

M1 MABS

Contrôle continu de Traitement des données biologiques – octobre 2011

Instructions

Commencez par enregistrer ce fichier sur le bureau en lui donnant un nom de la forme *Nom_Prénom_Groupe_CC_TDB_oct_2011.doc* ; Exemple si je m'appelle Bob Marley et que je suis dans le groupe 2, mon fichier s'appellera Marley_Bob_Groupe2_CC_TDB_oct_2011.doc

Tout manquement à cette règle sera sanctionné à hauteur de 2 points sur la note finale.

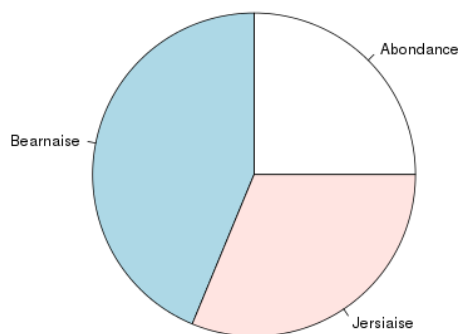
Pour chaque question, s'il y a lieu :

1. donnez la ou les commandes permettant d'y répondre
2. copiez/collez la réponse obtenue dans R (marche aussi bien pour les graphiques que le texte)
3. ajoutez un commentaire ou une explication.

Exemple :

Question 0 : Réalisez un camembert des mesures effectuées sur les races de vaches.

```
# Question 0  
pie(summary(vaches))
```



```
# Dans un premier temps, la fonction summary appliquée à vaches  
# permet d'obtenir les effectifs par race de vaches  
# ensuite, ces effectifs sont passé à la fonction pie qui trace un  
# camembert
```

Envoyez 5 minutes avant la fin de l'examen votre fichier de réponse aux adresses :

barriot@biotoul.fr

bonhomme@lrsv.ups-tlse.fr

Conseil :

-mettez-vous en copie du mail afin de garder une trace de votre travail

- laissez le fichier sur le « bureau » de l'ordinateur ET laissez l'ordinateur allumé

Vous êtes responsables de l'envoi du fichier en fin d'examen. Etant donné qu'il est facile d'antidater un fichier ainsi que son envoi, tout fichier non reçu en fin de séance recevra une note de 0/20.

Partie 1

La réaction hypersensible (HR) est une réponse d'un tissu végétal (ex : feuille) face à une attaque par un microorganisme parasite. Elle consiste en une nécrose locale du tissu. Lors de son stage en laboratoire, une étudiante a mesuré l'importance de la HR de plantes d'*Arabidopsis thaliana* mises en contact avec le champignon *Colletotrichum trifolii*, afin d'évaluer le degré de sensibilité d'*Arabidopsis* face à une attaque par *Colletotrichum*.

Deux jours après inoculation des spores sur une feuille par plante, pour 40 plantes, l'étudiante a mesuré la HR (surface en cm²). Elle obtient les données suivantes :

0.69, 3.71, 2.44, 0.81, 4.45, 3.70, 4.10, 3.33, 1.84, 3.25, 1.73, 2.93, 3.26, 1.27, 1.27, 4.15, 4.59, 3.07, 1.33, 1.54, 3.19, 3.37, 0.13, 1.02, 1.13, 0.69, 1.73, 0.59, 2.84, 2.67, 2.21, 1.70, 2.74, 1.42, 0.03, 1.38, 3.02, 3.10, 4.69, 1.60

Question 1 : Combien y-a-t-il : de variables ? De quelle(s) nature(s) ? Combien y a-t-il d'observations ?

Réponse en texte libre :

1 variable quantitative (la HR), et 40 observations.

Question 2 : Calculez la moyenne, la médiane et la variance de ces données.

Commandes réalisées dans R :

```
hr=c(0.69, 3.71, 2.44, 0.81, 4.45, 3.70, 4.10, 3.33, 1.84, 3.25,
1.73, 2.93, 3.26, 1.27, 1.27, 4.15, 4.59, 3.07, 1.33, 1.54, 3.19,
3.37, 0.13, 1.02, 1.13, 0.69, 1.73, 0.59, 2.84, 2.67, 2.21, 1.70,
2.74, 1.42, 0.03, 1.38, 3.02, 3.10, 4.69, 1.60)
```

Pour la moyenne :

```
mean(hr)
```

Résultats obtenus dans R :

```
[1] 2.31775
```

Médiane :

```
median(hr)
```

```
[1] 2.325
```

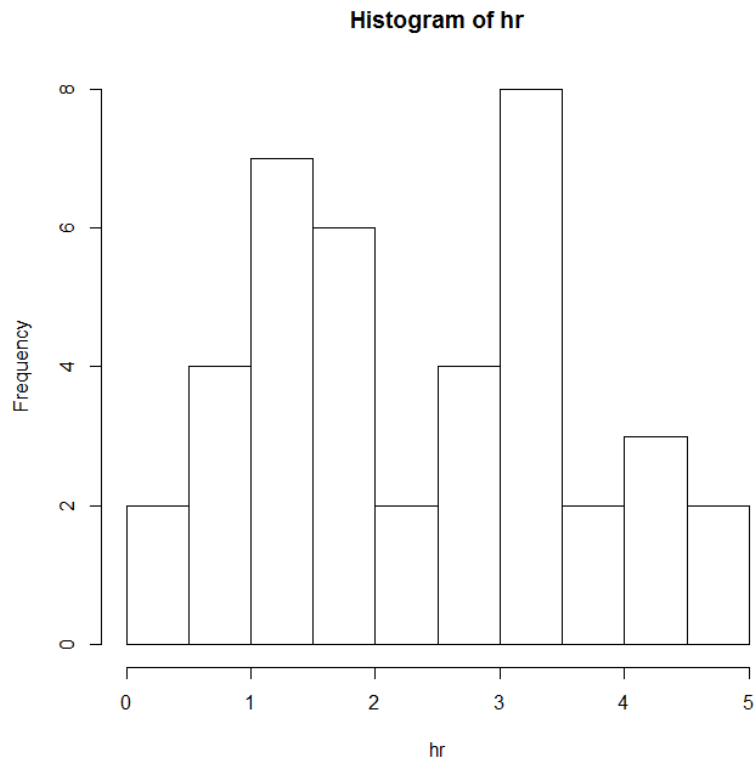
Variance :

```
var(hr)
```

```
[1] 1.640172
```

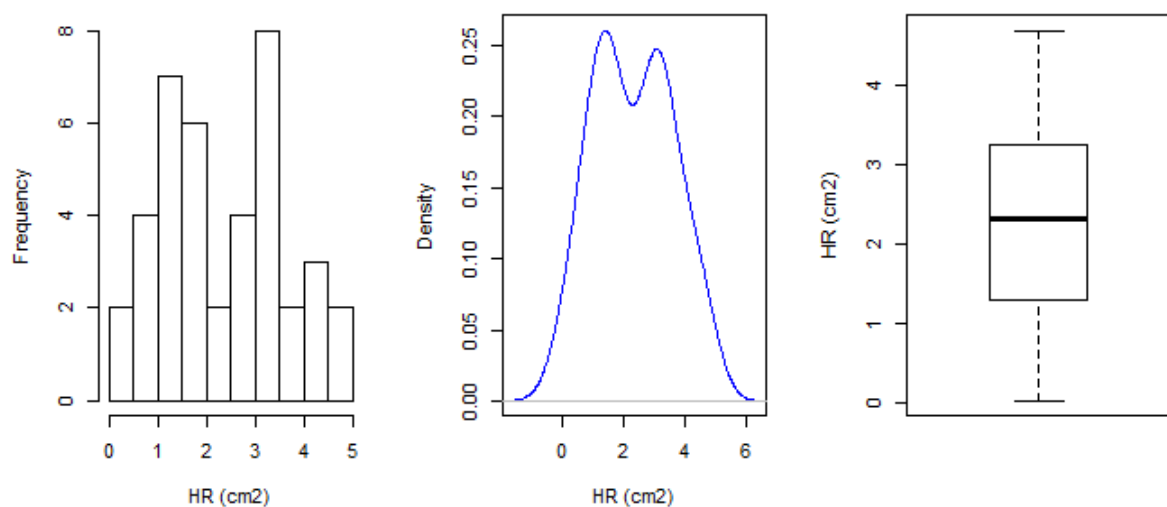
Question 3 : Donnez des représentations graphiques de ces données : histogramme (pour lequel vous choisirez un nombre d'intervalles vous semblant le plus approprié), courbe de densité et boîte à moustaches.

```
hist(hr,breaks=15) #15 met mieux en valeur le mélange
```



```
par(mfrow=c(1,3))
hist(hr,main="histogramme de la HR sur 40 plantes
d'Arabidopsis",xlab="HR (cm2)",breaks=15);
plot(density(hr),col="blue",main="densité de la HR sur 40 plantes
d'Arabidopsis",xlab="HR (cm2)");
boxplot(hr,main="boîte à moustaches de la HR sur 40 plantes
d'Arabidopsis",ylab="HR (cm2)")
```

ramme de la HR sur 40 plantes d'Amsité de la HR sur 40 plantes d'Arabioustaches de la HR sur 40 plantes c



Question 4 : Pourquoi la moyenne et la médiane sont-elles similaires dans ces données ?

La distribution est bimodale mais symétrique : homogénéité de la dispersion donc la moyenne et la médiane sont presque confondues

Question 5 : Concluez sur la variabilité de la sensibilité des plantes.

L'histogramme montre qu'il y a 2 types de comportement de HR dans cet échantillon de plantes : 7 plantes (17.5%) avec une $1 < HR < 1.5$ et 8 plantes (20%) avec une $3 < HR < 3.5$

Partie 2

En fait, l'étudiante a effectué ces mesures sur des individus de deux lignées bien distinctes d'*Arabidopsis thaliana* : Col0 et Ws4.

Mesures sur Col0 :

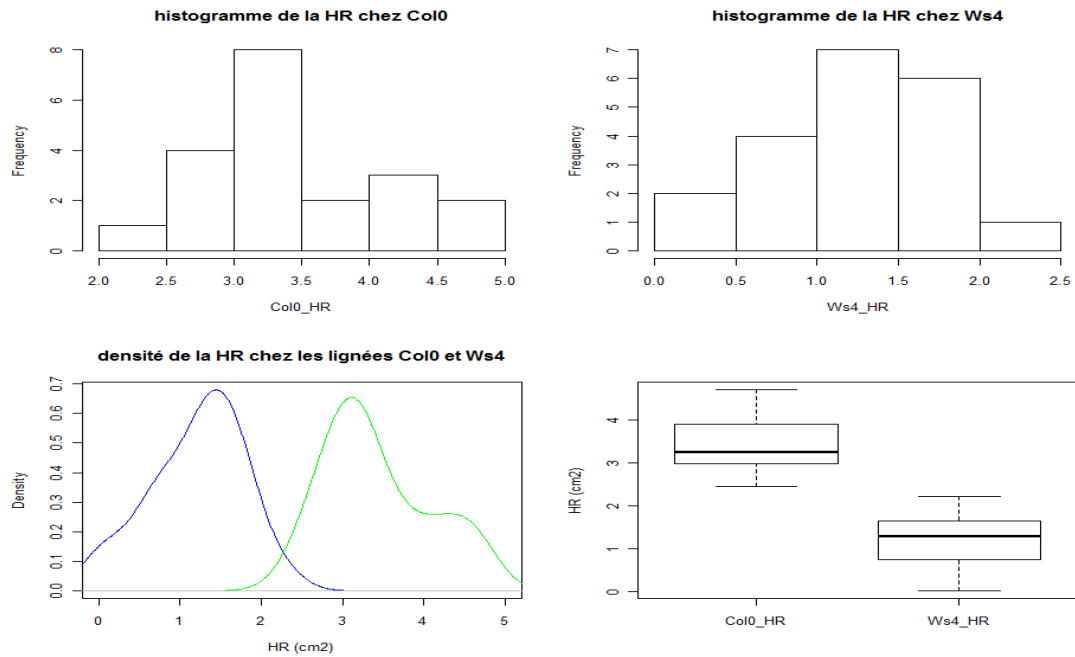
3.19, 4.59, 3.10, 2.67, 3.02, 2.74, 3.07, 3.37, 4.15, 3.25, 2.84, 3.71, 4.10, 3.70, 4.45, 2.93, 4.69, 3.33, 3.26, 2.44

Mesures sur Ws4 :

0.81, 1.73, 0.13, 1.13, 1.33, 2.21, 1.84, 1.60, 1.27, 0.03, 0.69, 1.02, 1.54, 1.42, 1.73, 1.27, 0.59, 1.70, 1.38, 0.69

Question 6 : Faites, comme précédemment, des représentations graphiques de ces données (et importez-les dans ce fichier). Affichez notamment les courbes de densité sur le même graphique, mais avec des couleurs différentes.

```
Col0_HR=c(3.19,4.59,3.10,2.67,3.02,2.74,3.07,3.37,4.15,3.25,2.84,3.71,4.10,3.70,4.45,2.93,4.69,3.33,3.26,2.44)
Ws4_HR=c(0.81,1.73,0.13,1.13,1.33,2.21,1.84,1.60,1.27,0.03,0.69,1.02,1.54,1.42,1.73,1.27,0.59,1.70,1.38,0.69)
par(mfrow=c(2,2))
hist(Col0_HR,main="histogramme de la HR chez Col0");hist(Ws4_HR,main="histogramme de la HR chez Ws4");
plot(density(Ws4_HR),col="blue",xlim=c(0,5),xlab="HR (cm2)",main="densité de la HR chez les lignées Col0 et Ws4");
lines(density(Col0_HR),col="green");
boxplot(Col0_HR,Ws4_HR, names=c("Col0_HR","Ws4_HR"),ylab="HR (cm2)")
```



Question 7 : L'étudiante se demande s'il y a une différence de HR entre Col0 et Ws4. Quel test statistique utiliseriez-vous pour tester cette différence ? Suppose-t-il de vérifier certaines hypothèses préalablement ? Effectuez les tests préalables si nécessaire puis effectuez le test de comparaison de ces 2 lignées. Que concluez-vous ?

Test de comparaison de moyenne : test de Student si normalité des échantillons et égalités des variances des 2 échantillons. Sinon, test de Wilcoxon (Mann et Whitney)

Tests de normalité :

`shapiro.test(Col0_HR)`

Shapiro-Wilk normality test

data: Col0_HR

W = 0.9296, p-value = 0.152

p-value = 0.152 > 0.05 donc on accepte l'hypothèse nulle → normalité

`shapiro.test(Ws4_HR)`

Shapiro-Wilk normality test

data: Ws4_HR

W = 0.9612, p-value = 0.5677

p-value = 0.5677 > 0 → normalité

Test d'homogénéité des variances :

`var.test(Col0_HR, Ws4_HR)`

F test to compare two variances

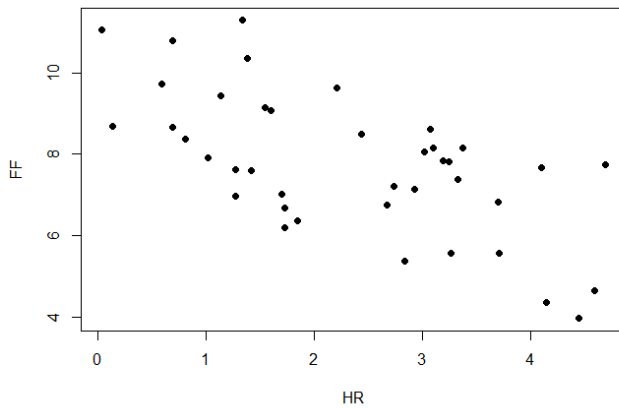
data: Col0_HR and Ws4_HR

F = 1.3286, num df = 19, denom df = 19, p-value = 0.5418

Question 8 : Comment feriez-vous pour comparer HR et FF ? En d'autres termes, analysez s'il y a une relation entre HR et FF et représentez graphiquement vos analyses.

On commence par inspecter visuellement les données avec un nuage de points :

```
plot(HR, FF)
```



Le graphique suggère une anti-corrélation (ou corrélation négative)

On recherche une relation entre ces deux variables (quantitatives), mesurées chez chacun des individus : calcul de la covariance et de la corrélation :

```
HR=c(0.03, 3.02, 1.73, 1.38, 1.54, 4.69, 0.69, 1.42, 4.59, 0.81, 3.33, 1.84, 1.02, 2.21, 2.44, 0.13, 2.67, 2.74, 3.71, 4.10, 2.84, 2.93, 1.13, 1.27, 1.27, 4.15, 4.45, 3.25, 3.26, 1.73, 3.19, 1.70, 0.59, 0.69, 1.60, 3.37, 3.70, 1.33, 3.07, 3.10)
```

```
FF=c(11.05, 8.06, 6.20, 10.35, 9.15, 7.75, 8.66, 7.61, 4.65, 8.37, 7.37, 6.37, 7.92, 9.64, 8.50, 8.68, 6.76, 7.20, 5.56, 7.67, 5.37, 7.14, 9.44, 7.62, 6.98, 4.37, 3.96, 7.82, 5.56, 6.68, 7.84, 7.02, 9.72, 10.79, 9.08, 8.16, 6.83, 11.29, 8.61, 8.15)
```

```
cov(HR, FF)
```

```
[1] -1.426305
```

Covariance : -1.426305

```
cor(HR, FF)
```

```
[1] -0.6423693
```

Corrélation : -0.6423693 → corrélation négative : quand le niveau de HR augmente, le flétrissement diminue.

Régression linéaire :

```
reg=lm(FF ~ HR)
```

```
reg
```

Call:

```
lm(formula = FF ~ HR)
```

Coefficients:

```
(Intercept)          HR  
    9.7643    -0.8696
```

droite d'équation $y = -0.87 + 9.76$

significativité des coefficients :

```
summary(reg)
```

Call:

```
lm(formula = FF ~ HR)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.0599	-1.0606	0.1035	0.9618	2.6823

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.7643	0.4444	21.973	< 2e-16 ***
HR	-0.8696	0.1683	-5.167	7.88e-06 ***

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

Residual standard error: 1.346 on 38 degrees of freedom

Multiple R-squared: 0.4126, Adjusted R-squared: 0.3972

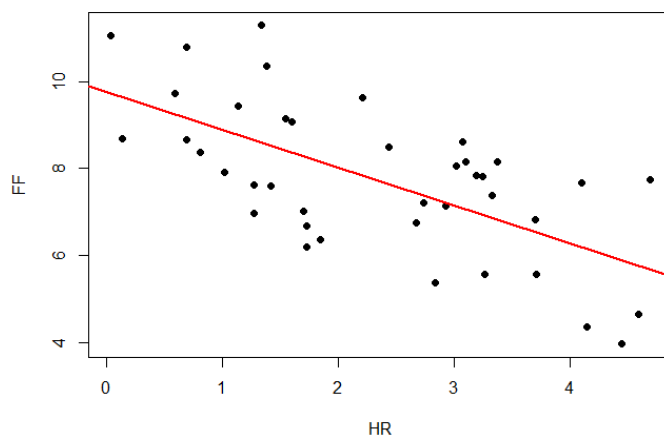
F-statistic: 26.7 on 1 and 38 DF, p-value: 7.876e-06

pende (-0.87) : p-value= 7.88e-06 → la pente est significativement différente de 0 (il y a une relation entre les variables)

ordonnée à l'origine (9.76) : 2e-16 → significativement différent de 0

Ajout de la droite sur le graphique :

```
abline(reg, col="red", lwd=2)
```



Calcul du coefficient de détermination (r^2) :

```
cor(HR, FF)^2  
[1] 0.4126384
```

$\text{cor}(\text{HR}, \text{FF})^2 = R^2 = 0.41$: On valide cette droite, mais elle n'explique que 41% de la variance des données.

Question 9 : Que concluez-vous ?

On conclut qu'il y a une relation linéaire entre le niveau de HR et le flétrissement des feuilles.

Une HR forte empêche le pathogène d'établir un point d'entrée dans la plante, et donc la plante ne montre pas de symptôme fort d'invasion par celui-ci (flétrissement faible)

A contrario, une HR faible n'est pas suffisante pour empêcher l'entrée du microorganisme dans la plante, et donc la plante est colonisée plus facilement par le microorganisme (flétrissement fort).

Il faut noter que certaines plantes peuvent individuellement présenter un comportement différents (d'où le R^2 faible) : d'autres facteurs sont donc probablement impliqués dans la colonisation de la

plante par le microorganisme ainsi que la réponse systémique des plantes face aux pathogènes. Toutefois le modèle linéaire décrit bien le comportement global des plantes face à l'invasion par *Colletotrichum*.