

Densité de population et ingestion de nourriture chez un insecte vecteur de la maladie de Chagas

F. Menu, A.B. Dufour, E. Desouhant et I. Amat

La fiche permet de se familiariser avec la planification et l'analyse expérimentale. Les questions biologiques auxquelles nous devons répondre nécessitent une planification expérimentale rigoureuse et l'utilisation d'outils statistiques adaptés comme le test de Levène et l'Analyse de variance (ANOVA). Les données présentées sont issues d'une étude de collègues argentins avec qui nous collaborons : E H. Weir, J.E. Rabinovich et Martin Torres. Les données utilisées dans cette fiche ne représentent qu'une petite partie des données originales.

Table des matières

1	La problématique	3
2	Les punaises	4
3	Protocole expérimental	4
4	Exploration des données	5
5	Effet de la densité sur la quantité de sang ingérée par les larves de Triatomines : dispositif expérimental aléatoire	8
5.1	Effet de la densité d'élevage sur la quantité moyenne de sang ingérée (ANOVA I)	8
5.2	Effet de la densité d'élevage sur la variance de sang ingérée	11
5.2.1	Homoscédasticité	11
5.2.2	Normalité : une autre condition d'application	14
5.2.3	Indépendance des résidus : dernière condition d'application	15
5.3	Rejeter ou non une hypothèse nulle, c'est une question de puissance du test! - Illustration d'un calcul de puissance sur un test de comparaison de deux moyennes	15

6	Effet de la densité et des répliques sur la quantité de sang ingérée par les larves de Triatomines : dispositif expérimental en blocs aléatoires complets	19
6.1	Exercice : ANOVA à 1 et 2 facteurs avec taille d'échantillon constante	19
6.1.1	ANOVA I	19
6.1.2	ANOVA II	21
6.2	ANOVA à 2 facteurs avec taille d'échantillon non constante (Fichier de données complet sans sélection de 5 valeurs par bloc) . .	22
7	Effet de la densité sur la quantité de sang ingérée par les adultes de Triatomines	23
7.1	Charger les données	23
7.2	Analyse des données (ANOVA III)	26
	Références	27

1 La problématique

Plusieurs espèces de punaises hématophages appartenant à la famille des *Reduviidae* et vivants en Amérique Latine, constituent des vecteurs de la maladie de Chagas. Cette maladie, mortelle pour l'homme (appelée aussi trypanosomiase) est transmise par un parasite protozoaire : *Trypanosoma cruzi* et constitue un problème de santé publique majeur en Amérique Latine. L'Organisation Mondiale de la Santé (OMS) a estimé qu'environ 100 millions de personnes sont exposées au risque d'infection et que 16 à 18 millions sont infectées. Actuellement, aucun traitement thérapeutique acceptable n'est disponible et le développement d'un vaccin est encore illusoire. Le contrôle de la transmission de la maladie est donc basé sur une lutte contre les insectes vecteurs (lutte insecticide et lutte biologique) et sur le contrôle des banques de sang. Plusieurs espèces de punaises transmettent la maladie, citons *Triatoma dimidiata*, *T. infestans* et *Rhodnius prolixus*. Ces punaises se nourrissent du sang de nombreux hôtes sauvages et domestiques : mammifères (dont l'homme) et oiseaux. Une lutte efficace contre ces punaises nécessite une bonne connaissance de leurs dynamiques de populations. Or, en dépit de nombreuses études sur la physiologie, l'éthologie et l'écologie de certaines de ces punaises comme *Rhodnius prolixus*, les mécanismes qui régulent leurs populations sont peu connus.

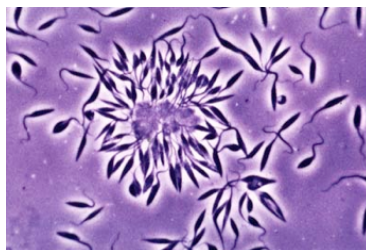
Des mécanismes de régulation de populations dépendant de la densité ont été montrés chez d'autres insectes hématophages comme les moustiques (Gillett, 1967 [1]) et les mouches tsé-tsé (Langley, 1967 [2]). Ces processus font intervenir la notion d' "inaccessibilité" des hôtes par les insectes en raison d'un comportement d'énervement et d'agressivité des hôtes envers les insectes quand leur densité de population atteint un niveau élevé. Les travaux de Schofield [3] suggèrent des mécanismes similaires chez *Triatoma infestans*. Ces mécanismes ne sont pas liés à la disponibilité en hôte (celle-ci étant assurée dans les maisons, les étables et les poulaillers des zones rurales), mais dépendraient de l'accessibilité des hôtes par les punaises. Cette hypothèse suppose qu'une augmentation de densité en punaises produit un état de nuisance et d'irritation des hôtes entraînant une série de conséquences néfastes pour les punaises (e.g. mortalité par prédation, baisse de la fécondité et augmentation du temps de développement). En comparant deux densités de populations de *T. infestans* (10 et 40 punaises par souris), Gadea (1984), a montré une différence significative dans la fécondité mais pas dans la quantité moyenne de sang ingérée par les femelles. Au contraire, Schofield [4] a montré qu'une réduction dans la quantité de sang absorbé induit non seulement une diminution du nombre moyen d'oeufs pondus par des punaises *T. infestans* mais aussi une augmentation dans le temps de développement des stades larvaires.

L'objectif de cette fiche est de tester chez *R. prolixus*, en conditions de laboratoire, l'influence de la densité d'élevage sur la distribution de la quantité de sang absorbée par les punaises (type de distribution, moyenne, variance). La quantité de nourriture ingérée sera mesurée après 3 jours. Cette quantité conditionne le succès reproducteur des punaises car elle influence le taux de survie, la fécondité et la mue des insectes.



2 Les punaises

Les triatomines (*Triatoma*) et les punaises du genre *Rhodnius* sont des punaises hématophages dont le cycle ne peut s'effectuer sans consommation régulière de sang (cf. photos). Elles se nourrissent sur des animaux sauvages comme des opossums, des mulots, mais aussi sur l'homme et ses animaux domestiques (e.g. poules, chiens, chats, vaches, ânes...). Ces punaises présentent une activité nocturne. Dans la journée, elles sont réfugiées dans des interstices naturels ou dans les fissures des murs ou les toits en chaume des habitations ainsi que dans les murets en pierre qui ferment les jardins. La nuit, les punaises se déplacent vers leurs hôtes et transpercent leur peau à l'aide de leur long rostre et absorbent leur sang. La contamination par le Trypanosome *T. cruzi* s'effectue à partir des excréments des punaises qu'elles laissent sur la peau de leur hôte. En phase terminale de la maladie, le Trypanosome s'attaque aux cellules cardiaques et induit des infarctus, suivis de la mort de l'homme. Cette phase terminale apparaît après une longue période de latence (jusqu'à une douzaine d'années).

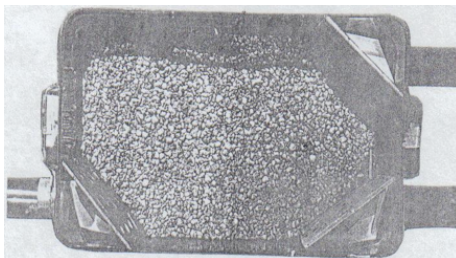


3 Protocole expérimental

L'expérience a été réalisée en Argentine. Des hamsters mâles (125-150 g) ont été utilisés comme hôte. Chez les punaises *R. prolixus*, deux stades ont été étudiés : le cinquième stade larvaire et le stade adulte (le sexe des punaises adultes était contrôlé). Les punaises étaient élevées dans une salle climatisée à $28 \pm 1^\circ\text{C}$ et $60 \pm 10\%$ d'Humidité Relative. Elles n'étaient pas nourries après leur mue (mue de quatrième stade larvaire ou mue imaginale). Les hamsters sont restés dans la pièce d'élevage des punaises durant une semaine avant le début de l'expérience (acclimatation). Pour chaque réplique de l'expérience (cf. ci-après), des hamsters différents étaient utilisés.

L'expérience a été menée dans un système composé de quatre boîtes en plastique interconnectées ($40 \times 20 \times 10$ cm) : une boîte centrale où les punaises et le hamster étaient introduits au début de chaque répétition et trois boîtes latérales. Des tuyaux en entonnoirs permettaient aux punaises de se disperser seulement de la boîte centrale vers les boîtes latérales. Un hamster était aussi enfermé dans chacune des boîtes latérales ; il était isolé de la connexion avec la boîte centrale par un grillage métallique rigide vertical. Les hamsters des boîtes latérales étaient utilisés pour produire des stimuli de dispersion des punaises de la boîte centrale vers les boîtes latérales. Le fond des boîtes était recouvert de papier-filtre pour absorber les excréments des hamsters et des punaises. Dans les trois coins de la boîte centrale, des cloisons perforées verticales en bois (cf. figure) offraient un refuge aux punaises (contre la prédation par le hamster quand les

punaises ne se nourrissaient pas) et permettaient aux punaises de rejoindre le hamster pour s'alimenter.



Quatre niveaux de densités de punaises (10, 20, 30 et 40) et 4 répliques par niveau de densité ont été réalisés. Chaque essai (i.e. une réplique d'un niveau de densité donné) durait 3 jours. La variable à expliquer par la densité est la quantité de sang ingérée durant ces trois jours. Cette quantité a été calculée en pesant avant et après l'expérience les punaises (précision ± 1 mg). Pour leur deuxième pesée, les punaises étaient identifiées grâce à des marques de couleurs différentes (sur chaque patte et sur le pronotum) réalisées en début d'expérience. Les punaises retrouvées mortes et/ou qui avaient quitté la boîte centrale, étaient remplacées par le même type d'insectes afin de garder des densités constantes durant l'expérience. Ces punaises n'ont pas été utilisées pour les mesures de quantité de sang ingéré. Deux séries d'expériences ont été réalisées dans les mêmes conditions. Une des séries a utilisé des larves de cinquième stade, l'autre des adultes. Chacune des quatre répliques (4 niveaux de densité par réplique) de l'expérience a duré 3 jours (1 réplique par semaine). Les répliques constituent donc quatre blocs temporels. Pour le niveau de densité "40", une cinquième réplique a été réalisée le même jour que la quatrième réplique. Plutôt que d'éliminer de l'analyse cette cinquième réplique (présente uniquement dans la modalité "40"), les données de la quatrième et de la cinquième réplique ont été regroupées et constituent la "quatrième" réplique de la densité "40". Les punaises ont été affectées aléatoirement à chaque niveau de densité et à chaque bloc.

4 Exploration des données

Cette première partie est destinée à explorer le jeu de données via des statistiques descriptives et des représentations graphiques. Le fichier `larve.txt` (à l'adresse suivante : <http://pbil.univ-lyon1.fr/R/donnees/larve.txt>) contient la quantité individuelle de sang (en mg) ingérée par les punaises selon la densité d'élevage (10, 20, 30 et 40 punaises) et la réplique (1, 2, 3 ou 4 = blocs temporels).

```
larve <- read.table("http://pbil.univ-lyon1.fr/R/donnees/larve.txt",  
                  h = T)  
attach(larve)
```

On va considérer la variable densite comme une variable qualitative :

```
densite <- factor(densite)
```

Nombre de données par colonne :

```
length(ingestion)
[1] 217
```

Nombre de données en fonction de la densité :

```
summary(densite)
 10  20  30  40
 36  28  44 109
```

La même commande (`summary`) ne donne pas le même type de sortie pour la variable `ingestion`, pourquoi ? Interpréter les sorties.

```
summary(ingestion)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00   36.20   90.10   80.22  118.10  205.80
```

Estimation de la variance :

```
var(ingestion)
[1] 2673.353
```

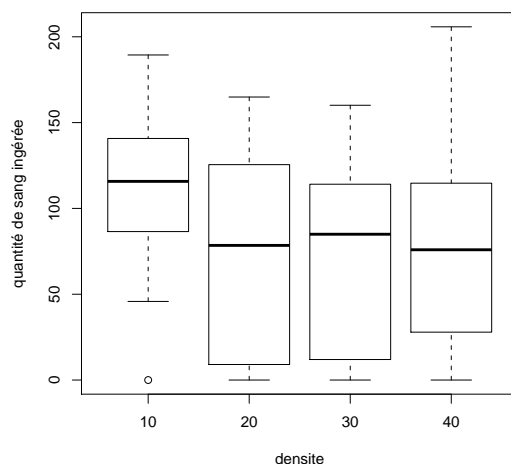
Est-ce la variance estimée (i.e. de la population) ou celle de l'échantillon ?

```
tapply(ingestion, densite, mean)
      10      20      30      40
111.00278 75.92857 74.02045 73.65229
tapply(ingestion, densite, var)
      10      20      30      40
1707.089 3473.600 2572.917 2520.854
```

Allez voir l'aide de la fonction (`?tapply`).

Un résumé graphique des données :

```
boxplot(ingestion ~ densite, xlab = "densite", ylab = "quantité de sang ingérée")
```



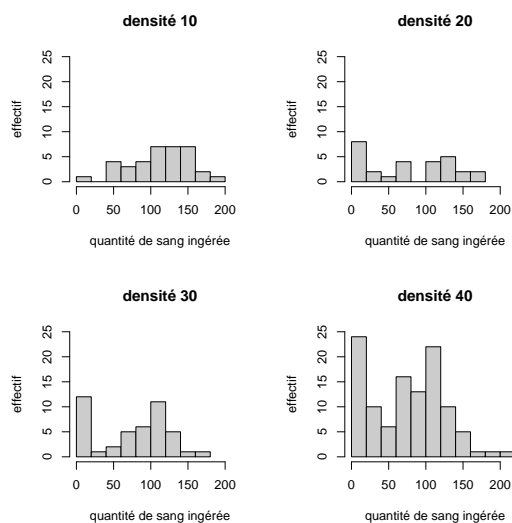
Que représentent les 5 traits horizontaux pour chaque boîte à moustaches ? Avez-vous déjà des hypothèses biologiques pour expliquer l'effet de la densité en punaises sur la quantité de sang ingérée ?

Distribution des données (quantités de sang ingérées) pour les 4 densités :

```

par(mfrow = c(2, 2))
hist(ingestion[densite == "10"], col = grey(0.8), main = "densité 10",
     xlim = c(0, 225), ylim = c(0, 25), xlab = "quantité de sang ingérée",
     ylab = "effectif")
hist(ingestion[densite == "20"], col = grey(0.8), main = "densité 20",
     xlim = c(0, 225), ylim = c(0, 25), xlab = "quantité de sang ingérée",
     ylab = "effectif")
hist(ingestion[densite == "30"], col = grey(0.8), main = "densité 30",
     xlim = c(0, 225), ylim = c(0, 25), xlab = "quantité de sang ingérée",
     ylab = "effectif")
hist(ingestion[densite == "40"], col = grey(0.8), main = "densité 40",
     xlim = c(0, 225), ylim = c(0, 25), xlab = "quantité de sang ingérée",
     ylab = "effectif")

```



Et si on tient compte des répliques pour décrire les données...

```

tapply(ingestion, list(densite, replique), mean)
      1      2      3      4
10 79.21250 98.98000 133.54000 129.65000
20 68.00000 59.40000 109.14000 75.54444
30 69.60833 29.81429 103.34286 69.64545
40 70.68750 80.56667 80.13333 69.68148

tapply(ingestion, list(densite, replique), var)
      1      2      3      4
10 1723.430 1958.175 953.5204 583.9143
20 5476.654 4683.748 2326.9730 1883.2003
30 2865.761 2276.591 1369.2257 2149.1867
40 1722.242 1920.096 3317.2224 2808.9012

```

Dans une première partie nous allons nous intéresser uniquement au facteur densité (paragraphe 5) puis nous inclurons dans nos modèles statistiques les deux facteurs explicatifs. Pour rappel, l'objectif de cette fiche est de comparer un dispositif simple aléatoire avec un dispositif en blocs aléatoires complets.

5 Effet de la densité sur la quantité de sang ingérée par les larves de Triatomines : dispositif expérimental aléatoire

5.1 Effet de la densité d'élevage sur la quantité moyenne de sang ingérée (ANOVA I)

```
options(show.signif.stars = F)
```

Pour commencer allez voir l'aide pour la fonction `lm`.

```
help("lm")
m0 <- lm(ingestion ~ densite)
summary(m0)

Call:
lm(formula = ingestion ~ densite)
Residuals:
    Min       1Q   Median       3Q      Max
-111.003  -49.103   4.948   39.848  132.148

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  111.003     8.364   13.272 < 2e-16
densite20    -35.074    12.645   -2.774 0.006034
densite30    -36.982    11.278   -3.279 0.001216
densite40    -37.350     9.647   -3.872 0.000144

Residual standard error: 50.18 on 213 degrees of freedom
Multiple R-squared:  0.07104,    Adjusted R-squared:  0.05796
F-statistic:  5.43 on 3 and 213 DF,  p-value: 0.001287
```

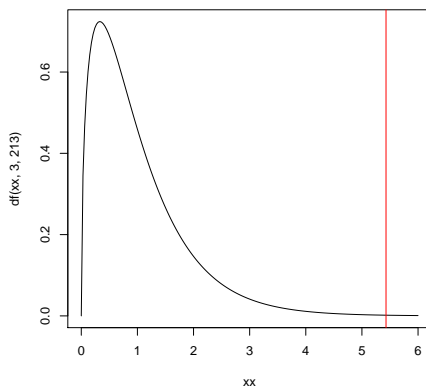
Retrouve-t-on les moyennes précédemment calculées pour chacune des densités ?
Pour obtenir le tableau de décomposition de la variance :

```
anova(lm(ingestion ~ densite))
Analysis of Variance Table
Response: ingestion
      Df Sum Sq Mean Sq F value    Pr(>F)
densite  3  41021  13673.8   5.4295 0.001287
Residuals 213 536423   2518.4
```

Retrouvez la valeur du R^2 (issu de la commande `summary(m0)`) à l'aide des valeurs du tableau d'ANOVA.

On trace la distribution théorique de la statistique F (à 3 et 213 ddl) sous l'hypothèse nulle (H_0). On positionne la valeur du F observé sur ce graphe. Retrouvez la valeur de la probabilité critique.

```
xx <- seq(0, 6, le = 200)
plot(xx, df(xx, 3, 213), type = "l")
abline(v = 5.4295, col = "red")
```

On peut retrouver la valeur de la probabilité critique à l'aide de la commande suivante :

```
1 - pf(5.4295, 3, 213)
[1] 0.001286680
```

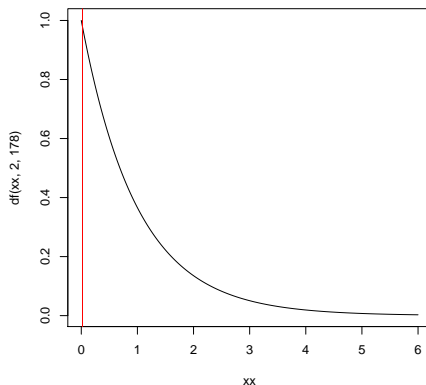
et la valeur seuil du F avec :

```
qf(0.95, 3, 213)
[1] 2.646988
```

Retrouver la signification statistique des fonctions `pf` et `qf`.

D'après les hypothèses que vous avez formulées à partir des `boxplot`, on se propose de tester si la significativité du test d'ANOVA provient de la différence entre la densité 10 et les densités 20, 30 et 40.

```
anova(lm(ingestion[densite != "10"] ~ densite[densite != "10"]))
Analysis of Variance Table
Response: ingestion[densite != "10"]
          Df Sum Sq Mean Sq F value Pr(>F)
densite[densite != "10"]  2    116    57.87  0.0216 0.9786
Residuals                178 476675 2677.95
xx <- seq(0, 6, le = 200)
plot(xx, df(xx, 2, 178), type = "l")
abline(v = 0.0216, col = "red")
```



Une autre façon de faire cette comparaison serait de modifier les contrastes. Les contrastes qu'utilise \mathbb{R} par défaut :

```
contrasts(densite)
  20 30 40
10  0  0  0
20  1  0  0
30  0  1  0
40  0  0  1
```

Les densités 20, 30 et 40 sont comparées à la densité 10 (voir `summary(m0)` et la signification de l'intercept et des coefficients du modèle).

Modifions les contrastes pour comparer les densités 20 et 30, la densité 40 aux densités 20 et 30, et enfin, la densité 10 aux 3 autres.

```
newcontrasts <- matrix(c(0, -1/2, 1/2, 0, 0, -1/3, -1/3, 2/3, 3/4,
  -1/4, -1/4, -1/4), byrow = F, nrow = 4)
contrasts(densite) <- newcontrasts
contrasts(densite)
  [,1] [,2] [,3]
10  0.0  0.0000000  0.75
20 -0.5 -0.3333333 -0.25
30  0.5 -0.3333333 -0.25
40  0.0  0.6666667 -0.25

m1 <- lm(ingestion ~ densite)
summary(m1)

Call:
lm(formula = ingestion ~ densite)
Residuals:
    Min       1Q   Median       3Q      Max
-111.003  -49.103   4.948   39.848  132.148

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    83.651     3.875   21.588 < 2e-16
densite1       -1.908     12.132  -0.157  0.875171
densite2       -1.322     7.739  -0.171  0.864511
densite3       36.469     9.427   3.868  0.000146

Residual standard error: 50.18 on 213 degrees of freedom
Multiple R-squared:  0.07104,    Adjusted R-squared:  0.05796
F-statistic:  5.43 on 3 and 213 DF,  p-value: 0.001287

anova(m1)

Analysis of Variance Table
Response: ingestion
          Df Sum Sq Mean Sq F value    Pr(>F)
densite    3  41021  13673.8    5.4295 0.001287
Residuals 213 536423   2518.4
```

Lorsque l'on spécifie les contrastes, l'intercept devient la moyenne "globale" de la quantité de sang ingérée (ici, la moyenne des moyennes de la quantité de sang ingérée aux quatre densités). Le premier coefficient correspond à la différence entre les ingestions moyennes aux densités 20 et 30. Le second coefficient correspond à la différence entre l'ingestion moyenne à la densité 40 et la moyenne des ingestions aux densités 20 et 30. Enfin, le troisième coefficient correspond à la différence entre l'ingestion moyenne à la densité 10 et la moyenne du groupe de densités 20, 30 et 40.

```
sum(tapply(ingestion, densite, mean))/4
[1] 83.65102

tapply(ingestion, densite, mean)
      10      20      30      40
111.00278  75.92857  74.02045  73.65229
74.02045 - 75.92857
[1] -1.90812

(73.65229 - 1/2 * (75.92857 + 74.02045))
```

```
[1] -1.32222
      111.00278 - 1/3 * (75.92857 + 74.02045 + 73.65229)
[1] 36.46901
```


Comparer ces résultats à ceux de `m0`. Interpréter, et discuter de l'intérêt d'un plan expérimental équilibré (même nombre de répétitions dans chaque modalité testée).

5.2 Effet de la densité d'élevage sur la variance de sang ingérée

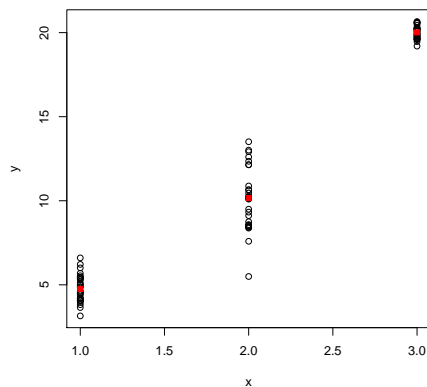
5.2.1 Homoscédasticité

Test de Levene : Exercice Le principe du test de Levene repose sur la comparaison des moyennes des valeurs absolues des résidus par rapport à la médiane ou à la moyenne. Le choix de la médiane ou de la moyenne dépend de la symétrie de la distribution des données. Pour comprendre le test de Levene, créons un jeu de données. Chaque colonne a 30 données et correspond par exemple à la quantité de sang ingérée pour 3 densités différentes A, B et C.

```
y1 <- rnorm(30, 5, 1)
y2 <- rnorm(30, 10, 2)
y3 <- rnorm(30, 20, 0.4)
y <- c(y1, y2, y3)
densiteC <- c(rep("C", 30))
densiteB <- c(rep("B", 30))
densiteA <- c(rep("A", 30))
x <- c(densiteA, densiteB, densiteC)
x <- as.factor(x)
x
[1] A A A A A A A A A A A A A A A A A A A A A A A A A A A A B B B B B B B B B B
[41] B B B B B B B B B B B B B B B B B B B B C C C C C C C C C C C C C C C C C C
[81] C C C C C C C C C C
Levels: A B C
```

Pour pouvoir représenter tous les points il faut que `x` soit une variable quantitative (sinon  produit des `boxplot`).

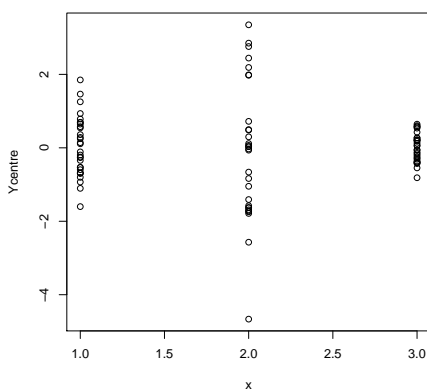
```
x = as.numeric(x)
x
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2
[41] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
[81] 3 3 3 3 3 3 3 3 3 3
plot(x, y)
points(3, median(y[x == 3]), col = "red", pch = 22, bg = "red")
points(2, median(y[x == 2]), col = "red", pch = 22, bg = "red")
points(1, median(y[x == 1]), col = "red", pch = 22, bg = "red")
```



Commenter ce graphique.

Centrons les données par rapport à la médiane de chaque groupe. On obtient donc les résidus par rapport à la médiane.

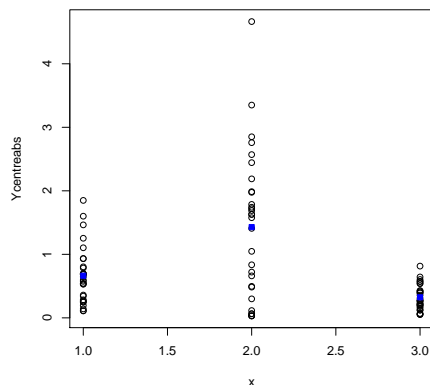
```
Ycentre = c(y1 - median(y1), y2 - median(y2), y3 - median(y3))
plot(x, Ycentre)
```



Commenter ce graphique.

Prenons maintenant la valeur absolue de chacun des nuages de points (sur le graphique ci-dessous ont été ajoutées les moyennes de chacun des nouveaux nuages de points obtenus).

```
Ycentreabs = c(abs(y1 - median(y1)), abs(y2 - median(y2)), abs(y3 -
  median(y3)))
plot(x, Ycentreabs)
points(3, mean(abs(y3 - median(y3))), col = "blue", pch = 22, bg = "blue")
points(2, mean(abs(y2 - median(y2))), col = "blue", pch = 22, bg = "blue")
points(1, mean(abs(y1 - median(y1))), col = "blue", pch = 22, bg = "blue")
```



Que décrivent les moyennes de ces 3 nouveaux nuages de points (en bleu sur la figure)? Comprenez vous mieux maintenant que pour comparer des variances de différents nuages de points, on peut comparer les moyennes des résidus par rapport à la médiane, résidus pris en valeur absolue?

Revenons à nos données Pour chaque densité on crée un nouveau vecteur (par exemple pour la densité 10, `ing10a`) constitué de la valeur absolue (`abs`) de la différence entre chaque valeur du vecteur initiale et la médiane de la densité considérée (`median`). Puis, on réunit ces 4 vecteurs en un seul (`ingabs`).

```
ing10a <- abs(ingestion[densite == "10"] - median(ingestion[densite ==
"10"]))
ing10a
[1] 115.75 13.35 27.75 6.35 54.15 69.95 4.85 22.55 65.05 5.95 64.45
[12] 43.65 40.75 9.05 24.55 58.65 37.75 53.85 12.25 1.75 14.25 0.75
[23] 44.05 73.65 0.75 17.95 27.75 54.15 50.25 30.75 12.45 24.25 26.45
[34] 4.25 7.35 25.45

ing20a <- abs(ingestion[densite == "20"] - median(ingestion[densite ==
"20"]))
ing30a <- abs(ingestion[densite == "30"] - median(ingestion[densite ==
"30"]))
ing40a <- abs(ingestion[densite == "40"] - median(ingestion[densite ==
"40"]))
ingabs <- c(ing10a, ing20a, ing30a, ing40a)
```


Analysons ce nouveau jeu de données.

```
anova(lm(ingabs ~ densite))
Analysis of Variance Table
Response: ingabs
          Df Sum Sq Mean Sq F value Pr(>F)
densite    3  6270 2090.03  2.8134 0.04026
Residuals 213 158232  742.88

summary(lm(ingabs ~ densite))
Call:
lm(formula = ingabs ~ densite)
Residuals:
    Min       1Q   Median       3Q      Max
-50.39 -22.13  -3.53   22.57  87.37

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  41.934     2.105  19.925 < 2e-16
densite1    -9.724     6.589  -1.476  0.14146
densite2    -4.143     4.203  -0.986  0.32541
densite3   -13.434     5.120  -2.624  0.00933
```

```
Residual standard error: 27.26 on 213 degrees of freedom
Multiple R-squared: 0.03812, Adjusted R-squared: 0.02457
F-statistic: 2.813 on 3 and 213 DF, p-value: 0.04026
```

Quelle information biologique peut on extraire de l'étude de la variance ?
 Pour réaliser le test de Levene directement sur , vous devez charger la bibliothèque `car`. Pour cela, allez dans le menu "packages" puis "installer les packages". Consulter la fiche TDR11 où toute la procédure est expliquée.

```
library(car)
levene.test(ingestion, densite)
Levene's Test for Homogeneity of Variance
  Df F value Pr(>F)
group 3 2.8134 0.04026
 213
```

Le test passe par une étape où l'on centre par rapport à la médiane. On aurait pu centrer par rapport à la moyenne. Quels sont les avantages d'utiliser la médiane ?
 Et si on enlève la densité 10 :

```
levene.test(ingestion[densite != "10"], densite[densite != "10"])
Levene's Test for Homogeneity of Variance
  Df F value Pr(>F)
group 2 1.3427 0.2638
 178
```

Interpréter ces résultats statistiquement et biologiquement. L'homoscédasticité est aussi une condition d'application des modèles linéaires.

5.2.2 Normalité : une autre condition d'application

```
help("shapiro.test")
shapiro.test(ingestion)
Shapiro-Wilk normality test
data: ingestion
W = 0.9418, p-value = 1.252e-07
```

Quelle information donne cette analyse ? Est-ce judicieux ?

```
shapiro.test(ingestion[densite == "10"])
Shapiro-Wilk normality test
data: ingestion[densite == "10"]
W = 0.9739, p-value = 0.5421
shapiro.test(ingestion[densite == "20"])
Shapiro-Wilk normality test
data: ingestion[densite == "20"]
W = 0.8867, p-value = 0.00565
shapiro.test(ingestion[densite == "30"])
Shapiro-Wilk normality test
data: ingestion[densite == "30"]
W = 0.8936, p-value = 0.000695
shapiro.test(ingestion[densite == "40"])
Shapiro-Wilk normality test
data: ingestion[densite == "40"]
W = 0.9463, p-value = 0.0002515
```

Interpréter les résultats ci-dessus statistiquement et biologiquement. Pour une illustration du théorème central limite : Cf. cours.

De manière approchée (Rappel de cours)

Calcul de puissance associée au test de comparaison de deux moyennes :

H0 : $\mu_1 = \mu_2$

H1 : $\mu_1 \neq \mu_2, \mu_1 - \mu_2 = \delta$

β est le risque de 2^e espèce = $\text{proba}(\text{accepter H0} \mid \text{H1})$

$$\beta = p \left(\frac{|\bar{X}_1 - \bar{X}_2|}{\sigma \sqrt{\frac{2}{n}}} < u_{1-\frac{\alpha}{2}} \right) \text{ sous H1}$$

$$\beta = p \left(-u_{1-\frac{\alpha}{2}} < \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{\frac{2}{n}}} < u_{1-\frac{\alpha}{2}} \right) \text{ sous H1}$$

$$\beta = p \left(-u_{1-\frac{\alpha}{2}} - \frac{\mu_1 - \mu_2}{\sigma \sqrt{\frac{2}{n}}} < \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{2}{n}}} < u_{1-\frac{\alpha}{2}} - \frac{\mu_1 - \mu_2}{\sigma \sqrt{\frac{2}{n}}} \right)$$

$$\beta = p \left(-u_{1-\frac{\alpha}{2}} - \delta_0 < Z < u_{1-\frac{\alpha}{2}} - \delta_0 \right) \text{ avec } Z \sim \mathcal{N}(0, 1) \text{ et } \delta_0 = \frac{\mu_1 - \mu_2}{\sigma \sqrt{\frac{2}{n}}}$$

$$\beta = p(u' < Z < u'') = \psi(u'') - \psi(u') \text{ avec } u' = -u_{1-\frac{\alpha}{2}} - \delta_0 \text{ et } u'' = u_{1-\frac{\alpha}{2}} - \delta_0$$

Donc la puissance vaut $1 - \beta = 1 - \psi(u'') + \psi(u')$

Approximation pour $\alpha < 0.1$ et $u'' < 0$: $1 - \beta = 1 - \psi(u'')$

Puissance associée au test de comparaison des quantités moyennes de sang ingérées aux densité 20 et 30 en fonction de δ (variances inconnues égales, approximation par la loi normale, approximation pour $\alpha < 0.1$ ($\alpha = 0.05$) et $u'' = u_{1-\frac{\alpha}{2}} - \delta_0 < 0$).

```
delta <- seq(20, 200, le = 100)
estvarcom <- (27 * var(ingestion[densite == 20]) + 43 * var(ingestion[densite ==
30]))/(28 + 44 - 2)
delta0 <- delta/sqrt(estvarcom * (1/28 + 1/44))
```

Calcul de u' et u''

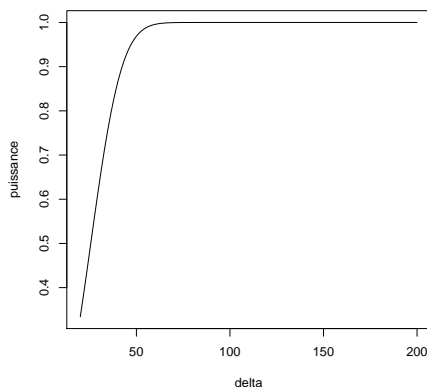
```
up <- -qnorm(0.975) - delta0
upp <- qnorm(0.975) - delta0
```


Notez l'utilisation de la fonction `qnorm` pour trouver $u_{1-\frac{\alpha}{2}}$. u'' (`upp`) est (presque) toujours négatif, utilisation de l'approximation. D'où la puissance :

```
puissance = 1 - pnorm(upp)
```

Notez l'utilisation de la fonction `pnorm(upp)` pour trouver $\psi(u'')$.

```
plot(delta, puissance, type = "l")
```



De manière exacte  possède une fonction capable de calculer la puissance d'un test de Student. Reproduire le graphique précédant avec la fonction `power.t.test`. Aller voir l'aide pour comprendre les arguments de cette fonction.

```
power.t.test(n = 40, delta = 15, sd = 52, sig.level = 0.05)
```

```
Two-sample t test power calculation
  n = 40
  delta = 15
  sd = 52
  sig.level = 0.05
  power = 0.2464259
  alternative = two.sided
```

NOTE: n is number in *each* group

```
power.t.test(n = 40, delta = 50, sd = 52, sig.level = 0.05)
```

```
Two-sample t test power calculation
  n = 40
  delta = 50
  sd = 52
  sig.level = 0.05
  power = 0.988892
  alternative = two.sided
```

NOTE: n is number in *each* group

```
power.t.test(n = 40, delta = 100, sd = 52, sig.level = 0.05)
```

```
Two-sample t test power calculation
  n = 40
  delta = 100
  sd = 52
  sig.level = 0.05
  power = 1
  alternative = two.sided
```

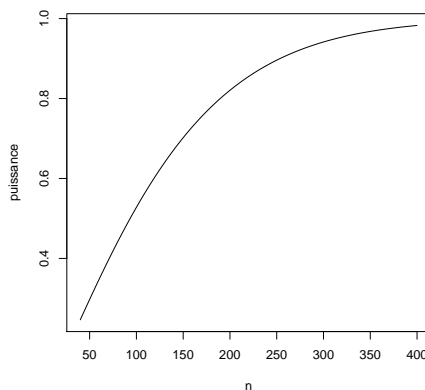
NOTE: n is number in *each* group

Comment varie la puissance si on augmente l'écart entre l'hypothèse nulle et l'alternative choisie? Regarder l'effet du changement des autres paramètres de la fonction (par exemple α).

```
power.t.test(n = 40, delta = 15, sd = 52, sig.level = 0.01)
power.t.test(n = 40, delta = 50, sd = 52, sig.level = 0.01)
power.t.test(n = 40, delta = 100, sd = 52, sig.level = 0.01)
```

Pour tracer le graphique :

```
n = 40:400
puissance = sapply(n, function(x) power.t.test(x, delta = 15, sd = 52,
  sig.level = 0.05)$power)
plot(n, puissance, type = "l")
```



Etudier graphiquement la variation de la fonction puissance en fonction du risque de première espèce, de l'écart type (**sd**) et de la taille des échantillons (**n**).

Si on revient à des problématiques de planification expérimentale, que peut on faire en tant qu'expérimentateur pour augmenter la puissance du test ?

Calculons maintenant la taille minimale de l'échantillon pour travailler avec une puissance $(1 - \beta)$ donnée.

De manière approchée (cf cours)

Calcul de la taille minimale de l'échantillon dans le cas d'un test de comparaison de deux moyennes (approximation pour $\alpha < 0.1$ et $u'' < 0$)

$$1 - \beta = 1 - \psi(u'') \text{ or } u'' < 0 \text{ donc } \psi(u'') = 1 - \psi(|u''|)$$

$$1 - \beta = \psi(|u''|) = \psi\left(\left|u_{1-\frac{\alpha}{2}} - \frac{\delta}{\sigma\sqrt{\frac{2}{n}}}\right|\right)$$

$$1 - \beta = \psi\left(\frac{\delta}{\sigma\sqrt{\frac{2}{n}}} - u_{1-\frac{\alpha}{2}}\right)$$

$$u_{1-\beta} = \frac{\delta}{\sigma\sqrt{\frac{2}{n}}} - u_{1-\frac{\alpha}{2}}$$


En ré-arrangeant les termes on obtient :

$$n = \frac{2\sigma^2(u_{1-\beta} + u_{1-\frac{\alpha}{2}})^2}{\delta^2}$$

```
nmin <- function(alpha, beta, delta, sigma) {
  (2 * sigma^2) * (qnorm(1 - beta) + qnorm(1 - (alpha/2)))^2/delta^2
}
nmin(0.05, 0.1, 15, 52)
[1] 252.5518
nmin(0.05, 0.1, 50, 52)
[1] 22.72966
nmin(0.05, 0.01, 50, 52)
[1] 39.74333
```

De manière exacte

```
power.t.test(delta = 15, sd = 52, sig.level = 0.05, power = 0.9)
power.t.test(delta = 15, sd = 52, sig.level = 0.05, power = 0.99)
power.t.test(delta = 15, sd = 52, sig.level = 0.01, power = 0.99)
power.t.test(delta = 15, sd = 52, sig.level = 0.01, power = 0.9)
power.t.test(delta = 50, sd = 52, sig.level = 0.05, power = 0.99)
```

Jusqu'à présent nous avons travaillé sur les densités 20 et 30. Pour analyser le jeu de données complet, on devrait calculer la puissance d'une ANOVA I. Il existe une fonction dans . Regarder l'aide : `help(■power.anova.test■)` ou `?power.anova.test()`.

Conclusions On peut gagner de la puissance en augmentant α et δ (est-ce intéressant?). En terme de planification expérimentale il peut être plus facile d'augmenter la taille des échantillons. On peut aussi jouer sur l'écart-type si la problématique biologique l'autorise.

6 Effet de la densité et des répliques sur la quantité de sang ingérée par les larves de Triatomines : dispositif expérimental en blocs aléatoires complets

Regardons un peu les données :

```
summary(densite)
10 20 30 40
36 28 44 109

summary(replique)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.00  2.00   3.00  2.77  4.00   4.00

table(densite, replique)
      replique
densite 1  2  3  4
   10  8 10 10  8
   20  8  6  5  9
   30 12  7 14 11
   40 16 21 18 54
```

Le nombre de répétitions pour chacune des combinaisons de modalités est variable. Cette variabilité peut poser des problèmes pour la réalisation d'une ANOVA II. Nous allons illustrer ces difficultés au travers un exercice en travaillant soit sur le jeu de données complet soit en réduisant ce jeu de données afin d'avoir le même nombre de répétitions par combinaison.

6.1 Exercice : ANOVA à 1 et 2 facteurs avec taille d'échantillon constante

L'objectif de cette partie est de tester les effets de la densité et de la réplique sur la quantité de sang ingéré. Cette question nécessite de mettre en oeuvre une ANOVA à deux facteurs (ANOVA II). Dans ce type de modèle l'égalité du nombre de données par modalité est importante. Pour des raisons pédagogiques nous allons réduire le nombre de données par modalité à 5 (cf. `table(densite, replique)` ; `densite = 20` et `replique = 3`) Ainsi nous allons discuter de la puissance de l'ANOVA I avec peu de données, puis nous allons comparer les résultats d'une ANOVA II avec plan équilibré (même nombre de données par modalité) avec ceux pour une analyse en plan non équilibré (sur l'ensemble du jeu de données).

6.1.1 ANOVA I

Echantillonnage de 5 valeurs d'ingestion pour chaque combinaison de modalités (`densite × replique`) afin d'avoir un plan équilibré. Méthode élémentaire : un exemple pour la densité 40 et la réplique 4.

```
x404 = ingestion[densite == "40" & replique == "4"]
help("sample")
sample(x404, 5)
[1] 124.5  0.0 74.0 72.2 67.3

sample(x404, 5)
[1] 20.1 78.6 0.0 63.9 13.4

ex404 = sample(x404, 5)
ex404
```

```
[1] 100.4 49.2 127.2 0.0 134.7
```

Voyons comment généraliser :

En raison de la fluctuations d'échantillonnage illustrée ci-dessus (fonction `sample`), on répète 500 fois l'échantillonnage pour effectuer l'ANOVA I et récupérer les 500 valeurs de la statistique observée (F).

Mais avant de faire 500 fois, regardons déjà ce qui se passe pour trois simulations :

```
w1 <- split(ingestion, densite:replique)
echa5 <- matrix((unlist(lapply(1:3, function(k) {
  unlist(lapply(w1, function(x) {
    sample(x, 5)
  })))
}))), nrow = 80)
edensite <- as.factor(rep(c("10", "20", "30", "40"), rep(20, 4)))
```

Visualisez les objets `w1`, `echa5` et `edensite` que vous venez de créer. `echa5` est un tableau de 80 lignes (16 modalités \times 5 données) et 3 colonnes (i.e. les 3 simulations). L'objet `edensite` contient la densité qui correspond à chacune de ces 80 lignes.

Réalisons l'ANOVA I sur la première simulation (première colonne du tableau `echa5`, i.e. `echa5[,1]`) et voyons comment récupérer la valeur de la probabilité critique associée au test.

```
anova(lm(echa5[, 1] ~ edensite))
Analysis of Variance Table
Response: echa5[, 1]
      Df Sum Sq Mean Sq F value Pr(>F)
edensite  3 27030  9009.9   3.1765 0.02882
Residuals 76 215572  2836.5
names(anova(lm(echa5[, 1] ~ edensite)))
[1] "Df"      "Sum Sq"  "Mean Sq" "F value" "Pr(>F)"
anova(lm(echa5[, 1] ~ edensite))[5]
      Pr(>F)
edensite 0.02882
Residuals
as.numeric(unlist(anova(lm(echa5[, 1] ~ edensite))[5])[1])
[1] 0.02881918
```

La dernière ligne de commande ci-dessus permet de récupérer la valeur numérique de la p-value (probabilité critique).

Les sorties de \mathbb{R} pour la réalisation de 3 simulations simultanément :

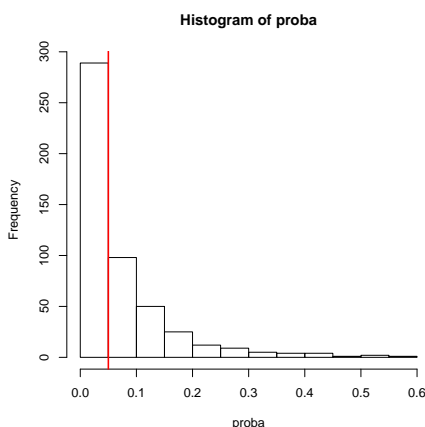
```
lapply(1:3, function(x) anova(lm(echa5[, x] ~ edensite)))
[[1]]
Analysis of Variance Table
Response: echa5[, x]
      Df Sum Sq Mean Sq F value Pr(>F)
edensite  3 27030  9009.9   3.1765 0.02882
Residuals 76 215572  2836.5

[[2]]
Analysis of Variance Table
Response: echa5[, x]
      Df Sum Sq Mean Sq F value Pr(>F)
edensite  3 14705  4901.7   1.7824 0.1576
Residuals 76 209002  2750.0

[[3]]
Analysis of Variance Table
Response: echa5[, x]
      Df Sum Sq Mean Sq F value Pr(>F)
edensite  3 29949  9982.9   3.9563 0.01120
Residuals 76 191769  2523.3
```

Par les commandes qui suivent, on réalise 500 simulations simultanées, on sauvegarde les p-values et on représente l'histogramme de ces 500 probabilités critiques (p-values).

```
echa5 <- matrix((unlist(lapply(1:500, function(k) {
  unlist(lapply(w1, function(x) {
    sample(x, 5)
  })))
}))), nrow = 80)
proba <- as.vector(unlist(lapply(1:500, function(x) as.numeric(unlist(anova(lm(echa5[,
  x] ~ edensite))[5])[1]))))
hist(proba)
abline(v = 0.05, col = "red", lwd = 2)
length(proba[proba < 0.05])
[1] 315
```



Combien de simulations ont donné une probabilité critique inférieure à 5% ? Discuter des résultats en termes de puissance. On vous avait bien dit que cet exemple n'avait qu'une valeur pédagogique !

6.1.2 ANOVA II

Créer un objet `ereplique`. L'objet `ereplique` contient la réplique (de 1 à 4) qui correspond à chacune des 80 lignes du tableau `echa5`. Parmi les 500 simulations (i.e. le jeu de données), on réalise, à titre d'exemple, une ANOVA II sur deux des colonnes (ici, la 1^{re} et la 100^e).

Expliquer la notion d'interaction. (L'interaction entre les deux variables explicatives `edensite` et `ereplique` est représentée par `:"`). On pourrait aussi taper : `anova(lm(echa5[,1] edensite*ereplique))`.

```
ereplique <- as.factor(rep(rep(1:4, rep(5, 4)), 4))
anova(lm(echa5[, 1] ~ edensite + ereplique + edensite:ereplique))
Analysis of Variance Table
Response: echa5[, 1]
          Df Sum Sq Mean Sq F value Pr(>F)
edensite   3  29857   9952.5   3.7280 0.01555
ereplique   3   21976   7325.2   2.7439 0.05026
edensite:ereplique  9  20813   2312.6   0.8663 0.55954
Residuals  64 170859   2669.7
anova(lm(echa5[, 100] ~ edensite + ereplique))
```

```

Analysis of Variance Table
Response: echa5[, 100]
      Df Sum Sq Mean Sq F value Pr(>F)
edensite  3  16081  5360.2   2.0150 0.11929
ereplique 3  30979 10326.3   3.8819 0.01241
Residuals 73 194187  2660.1
  anova(lm(echa5[, 100] ~ ereplique + edensite))
Analysis of Variance Table
Response: echa5[, 100]
      Df Sum Sq Mean Sq F value Pr(>F)
ereplique 3  30979 10326.3   3.8819 0.01241
edensite  3  16081  5360.2   2.0150 0.11929
Residuals 73 194187  2660.1

```

Pourquoi peut-on ôter du modèle l'interaction entre les deux facteurs? Que devient la tableau de décomposition de la variance lorsque l'on change l'ordre d'entrée des facteurs dans l'analyse?

6.2 ANOVA à 2 facteurs avec taille d'échantillon non constante (Fichier de données complet sans sélection de 5 valeurs par bloc)

```

  anova(lm(ingestion ~ densite * replique))
Analysis of Variance Table
Response: ingestion
      Df Sum Sq Mean Sq F value Pr(>F)
densite  3  41021 13673.8   5.6967 0.00092
replique  3  25718  8572.6   3.5715 0.01500
densite:replique  9  28245  3138.3   1.3075 0.23480
Residuals 201 482460  2400.3
  anova(lm(ingestion ~ densite + replique))
Analysis of Variance Table
Response: ingestion
      Df Sum Sq Mean Sq F value Pr(>F)
densite  3  41021 13673.8   5.6226 0.001002
replique  3  25718  8572.6   3.5250 0.015873
Residuals 210 510705  2431.9
  anova(lm(ingestion ~ replique + densite))
Analysis of Variance Table
Response: ingestion
      Df Sum Sq Mean Sq F value Pr(>F)
replique  3  27819  9272.9   3.8130 0.010868
densite  3  38920 12973.5   5.3346 0.001463
Residuals 210 510705  2431.9

```

En retournant au jeu de données complet et son plan déséquilibré, est-il statistiquement intéressant de garder l'interaction entre les 2 variables explicatives? Que se passe-t-il si l'on change l'ordre d'introduction des facteurs dans le modèle?

- ★ Conclure sur l'effet de l'ordre d'introduction des facteurs. Aidez vous de la fiche BS7 "Analyse de la variance" (pages 15-20).
- ★ Discuter de l'intérêt d'un plan équilibré du point de vue statistique. Peut on toujours constituer un plan équilibré dans une expérience?
- ★ Faire les graphiques correspondants à ces analyses. Discuter l'effet replique
- ★ Qu'apporte l'intégration de l'effet bloc dans l'analyse pour l'interprétation biologique? L'apport du facteur replique peut s'estimer par le rapport des carrés moyens des résiduelles du modèle d'anova I et d'anova II.
- ★ Plus généralement qu'apporte l'intégration d'un facteur supplémentaire dans une analyse statistique?

7 Effet de la densité sur la quantité de sang ingérée par les adultes de Triatomines

7.1 Charger les données

```

rhodniusad <- read.table("http://pbil.univ-lyon1.fr/R/donnees/rhodniusad.txt",
  h = T)
names(rhodniusad)
[1] "densite" "replique" "sexe" "ingestion"
neodensite <- as.factor(rhodniusad$densite)
neoreplique <- as.factor(rhodniusad$replique)
neosexse <- as.factor(rhodniusad$sexe)
neoringestion <- rhodniusad$ingestion
table(neodensite, neoreplique, neosexse)
, , neosexse = Hembra
  neoreplique
neodensite 1 2 3 4
           10 6 5 5 3
            20 8 8 10 7
            30 11 9 9 11
            40 14 17 14 14
, , neosexse = Macho
  neoreplique
neodensite 1 2 3 4
           10 4 5 5 2
            20 7 6 9 9
            30 14 12 10 11
            40 14 17 15 15
  tapply(neoringestion, list(neodensite, neoreplique, neosexse), mean)
, , Hembra
      1      2      3      4
10 59.73333 91.62000 105.06000 66.96667
20 88.16250 51.87500 57.95000 72.64286
30 53.10000 75.47778 67.90000 82.98182
40 27.95714 41.44118 35.17143 64.10000
, , Macho
      1      2      3      4
10 71.75000 52.42 57.28000 47.80000
20 68.08571 68.75 42.55556 66.11111
30 50.39286 46.70 69.21000 48.62727
40 28.67143 42.10 36.57333 45.30667
  tapply(neoringestion, list(neodensite, neoreplique, neosexse), var)
, , Hembra
      1      2      3      4
10 1028.0027 559.007 599.593 1118.573
20 1008.3484 2077.548 1578.461 1352.393
30 1626.0560 1052.649 1379.525 1061.976
40 823.2119 1043.470 1121.875 774.623
, , Macho
      1      2      3      4
10 147.49667 1003.1770 415.9970 462.0800
20 42.91810 1411.3750 914.6728 635.4711
30 805.23302 1129.3327 744.1854 545.5502
40 762.80374 762.4962 380.0321 443.0907

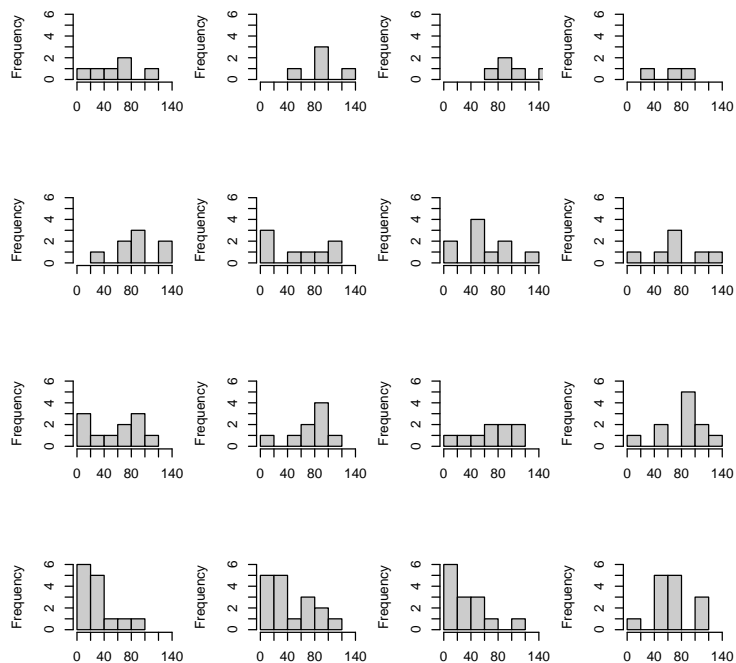
```

Distribution des quantités de sang ingérées par les femelles en fonction des répliques et de la densité de punaises. Pour regarder la distribution des données pour les femelles, on commence par découper le tableau `rhodniusad` en fonction du facteur `sexe`.

```

w4 <- split(rhodniusad, neosexse)
rhodniusadf = w4$Hembra
fdens <- as.factor(rhodniusadf$densite)
frepl <- as.factor(rhodniusadf$replique)
par(mfrow = c(4, 4), mar = c(3, 4, 4, 1))
tapply(rhodniusadf$ingestion, fdens:frepl, hist, col = grey(0.8),
  main = "", xlab = "", xlim = c(0, 140), ylim = c(0, 6))

```

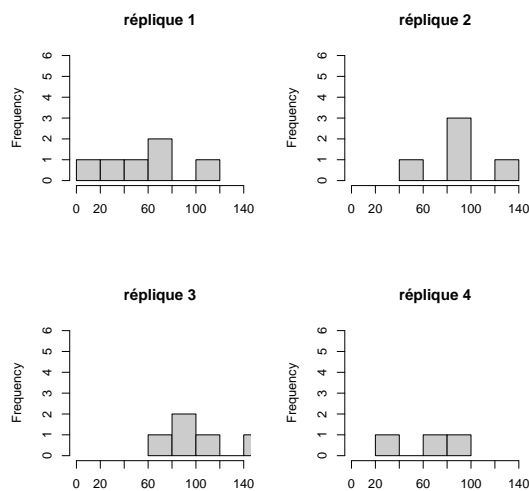


Si l'on souhaite faire les histogrammes un à un, voici les commandes à utiliser :

```

par(mfrow = c(2, 2))
hist(rhodniusadf$ingestion[fdens == "10" & frepl == "1"], col = grey(0.8),
     main = "réplique 1", xlab = "", xlim = c(0, 140), ylim = c(0,
6))
hist(rhodniusadf$ingestion[fdens == "10" & frepl == "2"], col = grey(0.8),
     main = "réplique 2", xlab = "", xlim = c(0, 140), ylim = c(0,
6))
hist(rhodniusadf$ingestion[fdens == "10" & frepl == "3"], col = grey(0.8),
     main = "réplique 3", xlab = "", xlim = c(0, 140), ylim = c(0,
6))
hist(rhodniusadf$ingestion[fdens == "10" & frepl == "4"], col = grey(0.8),
     main = "réplique 4", xlab = "", xlim = c(0, 140), ylim = c(0,
6))

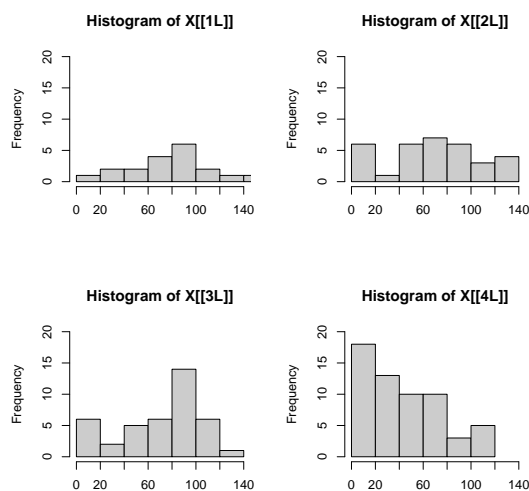
```

... et la même chose pour les densités 20, 30 et 40.

On propose de regrouper les 4 réplices, pour quelles raisons? On obtient la figure suivante :

```
par(mfrow = c(2, 2))
tapply(rhodniusadf$ingestion, fdens, hist, col = grey(0.8), xlab = "",
       xlim = c(0, 140), ylim = c(0, 20))
```



Refaire la même chose pour étudier la distribution des données chez les mâles. Formuler des hypothèses biologiques pour expliquer l'effet de la densité sur la quantité de sang ingérée chez les adultes et en fonction des blocs. Proposer des représentations graphiques.

7.2 Analyse des données (ANOVA III)

```

anova(lm(neoingestion ~ neodensite * neosexex * neoreplique))
Analysis of Variance Table
Response: neoingestion
              Df Sum Sq Mean Sq F value    Pr(>F)
neodensite    3  42228  14075.9  15.4703 2.52e-09
neosexex      1   7911   7911.1    8.6948 0.003466
neoreplique   3   6556   2185.4    2.4019 0.068004
neodensite:neosexex  3   3566   1188.8    1.3066 0.272573
neodensite:neoreplique  9  17896   1988.4    2.1854 0.023266
neosexex:neoreplique  3   3524   1174.7    1.2911 0.277761
neodensite:neosexex:neoreplique  9  11867   1318.5    1.4491 0.167053
Residuals    274 249304    909.9

step(lm(neoingestion ~ neodensite * neosexex * neoreplique))
Start:  AIC=2115.07
neoingestion ~ neodensite * neosexex * neoreplique
              Df Sum of Sq  RSS    AIC
- neodensite:neosexex:neoreplique  9    11867 261170 2111.3
<none>                                249304 2115.1

Step:  AIC=2111.3
neoingestion ~ neodensite + neosexex + neoreplique + neodensite:neosexex +
  neodensite:neoreplique + neosexex:neoreplique
              Df Sum of Sq  RSS    AIC
- neosexex:neoreplique  3   3524.2 264694 2109.4
- neodensite:neosexex  3   4322.8 265493 2110.3
<none>                                261170 2111.3
- neodensite:neoreplique  9  18376.8 279547 2114.1

Step:  AIC=2109.4
neoingestion ~ neodensite + neosexex + neoreplique + neodensite:neosexex +
  neodensite:neoreplique
              Df Sum of Sq  RSS    AIC
- neodensite:neosexex  3   3881.5 268576 2107.8
<none>                                264694 2109.4
- neodensite:neoreplique  9  17895.9 282590 2111.4

Step:  AIC=2107.85
neoingestion ~ neodensite + neosexex + neoreplique + neodensite:neoreplique
              Df Sum of Sq  RSS    AIC
<none>                                268576 2107.8
- neodensite:neoreplique  9  17580.8 286157 2109.3
- neosexex  1    8218.5 276794 2115.1

Call:
lm(formula = neoingestion ~ neodensite + neosexex + neoreplique +
    neodensite:neoreplique)

Coefficients:
(Intercept)          neodensite20          neodensite30
      68.701             14.947             -11.292
neodensite40          neosexexMacho          neoreplique2
     -35.186            -10.402              8.520
neoreplique3          neoreplique4  neodensite20:neoreplique2
      17.670             -5.240             -28.603
neodensite30:neoreplique2  neodensite40:neoreplique2  neodensite20:neoreplique3
     -0.952              4.936             -45.733
neodensite30:neoreplique3  neodensite40:neoreplique3  neodensite20:neoreplique4
     -1.015             -9.909             -3.588
neodensite30:neoreplique4  neodensite40:neoreplique4
      18.836             31.484

anova(lm(neoingestion ~ neodensite + neosexex + neoreplique + neodensite:neoreplique))
Analysis of Variance Table
Response: neoingestion
              Df Sum Sq Mean Sq F value    Pr(>F)
neodensite    3  42228  14075.9  15.1463 3.509e-09
neosexex      1   7911   7911.1    8.5128 0.003804
neoreplique   3   6556   2185.4    2.3516 0.072472
neodensite:neoreplique  9  17581   1953.4    2.1020 0.029358
Residuals    289 268576    929.3

summary(lm(neoingestion ~ neodensite + neosexex + neoreplique + neodensite:neoreplique))
Call:
lm(formula = neoingestion ~ neodensite + neosexex + neoreplique +
    neodensite:neoreplique)
Residuals:

```

```

      Min      1Q  Median      3Q      Max
-74.064 -20.069   1.075  20.810  73.123

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)      68.701     9.741   7.053 1.3e-11
neodensite20      14.947    12.448   1.201 0.23082
neodensite30     -11.292    11.420  -0.989 0.32361
neodensite40     -35.185    11.236  -3.132 0.00192
neosexMacho     -10.402     3.498  -2.974 0.00319
neoreplique2      8.520    13.638   0.625 0.53263
neoreplique3     17.670    13.638   1.296 0.19612
neoreplique4     -5.240    16.697  -0.314 0.75388
neodensite20:neoreplique2 -28.603    17.732  -1.613 0.10783
neodensite30:neoreplique2  -0.952    16.352  -0.058 0.95362
neodensite40:neoreplique2   4.936    15.701   0.314 0.75345
neodensite20:neoreplique3 -45.733    17.229  -2.654 0.00839
neodensite30:neoreplique3  -1.015    16.498  -0.062 0.95098
neodensite40:neoreplique3  -9.909    15.849  -0.625 0.53234
neodensite20:neoreplique4  -3.588    19.974  -0.180 0.85757
neodensite30:neoreplique4  18.836    18.928   0.995 0.32048
neodensite40:neoreplique4  31.484    18.548   1.697 0.09069

```

```

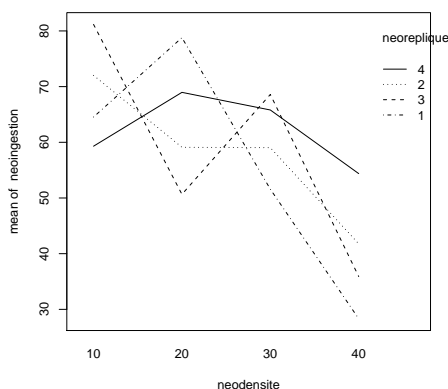
Residual standard error: 30.48 on 289 degrees of freedom
Multiple R-squared: 0.2166, Adjusted R-squared: 0.1733
F-statistic: 4.995 on 16 and 289 DF, p-value: 4.048e-09

```

```

help("interaction.plot")
interaction.plot(neodensite, neoreplique, neoringestion)

```



Références

- [1] J.D. Gillett. Natural selection and feeding speed in a blood-sucking insect. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 167(1008) :316–329, 1967.
- [2] P.A. Langley. Effect of ligaturing on puparium formation in larva of tsetse fly glossina morsitans westwood. *Nature*, 214(5086) :389–390, 1967.
- [3] C.J. Schofield. Density regulation of domestic populations of triatoma infestans in brazil. *Trans. R Soc. Trop. Med. Hyg.*, 74(6) :761–769, 1980.
- [4] C.J. Schofield. Population-dynamics and control of triatoma-infestans. *Annales De La Societe Belge De Medecine Tropicale*, 65 :149–164, 1985.