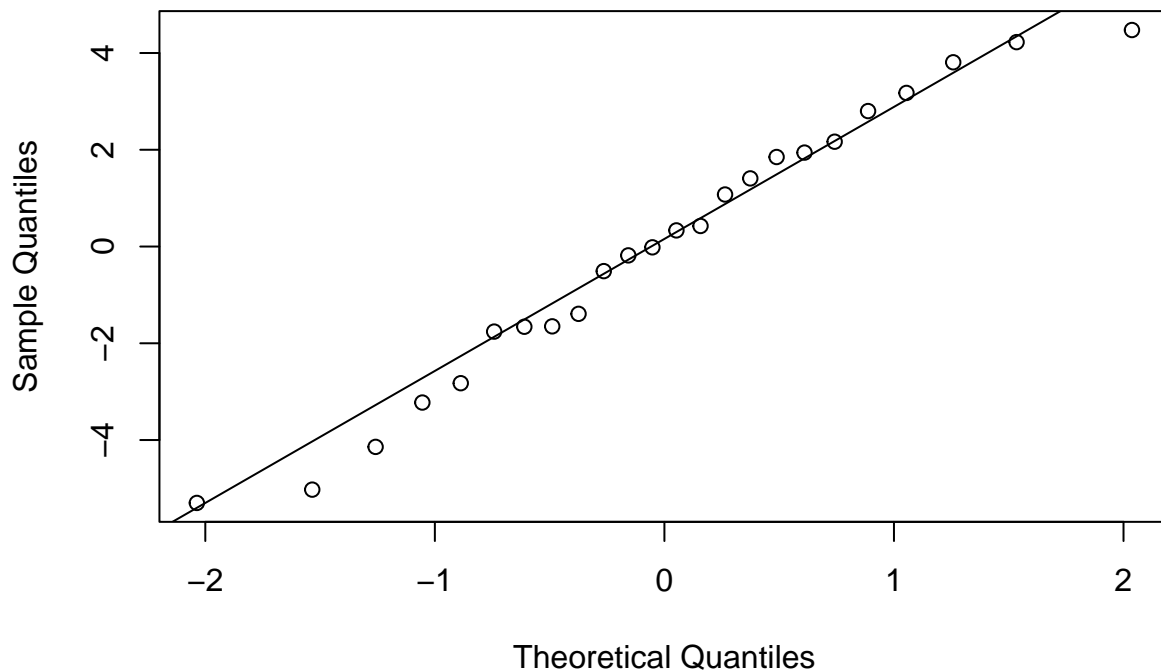
```
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Les résultats de ce test suggèrent que seule la présence de Potassium (K) et d'Azote (N) ont un impact sur rendement des différentes parcelles.

On va maintenant vérifier la validité du modèle en regardant la normalité des résidus.

```
qqnorm(res$residuals)  
qqline(res$residuals)
```

Normal Q-Q Plot



```
shapiro.test(res$residuals)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  res$residuals  
## W = 0.96875, p-value = 0.6363
```

Nos résidus sont bien normalement distribués. On peut donc valider les résultats de notre ANOVA.

Exercice 4

On souhaite analyser les effets de deux médicaments afin de savoir si ces derniers ont significativement augmenté le temps de sommeil des individus. Pour cela on mesure l'augmentation du temps de sommeil sur 10 individus suite à la prise de l'un ou l'autre des médicaments.

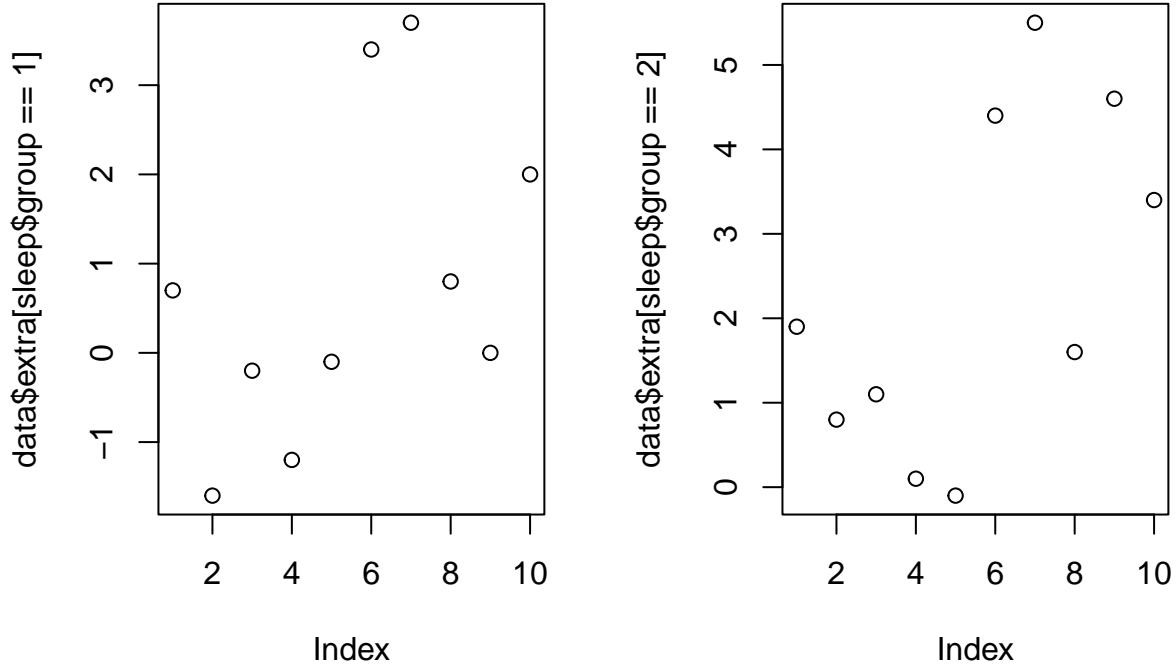
```
data <- sleep
```

On va donc formuler les hypothèses suivantes :

- H_0 : le médicament augmente significativement le temps de sommeil des individus
- H_1 : l'influence n'est pas significative.

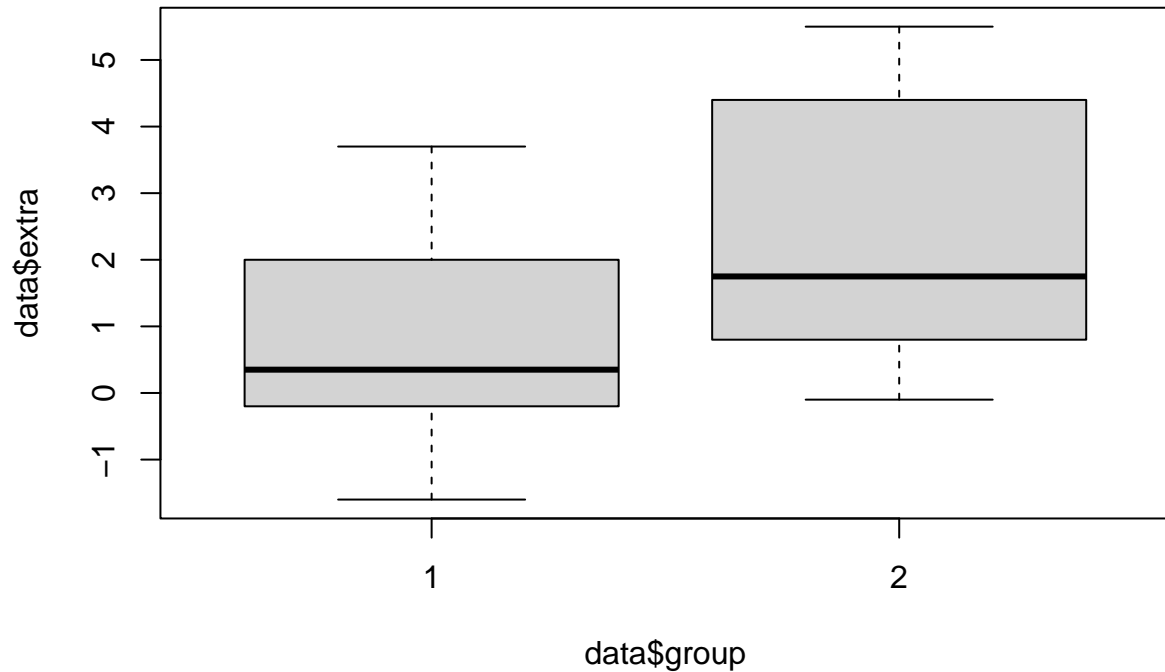
On va de plus faire un test unilatéral supérieur car on souhaite vérifier “l’augmentation du temps de sommeil”. On va d’abord regarder cela graphique pour les deux médicaments.

```
par(mfrow=c(1,2))
plot(data$extra[sleep$group==1])
plot(data$extra[sleep$group==2])
```



Il est difficile de se rendre compte de l’impact sur ce simple graphique. Peut-être qu’un boxplot serait plus visuel

```
boxplot(data$extra~data$group)
```



A priori le médicament 2 semble avoir un effet significatif contrairement au médicament numéro 1. Etant donnée la taille des échantillons (10), on va donc devoir faire des tests non paramétriques. Ici des de Wilcoxon

pour les individus de chaque groupe.

```
wilcox.test(sleep$extra[sleep$group==1],mu = 0,alternative = "greater")
```

```
## Warning in wilcox.test.default(sleep$extra[sleep$group == 1], mu = 0,
## alternative = "greater"): cannot compute exact p-value with zeroes
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data:  sleep$extra[sleep$group == 1]
## V = 31, p-value = 0.1716
## alternative hypothesis: true location is greater than 0
```

Le médicament 1 n'a donc pas un impact significatif sur le temps de sommeil des individus.

```
wilcox.test(sleep$extra[sleep$group==2],mu = 0,alternative = "greater")
```

```
## Warning in wilcox.test.default(sleep$extra[sleep$group == 2], mu = 0,
## alternative = "greater"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data:  sleep$extra[sleep$group == 2]
## V = 53.5, p-value = 0.004649
## alternative hypothesis: true location is greater than 0
```

Le médicament 2 a quant à lui un impact significatif sur le temps de sommeil des individus.

Même si cela n'est pas nécessaire, nous pourrions aussi chercher à comparer l'efficacité des deux médicaments pour savoir si l'un est plus efficace qu'un autre. Ce que nous pouvons faire à nouveau avec un test de Wilcoxon.

```
wilcox.test(sleep$extra[sleep$group==2],
            sleep$extra[sleep$group==1],
            mu = 0,
            alternative = "greater")
```

```
## Warning in wilcox.test.default(sleep$extra[sleep$group == 2],
## sleep$extra[sleep$group == : cannot compute exact p-value with ties
```

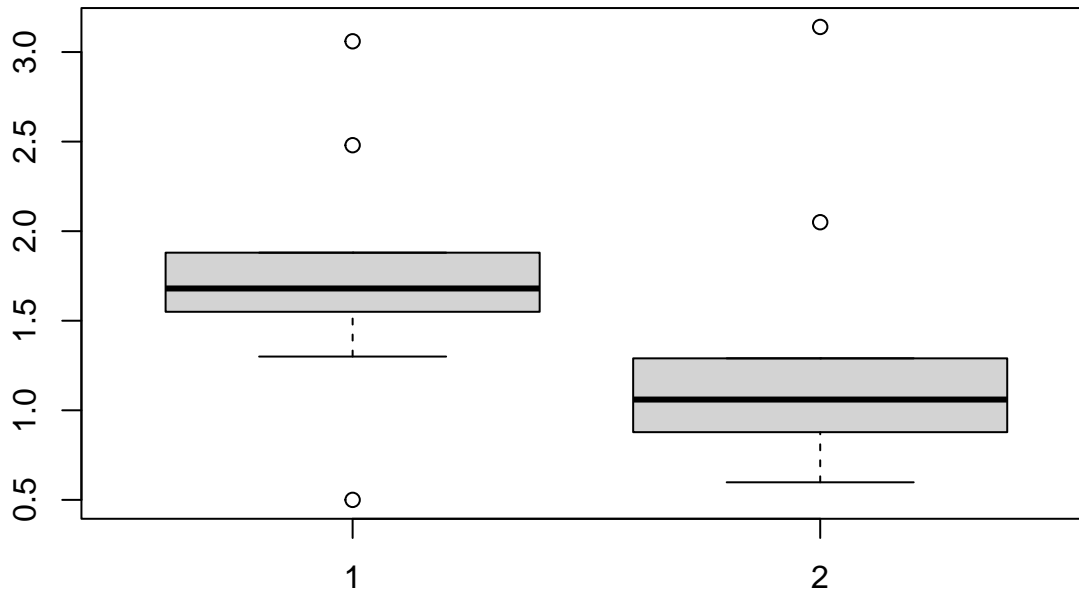
```
##
## Wilcoxon rank sum test with continuity correction
##
## data:  sleep$extra[sleep$group == 2] and sleep$extra[sleep$group == 1]
## W = 74.5, p-value = 0.03466
## alternative hypothesis: true location shift is greater than 0
```

On remarque bien, sans surprises, que le médicament 2 a un impact plus important sur l'allongement de la durée de sommeil que le médicament numéro 1.

Exercice 5

on souhaite mesurer l'effet d'un tranquillisant en évaluant l'échelle de dépression d'un patient à différents time points. Commençons par entrer les données et les représentées graphiquement

```
t1 <- c(1.83, 0.50, 1.62, 2.48, 1.68, 1.88, 1.55, 3.06, 1.30)
t2 <- c(0.878, 0.647, 0.598, 2.05, 1.06, 1.29, 1.06, 3.14, 1.29)
boxplot(t1,t2)
```



A priori, le tranquillisant semble avoir un effet sur les 9 patients sur lesquels portent cette étude. Pour confirmer cela, on formule le test d'hypothèses suivant :

- H_0 : le tranquillisant a une effet sur la réduction de l'anxiété du patient
- H_1 : le tranquillisant ne permet de réduire l'anxiété du patient.

On va à nouveau faire un test de wilcoxon, pour des raisons identiques à celles de l'exercice précédent. En revanche, on prendra soin d'indiquer que les deux échantillons ne sont pas indépendants (mesures répétées sur les mêmes patients).

```
wilcox.test(t1,t2,paired=TRUE,alternative="greater")
```

```
##
## Wilcoxon signed rank exact test
##
## data: t1 and t2
## V = 40, p-value = 0.01953
## alternative hypothesis: true location shift is greater than 0
```

Le tranquillisant a bien un effet, il permet de réduire l'anxiété chez les patients.