

# **TP – Formation Statistique de base**

## **Parties 4 et 5**

09-10/03/2020

Nathalie Vialaneix, Sandrine Laguerre

# **Partie 4 : Statistiques inférentielles (tests)**

# **Partie 4.1 : Variables quantitatives**

## **TO DO**

- **Trouver la moyenne de C14**

R Commander

Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide



Données : alldata

Script R R Markdown

```
rownames(alldata) <-  
alldata$Row.names <-  
MergedDataset <- me  
rownames(MergedData  
MergedDataset$Row.n  
alldata <- merge(alldata, lipid, all=TRU  
rownames(alldata) <- alldata$Row.names  
alldata$Row.names <- NULL
```

Résumés ▶

Tables de contingence ▶

Moyennes ▶

Proportions ▶

Variances ▶

Tests non paramétriques ▶

Analyse multivariée ▶

Ajustement de modèles ▶

Jeu de données actif

Statistiques descriptives...

Distributions de fréquences...

Dénombrer les observations manquantes

Tableau de statistiques...

Matrice de corrélations...

Test de corrélation...

Test de normalité...

Transformer vers la loi normale...

Données Statistiques

Variables (une ou plusieurs)

C14.0

C16.0

C16.1n.7

C16.1n.9

C16SR

C18.0

Résumer par groupes...



Aide



Réinitialiser



OK



Annuler



Appliquer

```
Rcmdr> library(abind, pos=17)
```

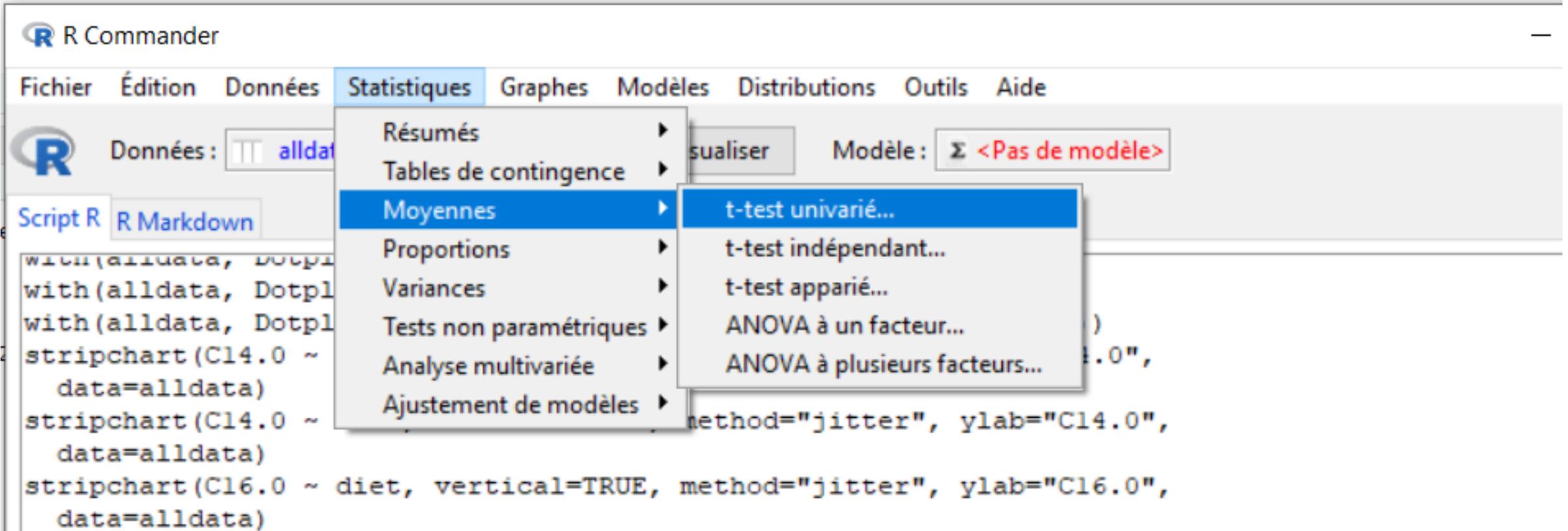
```
Rcmdr> library(e1071, pos=18)
```

```
Rcmdr> numsummary(alldata[, "c14.0", drop=FALSE], statistics=c("mean", "sd", "IQR",  
Rcmdr+ "quantiles"), quantiles=c(0,.25,.5,.75,1))  
  mean      sd    IQR   0%  25%  50%   75% 100%  n  
0.763 0.8005677 0.2825 0.22 0.35 0.42 0.6325 3.24 40
```

```
> |
```

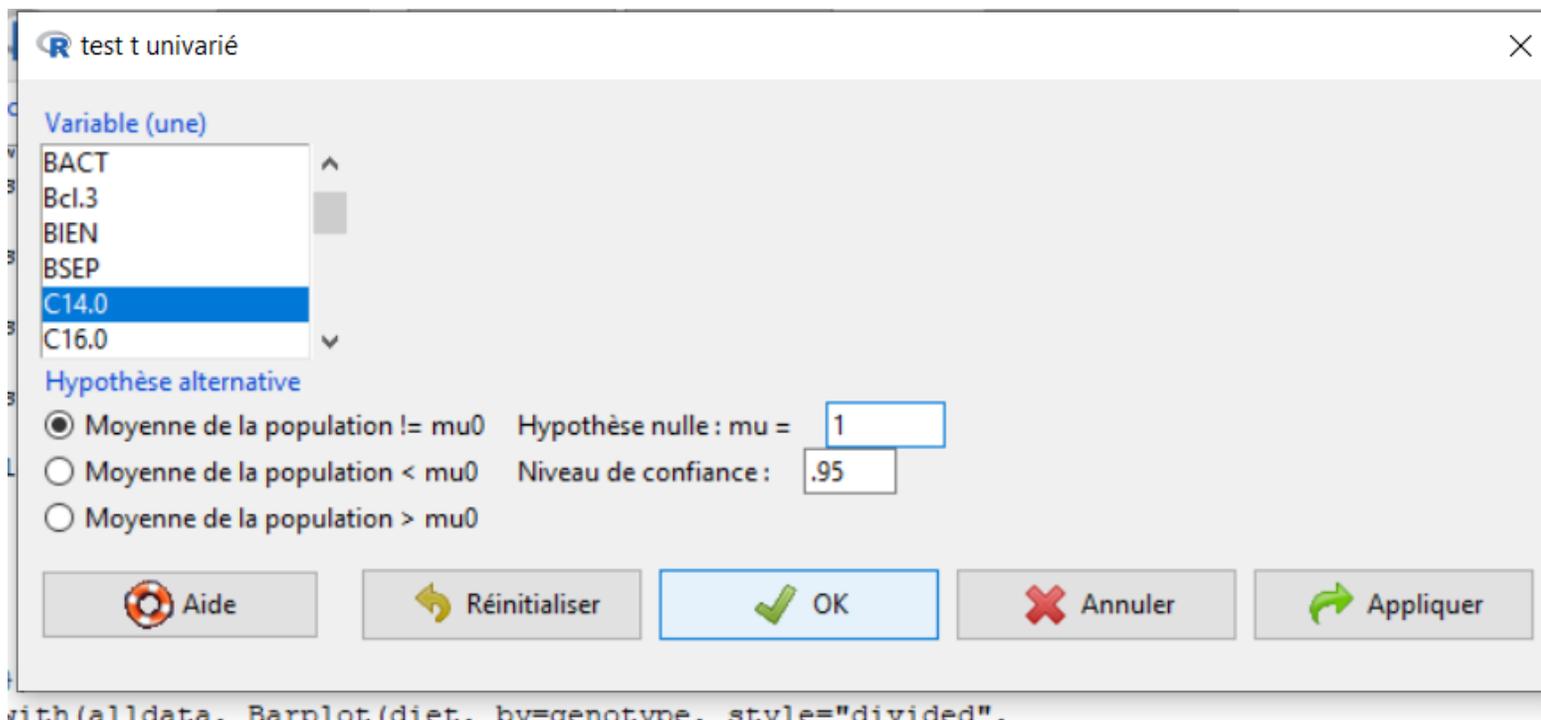
## Intervalle de confiance d'une variable quantitative

La moyenne de C14 est-elle significativement différente de 1?



The image shows the R Commander interface. The 'Statistiques' menu is open, and the 'Moyennes' option is selected. The 't-test univarié...' option is highlighted. The background shows the R Commander window with the 'Données' field set to 'alldata' and the 'Modèle' field set to ' $\Sigma$  <Pas de modèle>'. The console window shows the following R code:

```
with(alldata, Dotplot(C14.0 ~ diet, data=alldata))
with(alldata, Dotplot(C16.0 ~ diet, data=alldata))
stripchart(C14.0 ~ diet, vertical=TRUE, method="jitter", ylab="C14.0", data=alldata)
stripchart(C16.0 ~ diet, vertical=TRUE, method="jitter", ylab="C16.0", data=alldata)
```



```
with(alldata, Barplot(diet, by=genotype, style="divided").
```

```
Rcmdr> with(alldata, (t.test(C14.0, alternative='two.sided', mu=1,  
Rcmdr+   conf.level=.95)))
```

One sample t-test

```
data: C14.0  
t = -1.8723, df = 39, p-value = 0.06867  
alternative hypothesis: true mean is not equal to 1  
95 percent confidence interval:  
 0.506966 1.019034  
sample estimates:  
mean of x  
 0.763
```

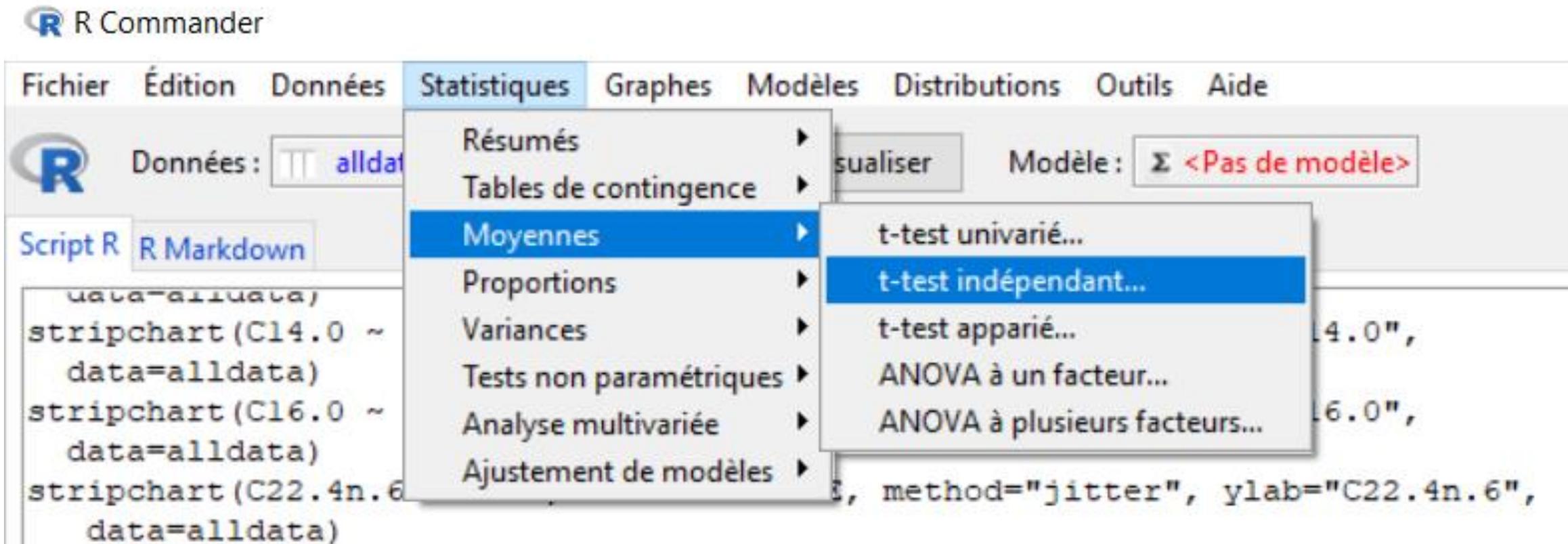
L'intervalle de confiance à 95% contient la valeur 1

P-value NS à 5% => on ne peut pas rejeter l'hypothèse nulle d'égalité de la moyenne de C14 à 1

# **Comparaison de moyenne pour 2 échantillons indépendants**

Les moyennes de C14 diffèrent t-elles selon le genotype?

Test paramétrique : t-test



R Commander

Fichier Édition Données **Statistiques** Graphes Modèles Distributions Outils Aide

Données:  Visualiser Modèle:

Script R R Markdown

- Résumés
- Tables de contingence
- Moyennes**
  - t-test univarié...
  - t-test indépendant...**
  - t-test apparié...
  - ANOVA à un facteur...
  - ANOVA à plusieurs facteurs...
- Proportions
- Variances
- Tests non paramétriques
- Analyse multivariée
- Ajustement de modèles

```
data=alldata)
stripchart(C14.0 ~
  data=alldata)
stripchart(C16.0 ~
  data=alldata)
stripchart(C22.4n.6
  data=alldata)
```

..., method="jitter", ylab="C22.4n.6",

Données Options

Groupes (un)

genotype

Variable réponse (une)

C14.0

C16.0

C16.1n.7

C16.1n.9

C16SR

C18.0



Aide



Réinitialiser



OK



Annuler



Appliquer

Données Options

Différence : ppar - wt

Hypothèse alternative

Bilatéral

Différence < 0

Différence > 0

Niveau de confiance

.95

Variances égales ?

Oui

Non



Aide



Réinitialiser



OK



Annuler



Appliquer

```
Rcmdr> t.test(C14.0~genotype, alternative='two.sided', conf.level=.95,  
Rcmdr+   var.equal=FALSE, data=alldata)
```

```
Welch Two Sample t-test
```

```
data: C14.0 by genotype
```

```
t = 1.8126, df = 26.386, p-value = 0.08129
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-0.05941223  0.95141223
```

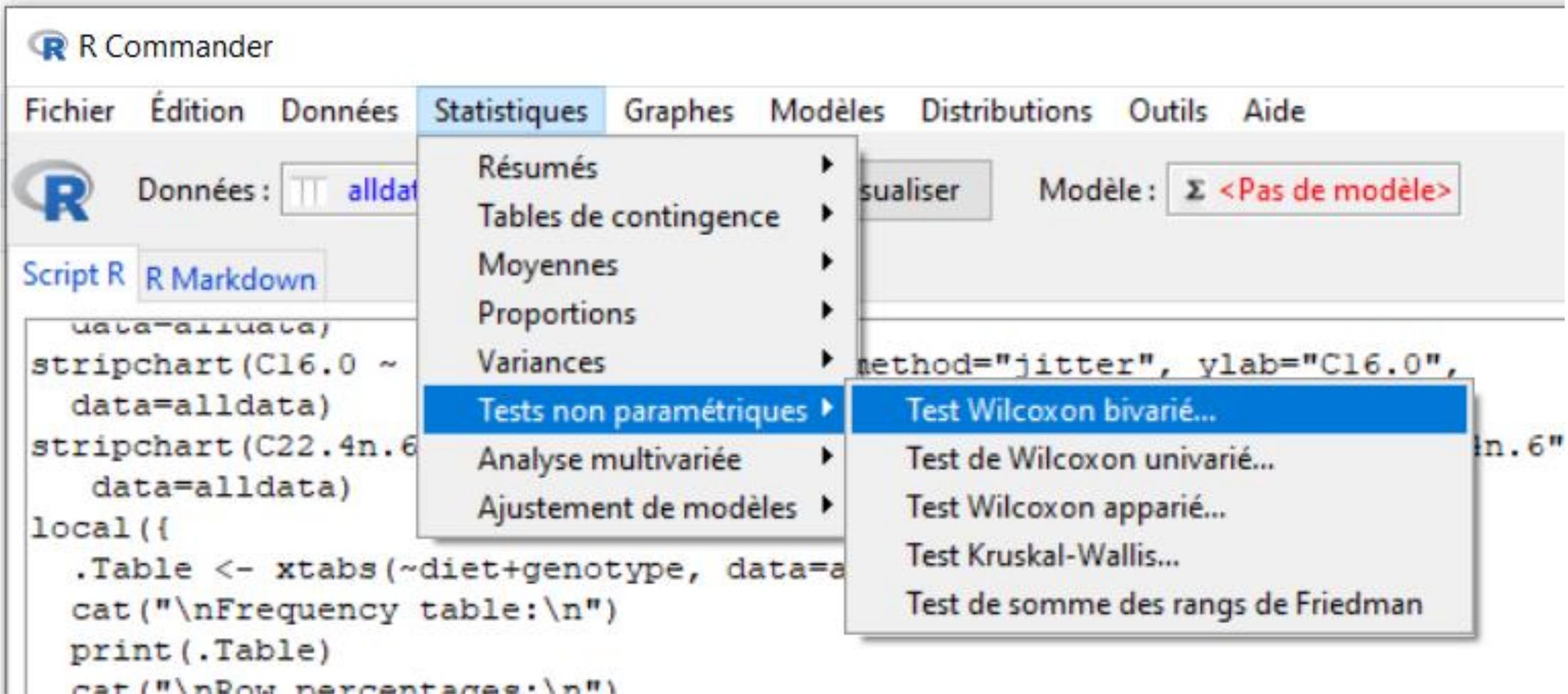
```
sample estimates:
```

mean in group ppar	mean in group wt
0.986	0.540

P-value NS à 5% => on ne peut pas rejeter l'hypothèse nulle d'égalité des moyennes de C14 selon le genotype

Les médianes de C14 diffèrent t-elles selon le genotype?

Test non paramétrique : test de Wilcoxon bivarié



R Commander

Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide

Données :  Visualiser Modèle :

Script R R Markdown

```
data=alldata)
stripchart(C16.0 ~
  data=alldata)
stripchart(C22.4n.6
  data=alldata)
local({
  .Table <- xtabs(~diet+genotype, data=a
  cat("\nFrequency table:\n")
  print(.Table)
  cat("\nRow percentages:\n")
```

method="jitter", ylab="C16.0", in.6"

- Résumés
- Tables de contingence
- Moyennes
- Proportions
- Variances
- Tests non paramétriques**
  - Test Wilcoxon bivarié...**
  - Test de Wilcoxon univarié...
  - Test Wilcoxon apparié...
  - Test Kruskal-Wallis...
  - Test de somme des rangs de Friedman
- Analyse multivariée
- Ajustement de modèles

Test de Wilcoxon bivarié

Données Options

Groupes (un)  
genotype

Variable réponse (une)  
BACT  
Bcl.3  
BIEN  
BSEP  
C14.0  
C16.0

Aide Réinitialiser OK Annuler Appliquer

Test de Wilcoxon bivarié

Données Options

Différence : ppar - wt

Hypothèse alternative Type de test

Bilatéral  Par défaut

Différence < 0  Exact

Différence > 0  Approximation normale

Approximation normale avec correction de continuité

Aide Réinitialiser OK Annuler Appliquer

```
Rcmdr> with(alldata, tapply(C14.0, genotype, median, na.rm=TRUE))
ppar    wt
0.57 0.35

Rcmdr> wilcox.test(C14.0 ~ genotype, alternative="two.sided", data=alldata)

    wilcoxon rank sum test with continuity correction

data:  C14.0 by genotype
w = 326, p-value = 0.0006808
alternative hypothesis: true location shift is not equal to 0
```

P-value très significative => on rejette l'hypothèse nulle d'égalité des 2 médianes

Les tests paramétrique et non paramétrique ne sont pas cohérent => on va tester les conditions d'application du t-test

## **TO DO**

- **Représenter la distribution en fonction du génotype**

Données:  

Script R R Markdown

```
data=alldata)
stripchart(C22.4n.6 ~ diet,
  data=alldata)
local({
  .Table <- xtabs(~diet+geno
  cat("\nFrequency table:\n"
  print(.Table)
  cat("\nRow percentages:\n"
  print(rowPercents(.Table))
})
with(alldata, Barplot(diet,
  legend.pos="topright", xla
with(alldata, (t.test(C14.0,
  conf.level=.95)))
with(alldata, (t.test(C14.0,
  conf.level=.95)))
t.test(C14.0~genotype, alter
  var.equal=FALSE, data=alld
with(alldata, tapply(C14.0,
wilcox.test(C14.0 ~ genotype
```

PaLETTE de couleurs...

Graphe indexé...

Graphe en points (dot plot)

Histogramme...

Graphe d'une variable numérique discrète...

Estimation de densité...

Graphe tiges et feuilles...

Boîte de dispersion...

Graphe quantile-quantile...

Boîte à moustaches pour symétrie...

Nuage de points...

Matrice de nuages de points...

Graphe en lignes...

Graphe XY conditionnel...

Graphe des moyennes...

Graphe en bande...

Graphe en barres...

Graphe en camembert...

Graphe 3D

Enregistrer le graphe dans un fichier...

le modèle&gt;

ab="C22.4n.6

lldata)

R Histogramme

Données Options

Variable (une)

- BACT
- Bcl.3
- BIEN
- BSEP
- C14.0
- C16.0

Grappe par : diet

R Groupes

Variable de regroupement (une)

- diet
- genotype

OK Annuler

R Histogramme

Données Options

Options du graphe

Nombre de classes : <auto>

Echelle des axes

- Fréquences
- Pourcentages
- Densité

Étiquettes du graphe

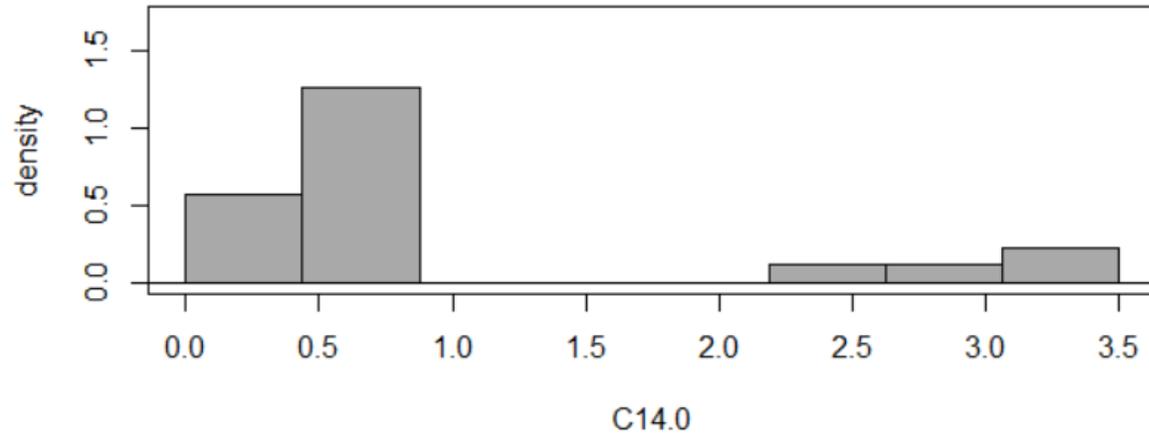
libellé de l'axe X <auto>

libellé de l'axe Y <auto>

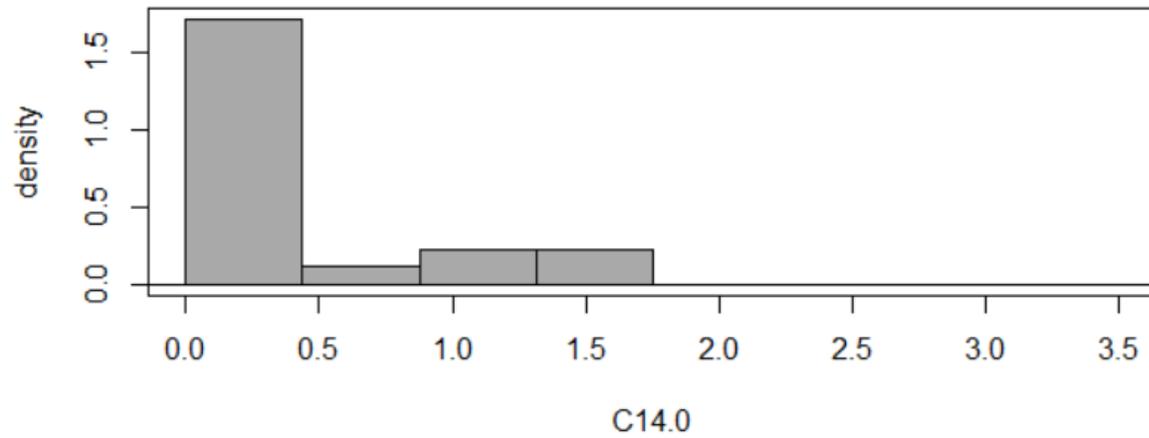
Titre du graphe <auto>

Aide Réinitialiser OK Annuler Appliquer

**genotype = ppar**



**genotype = wt**



Normalité?

# Test de normalité

R Commander

Fichier Édition Données **Statistiques** Graphes Modèles Distributions Outils Aide

Données : alldata

Script R R Markdown

```
.Table <- xtabs(  
cat("\nFrequency  
print(.Table)  
cat("\nRow percen  
print(rowPercents  
)  
})
```

W: R Test de normalité

Variable (une)

- BACT
- Bcl.3
- BIEN
- BSEP
- C14.0**
- C16.0

Test de normalité

Shapiro-Wilk     Anderson-Darling

Cramer-von Mises     Lilliefors (Kolmogorov-Smirnov)

Shapiro-Francia     Khi-2 de Pearson

Nombre de classes pour le Khi-2 de Pearson

Test by groups...

Aide Réinitialiser OK Annuler Appliquer

```
Rcmdr> normalityTest(C14.0 ~ genotype, test="shapiro.test", data=alldata)
```

```
-----
```

```
genotype = ppar
```

```
shapiro-wilk normality test
```

```
data: C14.0
```

```
W = 0.60338, p-value = 0.000003193
```

```
-----
```

```
genotype = wt
```

```
shapiro-wilk normality test
```

```
data: C14.0
```

```
W = 0.64591, p-value = 0.000009152
```

```
-----
```

```
p-values adjusted by the Holm method:
```

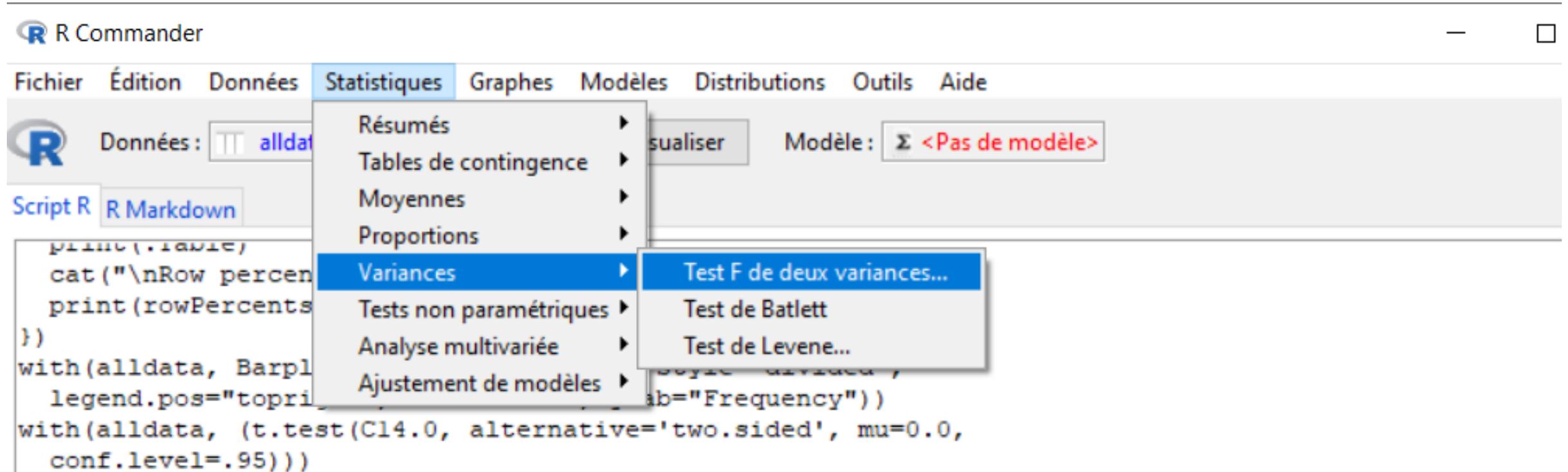
```
unadjusted adjusted
```

```
ppar 0.0000031926 0.0000063852
```

```
wt 0.0000091523 0.0000091523
```

P-value très significative => on rejette l'hypothèse nulle de normalité

## Test d'égalité des variances dans les 2 groupes comparés



R Commander

Fichier Édition Données **Statistiques** Graphes Modèles Distributions Outils Aide

Données:  Actualiser Modèle:  $\Sigma$  <Pas de modèle>

Script R R Markdown

```
print(.table)
cat("\nRow percent
print(rowPercents
})
with(alldata, Barplot
legend.pos="topri
with(alldata, (t.test(C14.0, alternative='two.sided', mu=0.0,
conf.level=.95)))
```

- Résumés
- Tables de contingence
- Moyennes
- Proportions
- Variations**
  - Test F de deux variances...**
  - Test de Batlett
  - Test de Levene...
- Tests non paramétriques
- Analyse multivariée
- Ajustement de modèles

Données Options

Groupes (un)

genotype

Variable réponse (une)

BACT

Bcl.3

BIEN

BSEP

C14.0

C16.0



Aide



Réinitialiser



OK



Annuler



Appliquer

Données Options

Ratio : ppar / wt

Hypothèse alternative

 Bilatéral Ratio < 1 Ratio > 1

Niveau de confiance : .95



Aide



Réinitialiser



OK



Annuler



Appliquer

```
Rcmdr> var.test(C14.0 ~ genotype, alternative='two.sided', conf.level=.95,  
Rcmdr+ data=alldata)
```

F test to compare two variances

```
data: C14.0 by genotype
```

```
F = 4.9423, num df = 19, denom df = 19, p-value = 0.001043
```

```
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
```

```
 1.95623 12.48652
```

```
sample estimates:
```

```
ratio of variances
```

```
 4.942318
```

P-value significative => on rejette l'hypothèse nulle d'égalité des 2 variances

# **Comparaison de moyenne pour 2 échantillons appariés**

## **TO DO**

- **Quelles sont les moyennes de ACAT1 et ACAT2?**
- **Que signifie un test de comparaison de moyenne pour échantillons appariés entre ces 2 variables?**

Données : 

Script R

R Markdown

```
legendu.pos= copy1  
with(alldata, (t.test(C14.0 ~ genotype, data=alldata,  
  conf.level=.95)))  
with(alldata, (t.test(C14.0 ~ genotype, data=alldata,  
  conf.level=.95)))  
t.test(C14.0~genotype, data=alldata, var.equal=FALSE, data=alldata)  
with(alldata, tapply(C14.0, genotype, median, na.rm=TRUE))  
wilcox.test(C14.0 ~ genotype, alternative="two.sided", data=alldata)
```

Résumés ▶

Tables de contingence ▶

Moyennes ▶

Proportions ▶

Variances ▶

Tests non paramétriques ▶

Analyse multivariée ▶

Ajustement de modèles ▶

Jeu de données actif

Statistiques descriptives...

Distributions de fréquences...

Dénombrer les observations manquantes

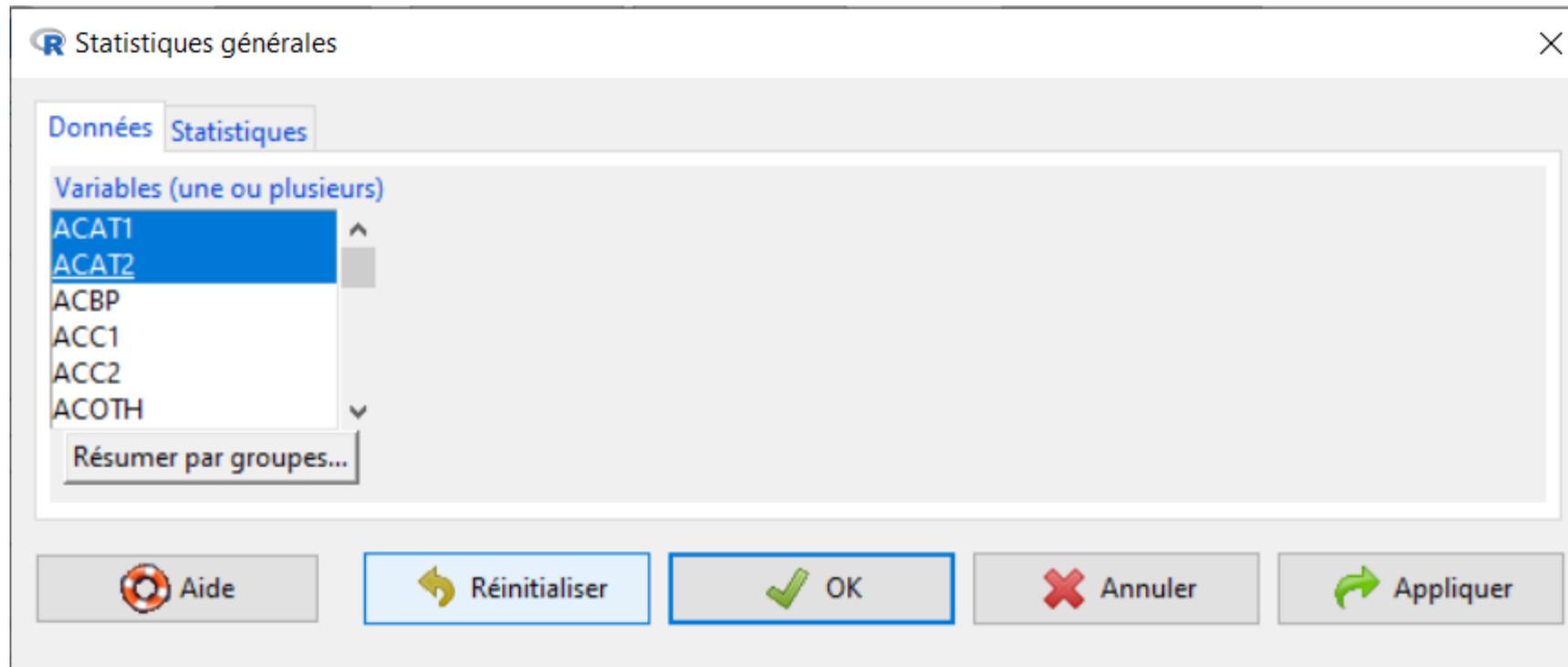
Tableau de statistiques...

Matrice de corrélations...

Test de corrélation...

Test de normalité...

Transformer vers la loi normale...



```
Rcmdr> numSummary(alldata[,c("ACAT1", "ACAT2"), drop=FALSE], statistics=c("mean",  
Rcmdr+ "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))  
      mean      sd  IQR  0%  25%  50%  75% 100%  n  
ACAT1 -0.65525 0.05325254 0.070 -0.75 -0.69 -0.660 -0.620 -0.52 40  
ACAT2 -0.76675 0.17303309 0.235 -1.10 -0.88 -0.795 -0.645 -0.39 40  
> |
```

## T test apparié (test paramétrique)

R Commander

Fichier Édition Données **Statistiques** Graphes Modèles Distributions Outils Aide

Données :  Actualiser Modèle :  $\Sigma$  <Pas de modèle>

Script R R Markdown

- Résumés
- Tables de contingence
- Moyennes**
  - t-test univarié...
  - t-test indépendant...
  - t-test apparié...**
  - ANOVA à un facteur...
  - ANOVA à plusieurs facteurs...
- Proportions
- Variances
- Tests non paramétriques
- Analyse multivariée
- Ajustement de modèles

```
with(alldata, Barplot(
  legend.pos="topright",
with(alldata, (t.test(
  conf.level=.95)))
with(alldata, (t.test(
  conf.level=.95)))
```

'two.sided', mu=1,

test t apparié

Données Options

Première variable (une)	Seconde variable (une)
ACAT1	ACAT1
ACAT2	ACAT2
ACBP	ACBP
ACC1	ACC1
ACC2	ACC2
ACOTH	ACOTH

test t apparié

Données Options

Hypothèse alternative Niveau de confiance

Bilatérale

Différence < 0

Différence > 0

.95

Aide Réinitialiser OK Annuler Appliquer

```
Rcmdr> with(alldata, (t.test(ACAT1, ACAT2, alternative='two.sided', conf.level=
Rcmdr+     paired=TRUE)))
```

Paired t-test

```
data: ACAT1 and ACAT2
```

```
t = 3.769, df = 39, p-value = 0.0005427
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
 0.0516612 0.1713388
```

```
sample estimates:
```

```
mean of the differences
```

```
      0.1115
```

Les 2 moyennes étaient proches (-0,65 et -0,77) et pourtant,  
P-value significative => on rejette l'hypothèse nulle d'égalité des 2 moyennes

# Test de Wilcoxon apparié (non paramétrique)

R Commander

Fichier Édition Données **Statistiques** Graphes Modèles Distributions Outils Aide

Données :  Actualiser Modèle :

Script R R Markdown

```
conf.level=.95)))  
with(alldata, (t.test  
  conf.level=.95)))  
t.test(C14.0~genoty  
  var.equal=FALSE,  
with(alldata, tappl  
wilcox.test(C14.0 ~ genotype, alternativ  
with(alldata, Hist(C14.0, groups=genotyp  
  breaks="Sturges", col="darkgray"))  
with(alldata, Hist(C14.0, groups=genotype, scale="density",
```

- Résumés
- Tables de contingence
- Moyennes
- Proportions
- Variances
- Tests non paramétriques**
  - Test Wilcoxon bivarié...
  - Test de Wilcoxon univarié...
  - Test Wilcoxon apparié...**
  - Test Kruskal-Wallis...
  - Test de somme des rangs de Friedman
- Analyse multivariée
- Ajustement de modèles

Test Wilcoxon apparié



Données Options

Première variable (une)

ACAT1  
ACAT2  
ACBP  
ACC1  
ACC2  
ACOTH

Seconde variable (une)

ACAT1  
ACAT2  
ACBP  
ACC1  
ACC2  
ACOTH

Test Wilcoxon apparié



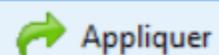
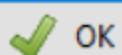
Données Options

Hypothèse alternative

- Bilatéral  
 Différence < 0  
 Différence > 0

Type de test

- Par défaut  
 Exact  
 Approximation normale  
 Approximation normale avec correction de continuité



```
Rcmdr> with(alldata, wilcox.test(ACAT1, ACAT2, alternative='two.sided',  
Rcmdr+   paired=TRUE))
```

```
    wilcoxon signed rank test with continuity correction
```

```
data:  ACAT1 and ACAT2
```

```
V = 623, p-value = 0.001174
```

```
alternative hypothesis: true location shift is not equal to 0
```

Les 2 moyennes étaient proches (-0,65 et -0,77) et pourtant, pour le test non paramétrique également p-value significative => on rejette l'hypothèse nulle d'égalité des 2 moyennes

## **TO DO**

- **Vérifier la normalité de ces 2 variables**

# Histogramme

R Commander

Fichier Édition Données Statistiques **Graphes** Modèles Distributions Outils Aide

Données :

Script R R Markdown

```
U.test(C14.0~genotype, data=alldata,
var.equal=FALSE, data=alldata)
with(alldata, tapply(C14.0,
wilcox.test(C14.0 ~ genotype
with(alldata, Hist(C14.0, gr
breaks="Sturges", col="dar
with(alldata, Hist(C14.0, gr
breaks="Sturges". col="dar
```

Palette de couleurs...

Graphe indexé...

Graphe en points (dot plot)

**Histogramme...**

Graphe d'une variable numérique discrète...

Estimation de densité...

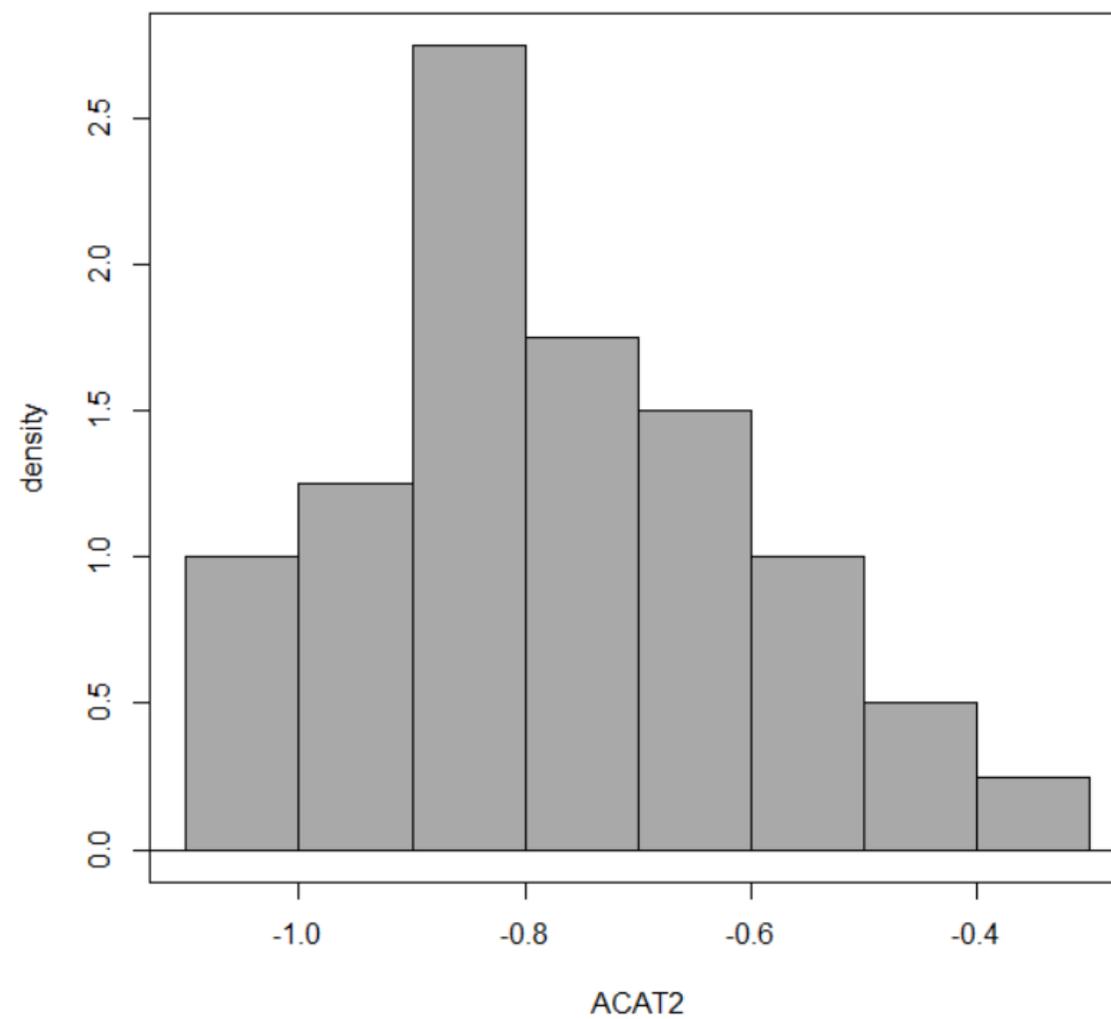
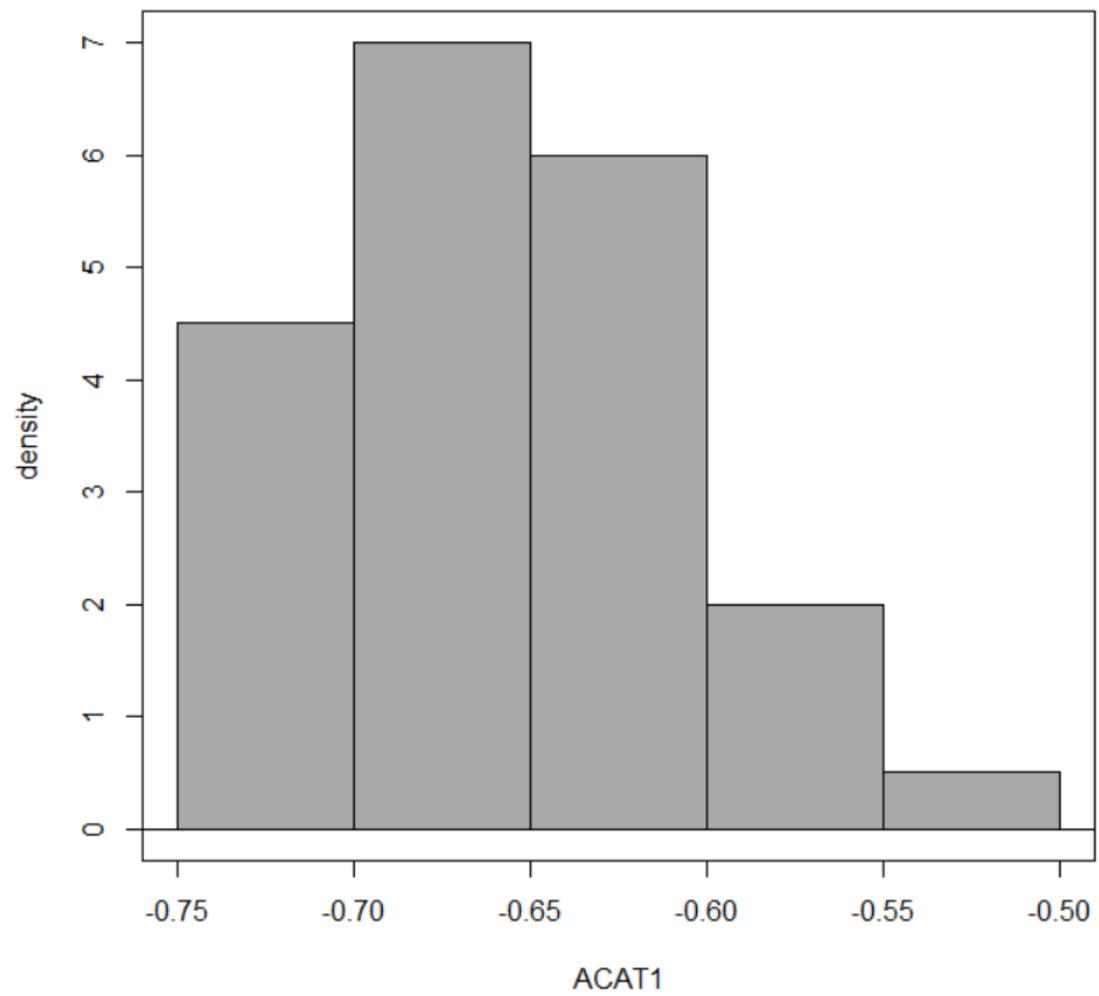
Graphe tiges et feuilles...

Boîte de dispersion...

Graphe quantile-quantile...

le modèle>

ldata)





Données :

Script R R Markdown

```
with(alldata, tapply(C14.0, genotype, var, na.rm=TRUE))
wilcox.test(C14.0 ~ genotype, data=alldata)
with(alldata, Hist(C14.0, breaks="Sturges", col="red", lty="n", las=1))
with(alldata, Hist(C14.0, breaks="Sturges", col="red", lty="n", las=1))
normalityTest(~C14.0, test="shapiro.test", data=alldata)
normalityTest(C14.0 ~ genotype, test="shapiro.test", data=alldata)
with(alldata, tapply(C14.0, genotype, var, na.rm=TRUE))
```

Résumés ▶

Tables de contingence ▶

Moyennes ▶

Proportions ▶

Variances ▶

Tests non paramétriques ▶

Analyse multivariée ▶

Ajustement de modèles ▶

Jeu de données actif

Statistiques descriptives...

Distributions de fréquences...

Dénombrer les observations manquantes

Tableau de statistiques...

Matrice de corrélations...

Test de corrélation...

Test de normalité...

Transformer vers la loi normale...

## Variable (une)

ACAT1  
ACAT2  
ACBP  
ACC1  
ACC2  
ACOTH

## Test de normalité

- Shapiro-Wilk       Anderson-Darling  
 Cramer-von Mises       Lilliefors (Kolmogorov-Smirnov)  
 Shapiro-Francia       Khi-2 de Pearson

Nombre de classes  
pour le Khi-2 de Pearson

```
Rcmdr> with(alldata, hist(ACAT1, scale="density", breaks="sturges",  
Rcmdr+   col="darkgray"))
```

```
Rcmdr> normalityTest(~ACAT1, test="shapiro.test", data=alldata)
```

```
shapiro-wilk normality test
```

```
data: ACAT1
```

```
W = 0.96444, p-value = 0.2368
```

p-value NS => on ne peut pas rejeter l'hypothèse nulle de normalité de ACAT1



```
Rcmdr> normalityTest(~ACAT2, test="shapiro.test", data=alldata)
```

```
shapiro-wilk normality test
```

```
data: ACAT2
```

```
W = 0.97672, p-value = 0.5696
```

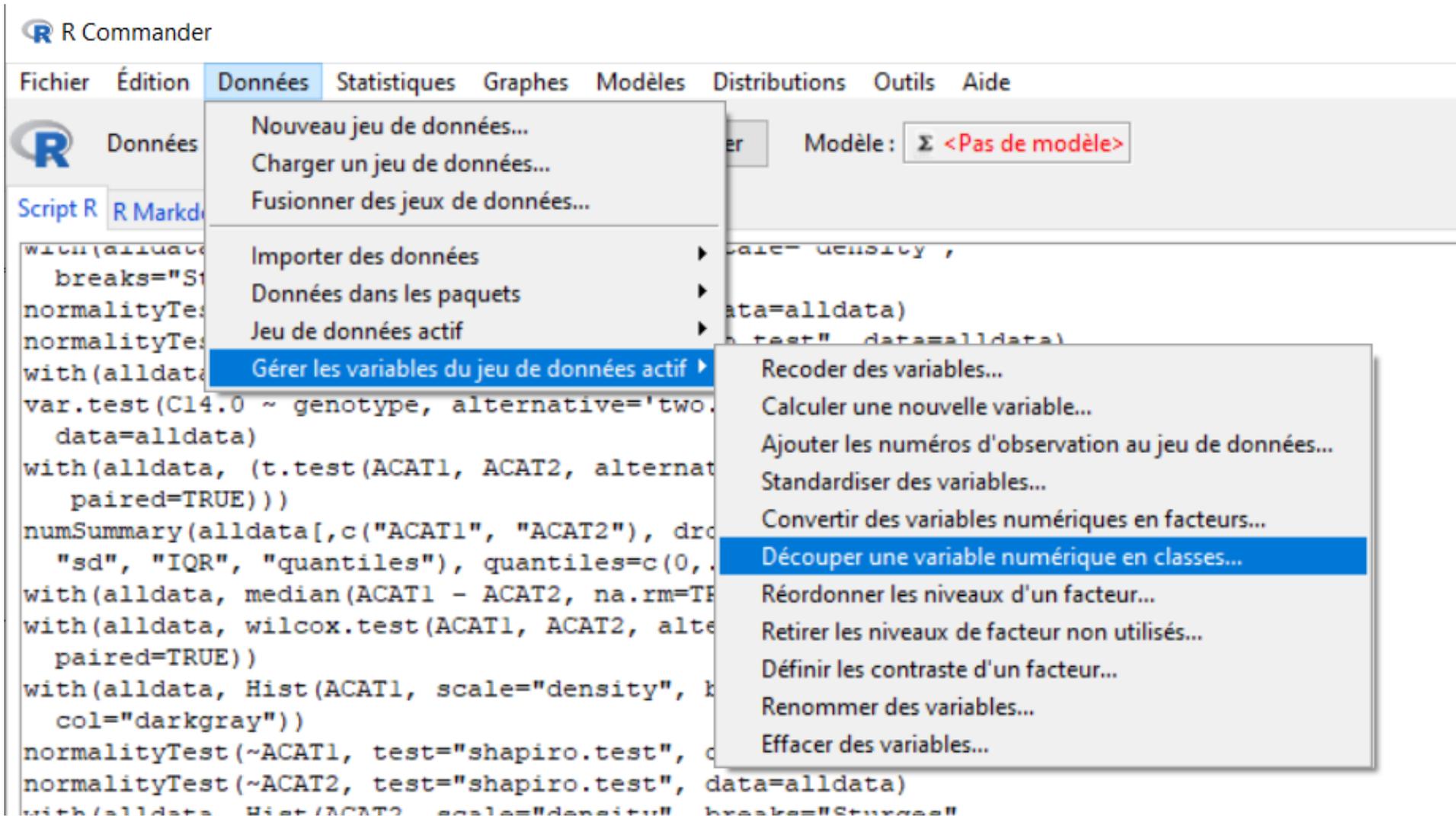
p-value NS => on ne peut pas rejeter l'hypothèse nulle de normalité de ACAT1

## **Partie 4.2 : Variables qualitatives**

# Test d'indépendance entre 2 variables qualitatives

Est-ce qu'être au dessus ou au dessous de la médiane pour C14 est corrélé à diet?

Discrétisation de la variable C14 par rapport à la médiane



Variable à découper (une)

BACT  
Bcl.3  
BIEN  
BSEP  
C14.0  
C16.0

Nom de la nouvelle variable

C14.0.bin

2

Nombre de classes :

Noms de niveaux

- Noms  
 Nombres  
 Etendues

Méthode de découpage

- Classes de taille égale  
 Classes de même effectif  
 Coupures de classes naturelles (avec k-means)



Aide



Réinitialiser

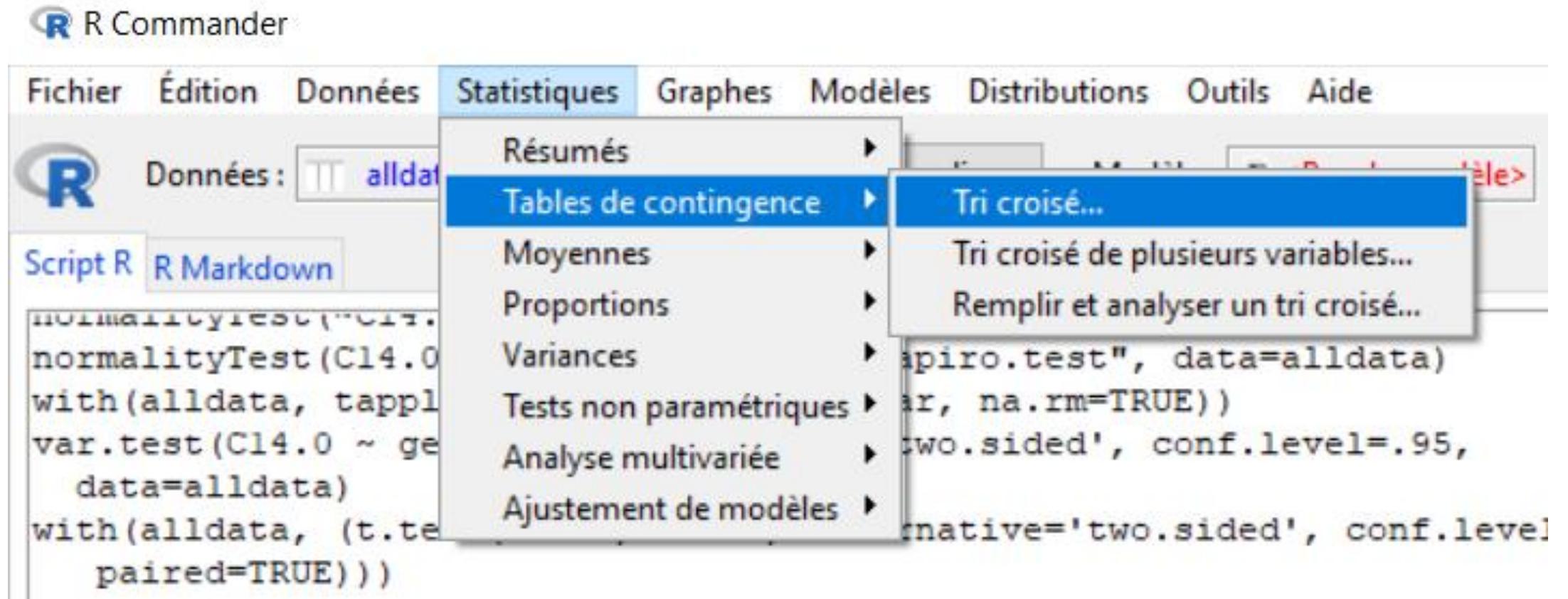


OK



Annuler

## Test d'indépendance du Chi-2



R Commander

Fichier Édition Données **Statistiques** Graphes Modèles Distributions Outils Aide

Données: alldata

Script R R Markdown

normalityTest("C14.0", data=alldata)  
with(alldata, tapply(C14.0, ge, normalityTest))  
var.test(C14.0 ~ ge, data=alldata)  
with(alldata, (t.test(C14.0 ~ ge, data=alldata, paired=TRUE)))

Résumés  
**Tables de contingence**  
Moyennes  
Proportions  
Variances  
Tests non paramétriques  
Analyse multivariée  
Ajustement de modèles

Tri croisé...  
Tri croisé de plusieurs variables...  
Remplir et analyser un tri croisé...

apiro.test", data=alldata)  
ar, na.rm=TRUE))  
two.sided', conf.level=.95,  
native='two.sided', conf.level

Données Statistiques

Variable en ligne (une)

C14.0.bin  
diet  
genotype

Variable en colonne (une)

C14.0.bin  
diet  
genotype

Expression de sélection

&lt;tous les cas valides&gt;

Données Statistiques

Calculer les pourcentages

- Pourcentages des lignes  Pourcentage des colonnes  
 Pourcentages du total  Pas de pourcentages

Test d'hypothèses

- Test Chi-deux d'indépendance  Composants de la statistique du Chi-deux  
 Imprimer les fréquences attendues  Test exact de Fisher



Aide



Réinitialiser



OK



Annuler



Appliquer

Frequency table:

	diet				
c14.0.bin	coc	fish	lin	ref	sun
[0.22,0.42]	0	7	4	3	6
(0.42,3.24]	8	1	4	5	2

Column percentages:

	diet				
c14.0.bin	coc	fish	lin	ref	sun
[0.22,0.42]	0	87.5	50	37.5	75
(0.42,3.24]	100	12.5	50	62.5	25
Total	100	100.0	100	100.0	100
Count	8	8.0	8	8.0	8

Pearson's Chi-squared test

data: .Table

X-squared = 15, df = 4, p-value = 0.004701

Expected counts:

	diet				
c14.0.bin	coc	fish	lin	ref	sun
[0.22,0.42]	4	4	4	4	4
(0.42,3.24]	4	4	4	4	4

Chi-square components:

	diet				
c14.0.bin	coc	fish	lin	ref	sun
[0.22,0.42]	4.00	2.25	0.00	0.25	1.00
(0.42,3.24]	4.00	2.25	0.00	0.25	1.00

RCmdrMsg: [27] AVIS:

RCmdrMsg+ 10 les fréquences attendues sont inférieures à 5

Les conditions d'application du test du Chi2 d'indépendance ne sont pas remplies

# Test exact de Fisher

R Table de contingence à double entrée ×

Données Statistiques

Variable en ligne (une) Variable en colonne (une)

C14.0.bin C14.0.bin

diet diet

genotype genotype

Expression de sélection

<tous les cas valides>

< >

R Table de contingence à double entrée ×

Données Statistiques

Calculer les pourcentages

Pourcentages des lignes  Pourcentage des colonnes

Pourcentages du total  Pas de pourcentages

Test d'hypothèses

Test Chi-deux d'indépendance  Composants de la statistique du Chi-deux

Imprimer les fréquences attendues  Test exact de Fisher

Aide Réinitialiser **OK** Annuler Appliquer

Frequency table:

	diet				
c14.0.bin	coc	fish	lin	ref	sun
[0.22,0.42]	0	7	4	3	6
(0.42,3.24]	8	1	4	5	2

Column percentages:

	diet				
c14.0.bin	coc	fish	lin	ref	sun
[0.22,0.42]	0	87.5	50	37.5	75
(0.42,3.24]	100	12.5	50	62.5	25
Total	100	100.0	100	100.0	100
Count	8	8.0	8	8.0	8

Fisher's Exact Test for Count Data

data: .Table

p-value = 0.003737

alternative hypothesis: two.sided

P-value significative => on rejette l'hypothèse nulle d'indépendance des 2 distributions

# **Partie 5 : Modèle linéaire**

**Partie 5.1 : Régression linéaire et  
corrélation entre 2 variables  
quantitatives**

# Matrice de corrélation

R Commander

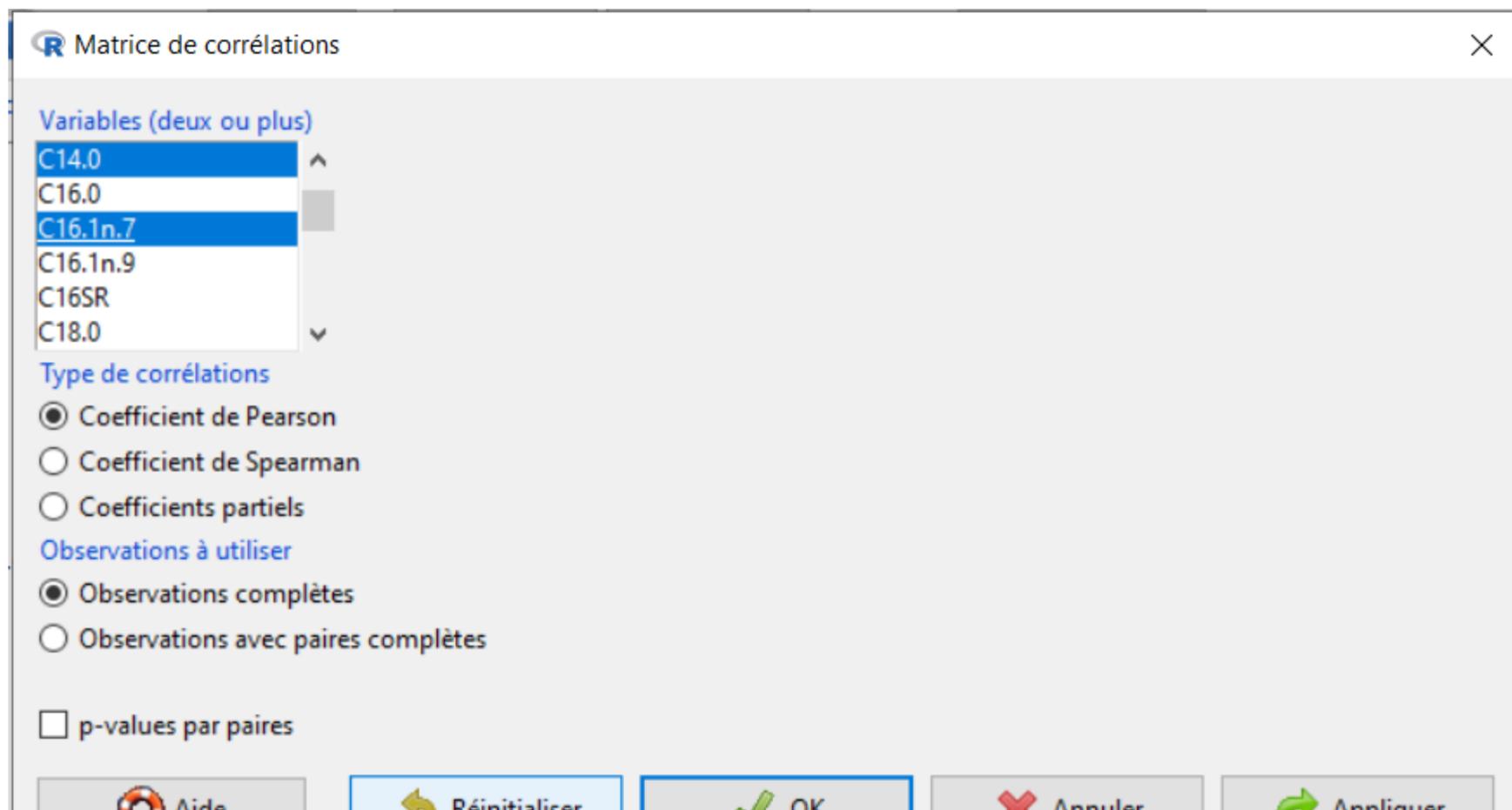
Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide

Données: alldat

Script R R Markdown

```
.Table <- table(x, y)
cat("\nFrequency")
print(.Table)
cat("\nColumn percents")
print(colPercents(.Table))
.Test <- chisq.test(.Table)
print(.Test)
cat("\nExpected counts:\n")
print(.Test$expected)
```

- Résumés
  - Jeu de données actif
  - Statistiques descriptives...
  - Distributions de fréquences...
  - Dénombrer les observations manquantes
  - Tableau de statistiques...
  - Matrice de corrélations...**
  - Test de corrélation...
  - Test de normalité...
  - Transformer vers la loi normale...
- Tables de contingence
- Moyennes
- Proportions
- Variances
- Tests non paramétriques
- Analyse multivariée
- Ajustement de modèles



```
Rcmdr> cor(alldata[,c("c14.0", "c16.1n.7")], use="complete")
```

```
      c14.0 c16.1n.7
c14.0  1.000000 0.971969
c16.1n.7 0.971969 1.000000
```

# Test de corrélation

R Commander

Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide



Données :

Script R R Markdown

```
cat( "\nChi-square test\n")
print(colPercents)
.Test <- chisq.test(x, y)
print(.Test)
cat( "\nExpected counts\n")
print(.Test$expected)
cat( "\nChi-square components:\n")
print(round(.Test$residuals^2, 2))
```

Résumés ▶

Tables de contingence ▶

Moyennes ▶

Proportions ▶

Variances ▶

Tests non paramétriques ▶

Analyse multivariée ▶

Ajustement de modèles ▶

Jeu de données actif

Statistiques descriptives...

Distributions de fréquences...

Dénombrer les observations manquantes

Tableau de statistiques...

Matrice de corrélations...

Test de corrélation...

Test de normalité...

Transformer vers la loi normale...

```
Rcmdr> with(alldata, cor.test(C14.0, C16.1n.7, alternative="two.sided",  
Rcmdr+   method="pearson"))
```

Pearson's product-moment correlation

```
data: C14.0 and C16.1n.7
```

```
t = 25.484, df = 38, p-value < 2.2e-16
```

```
alternative hypothesis: true correlation is not equal to 0
```

```
95 percent confidence interval:
```

```
 0.9472722 0.9851862
```

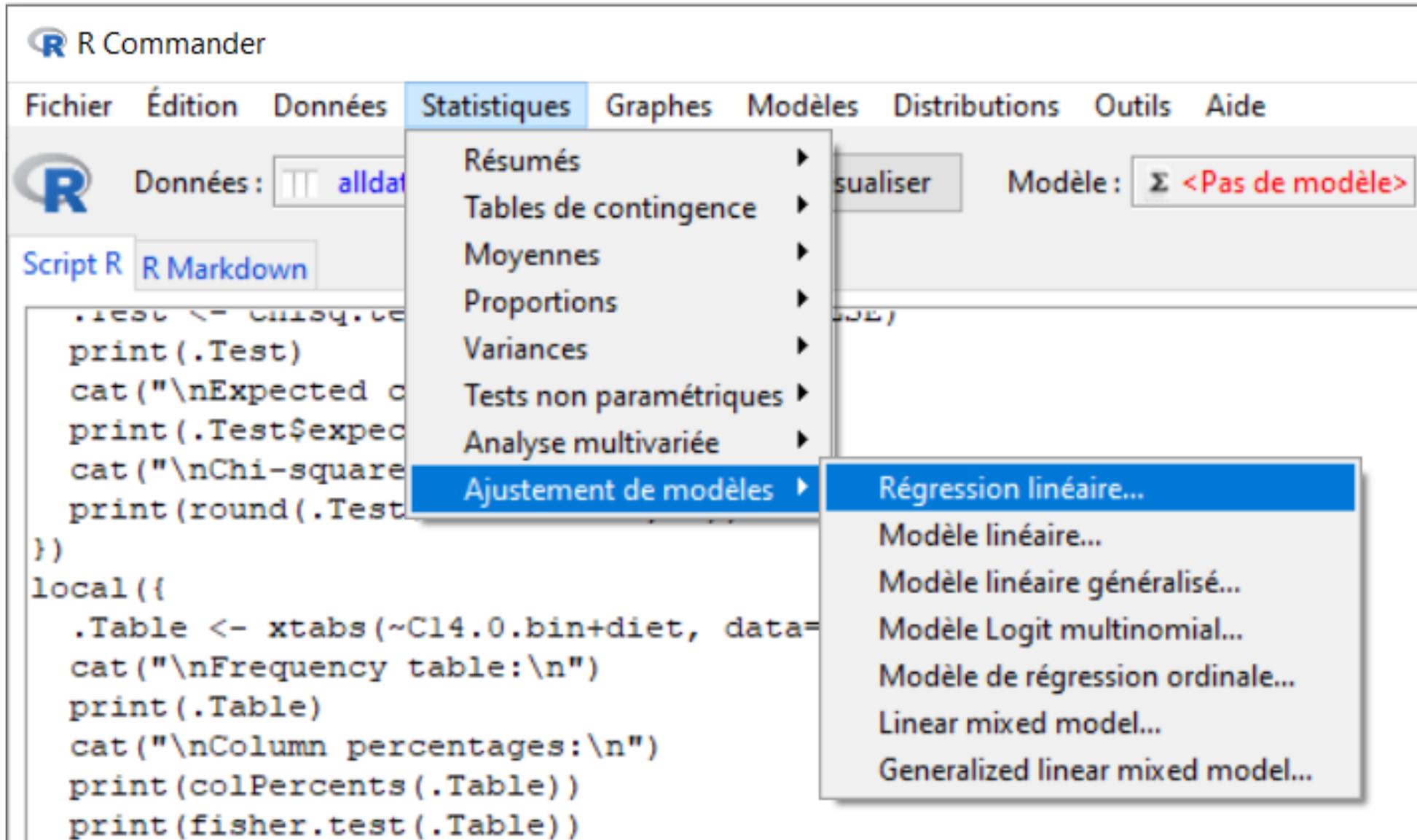
```
sample estimates:
```

```
cor
```

```
0.971969
```

P-value très significative => on rejette l'hypothèse nulle de corrélation = 0

# Régression linéaire



R Commander

Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide

Données :  Actualiser Modèle :  $\Sigma$  <Pas de modèle>

Script R R Markdown

```
.Test <- chiSq.test  
print(.Test)  
cat("\nExpected c  
print(.Test$expec  
cat("\nChi-square  
print(round(.Test  
)  
})  
local({  
  .Table <- xtabs(~C14.0.bin+diet, data=  
  cat("\nFrequency table:\n")  
  print(.Table)  
  cat("\nColumn percentages:\n")  
  print(colPercents(.Table))  
  print(fisher.test(.Table))
```

Résumés

Tables de contingence

Moyennes

Proportions

Variances

Tests non paramétriques

Analyse multivariée

Ajustement de modèles

Régression linéaire...

Modèle linéaire...

Modèle linéaire généralisé...

Modèle Logit multinomial...

Modèle de régression ordinale...

Linear mixed model...

Generalized linear mixed model...

Entrez un nom pour le modèle 

Variable réponse (une)

BACT  
Bcl.3  
BIEN  
BSEP  
C14.0  
C16.0

Variables explicatives (une ou plus)

C14.0  
C16.0  
C16.1n.7  
C16.1n.9  
C16SR  
C18.0

Expression de sélection



Aide



Réinitialiser



OK



Annuler



Appliquer

```
Rcmdr> RegModel.1 <- lm(c14.0~c16.1n.7, data=alldata)
```

```
Rcmdr> summary(RegModel.1)
```

```
call:
```

```
lm(formula = c14.0 ~ c16.1n.7, data = alldata)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-0.41401	-0.08533	-0.01660	0.09894	0.53002

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-0.38889	0.05433	-7.158	0.0000000151	***
c16.1n.7	0.26068	0.01023	25.484	< 2e-16	***

```
---
```

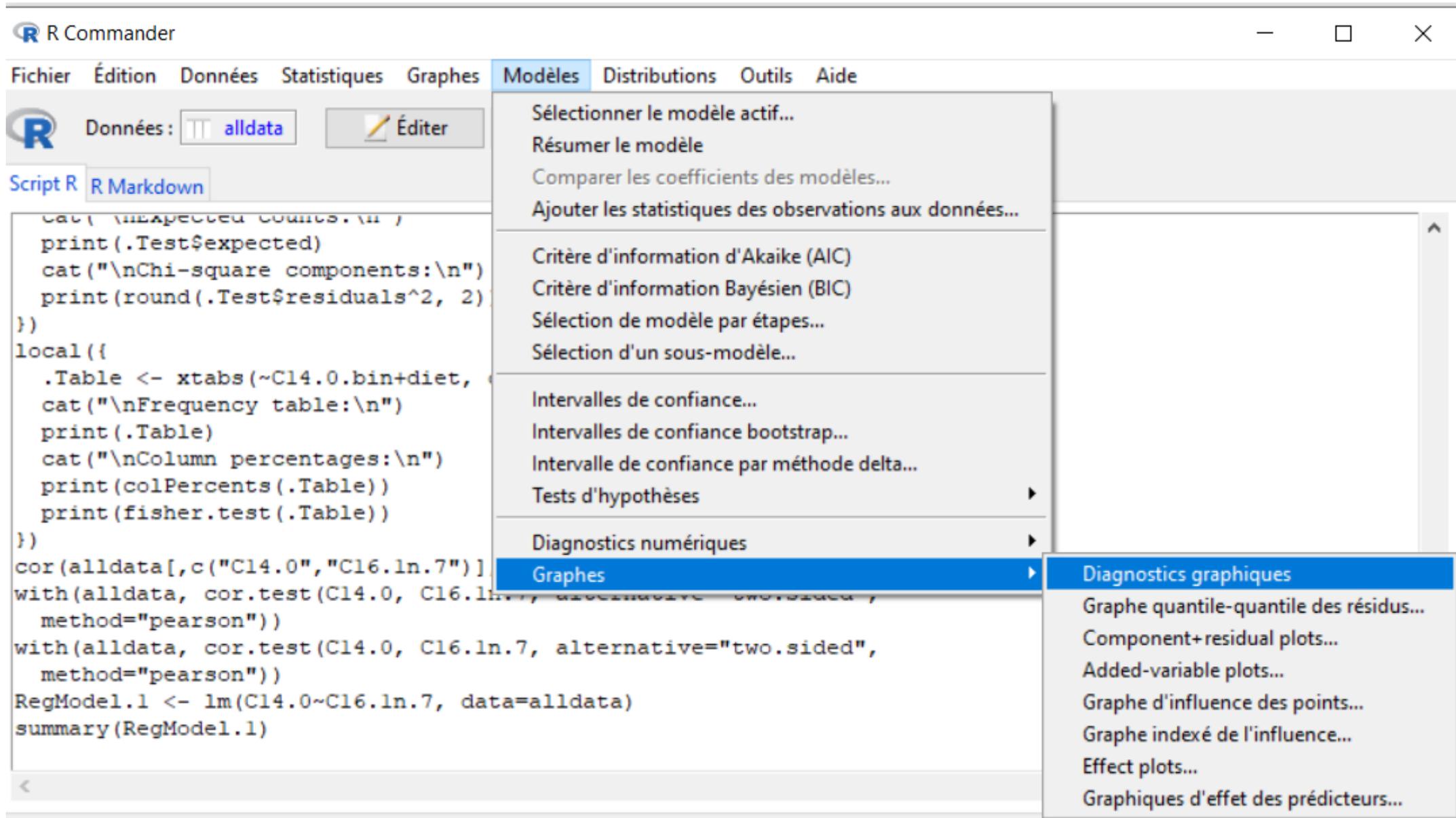
```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1907 on 38 degrees of freedom
```

```
Multiple R-squared:  0.9447,    Adjusted R-squared:  0.9433
```

```
F-statistic: 649.5 on 1 and 38 DF,  p-value: < 2.2e-16
```

# Graphes de diagnostic du modèle linéaire



R Commander

Fichier Édition Données Statistiques Graphes **Modèles** Distributions Outils Aide

Données: alldata Éditer

Script R R Markdown

```
cat("\nExpected counts:\n")
print(.Test$expected)
cat("\nChi-square components:\n")
print(round(.Test$residuals^2, 2))
})
local({
  .Table <- xtabs(~C14.0.bin+diet,
cat("\nFrequency table:\n")
print(.Table)
cat("\nColumn percentages:\n")
print(colPercents(.Table))
print(fisher.test(.Table))
})
cor(alldata[,c("C14.0", "C16.ln.7")])
with(alldata, cor.test(C14.0, C16.ln.7,
method="pearson"))
with(alldata, cor.test(C14.0, C16.ln.7, alternative="two.sided",
method="pearson"))
RegModel.1 <- lm(C14.0~C16.ln.7, data=alldata)
summary(RegModel.1)
```

Sélectionner le modèle actif...  
Résumer le modèle  
Comparer les coefficients des modèles...  
Ajouter les statistiques des observations aux données...

Critère d'information d'Akaike (AIC)  
Critère d'information Bayésien (BIC)  
Sélection de modèle par étapes...  
Sélection d'un sous-modèle...

Intervalles de confiance...  
Intervalles de confiance bootstrap...  
Intervalle de confiance par méthode delta...  
Tests d'hypothèses ▶

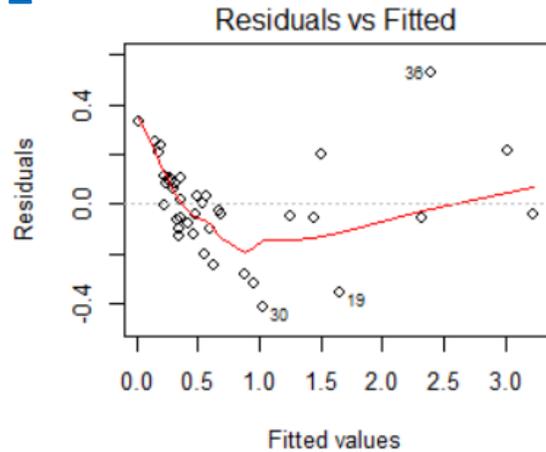
Diagnostics numériques ▶

**Graphes ▶**

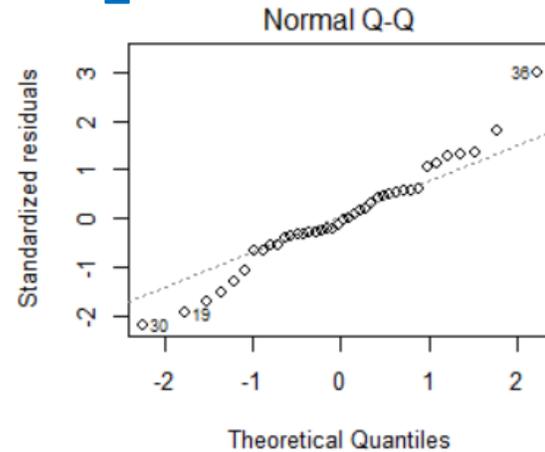
**Diagnostics graphiques**  
Graphe quantile-quantile des résidus...  
Component+residual plots...  
Added-variable plots...  
Graphe d'influence des points...  
Graphe indexé de l'influence...  
Effect plots...  
Graphiques d'effet des prédicteurs...

lm(C14.0 ~ C16.1n.7)

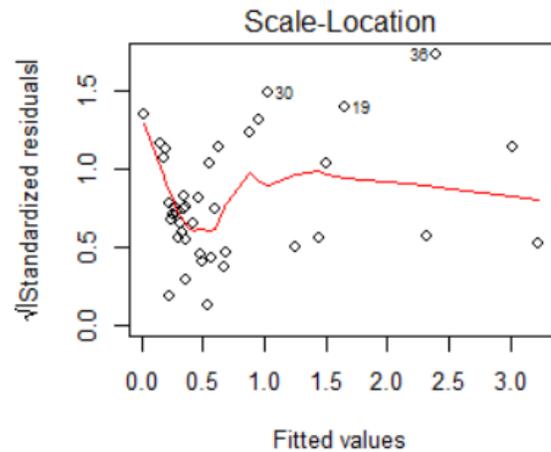
1



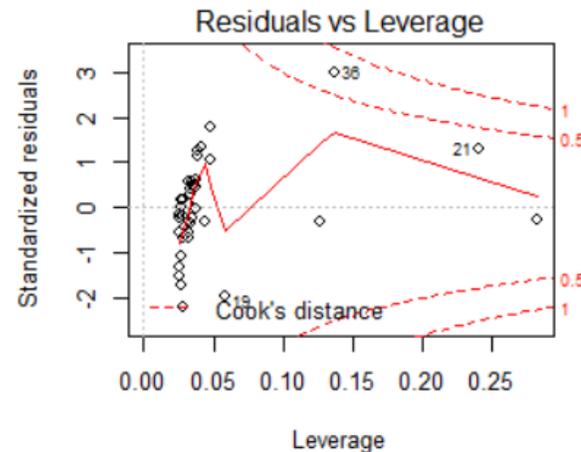
2



3



4



- 1) Les résidus doivent être distribués aléatoirement (pas de tendance à prendre des valeurs différentes selon  $Y$ , la dispersion doit être constante)
- 2) Normalité? : quantiles distribution empirique versus quantiles loi normale (doit être proche d'une droite)
- 3) Homoscédasticité
- 4) Points d'influence

**Partie 5.2 : ANOVA, relation entre une  
variable qualitative et une variable  
quantitative**

# ANOVA à 1 facteur

R Commander

Fichier Édition Données **Statistiques** Graphes Modèles Distributions Outils Aide

Données :  Visualiser Modèle :

Script R R Markdown

```
print(summary(.test))
})
local({
  .Table <- xtabs(~
cat("\nFrequency
print(.Table)
cat("\nColumn percentages:\n")
```

- Résumés
- Tables de contingence
- Moyennes**
  - t-test univarié...
  - t-test indépendant...
  - t-test apparié...
  - ANOVA à un facteur...**
  - ANOVA à plusieurs facteurs...
- Proportions
- Variations
- Tests non paramétriques
- Analyse multivariée
- Ajustement de modèles

 Analyse de la variance à un facteur

Entrez un nom pour le modèle

Groupes (un)

C14.0.bin  
diet  
genotype

Variable réponse (une)

ACAT1  
ACAT2  
ACBP  
ACC1  
ACC2  
ACOTH

- Comparaisons multiples des moyennes  
 Test F de Welch sans supposer l'égalité des variances



Aide



Réinitialiser



OK

## Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = c14.0 ~ diet, data = alldata)

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t )	
fish - coc == 0	-1.82625	0.20064	-9.102	<0.00001	***
lin - coc == 0	-1.69500	0.20064	-8.448	<0.00001	***
ref - coc == 0	-1.66375	0.20064	-8.292	<0.00001	***
sun - coc == 0	-1.75000	0.20064	-8.722	<0.00001	***
lin - fish == 0	0.13125	0.20064	0.654	0.965	
ref - fish == 0	0.16250	0.20064	0.810	0.926	
sun - fish == 0	0.07625	0.20064	0.380	0.995	
ref - lin == 0	0.03125	0.20064	0.156	1.000	
sun - lin == 0	-0.05500	0.20064	-0.274	0.999	
sun - ref == 0	-0.08625	0.20064	-0.430	0.993	

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
 (Adjusted p values reported -- single-step method)

## Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = c14.0 ~ diet, data = alldata)

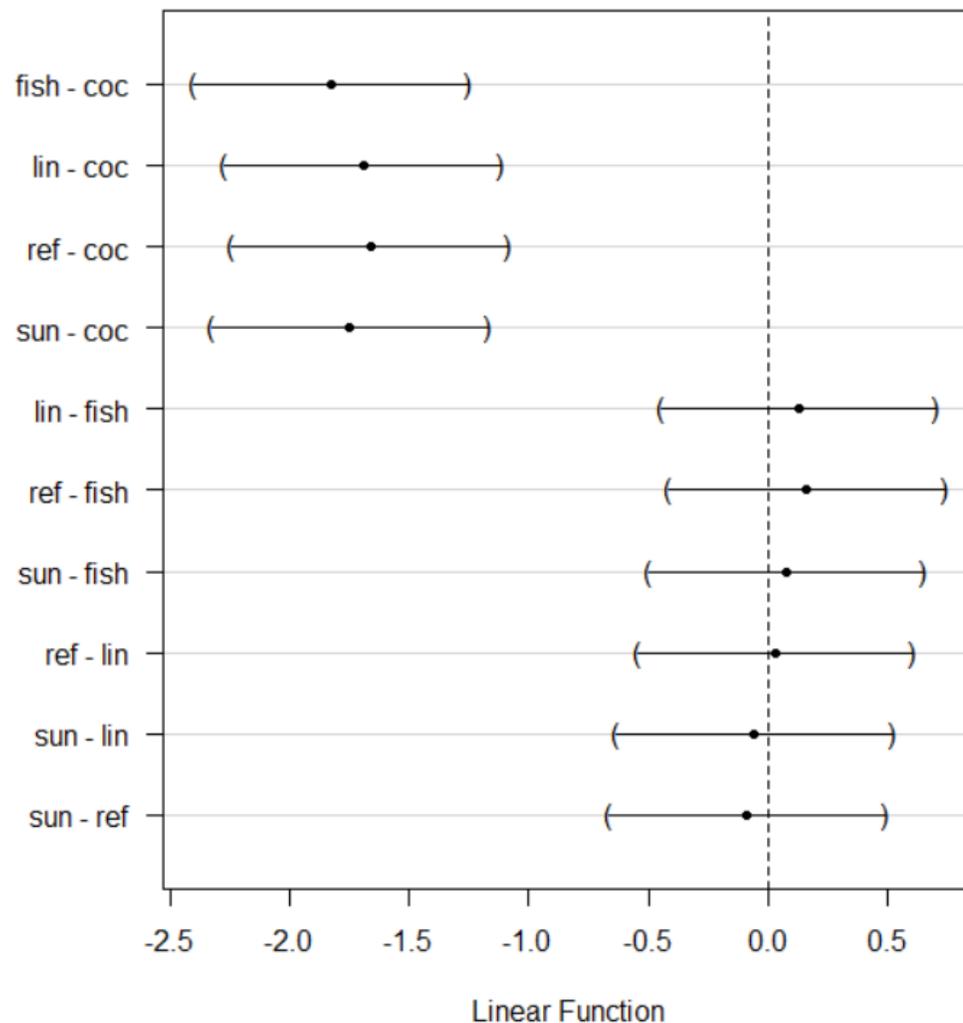
quantile = 2.8763  
 95% family-wise confidence level

Linear Hypotheses:

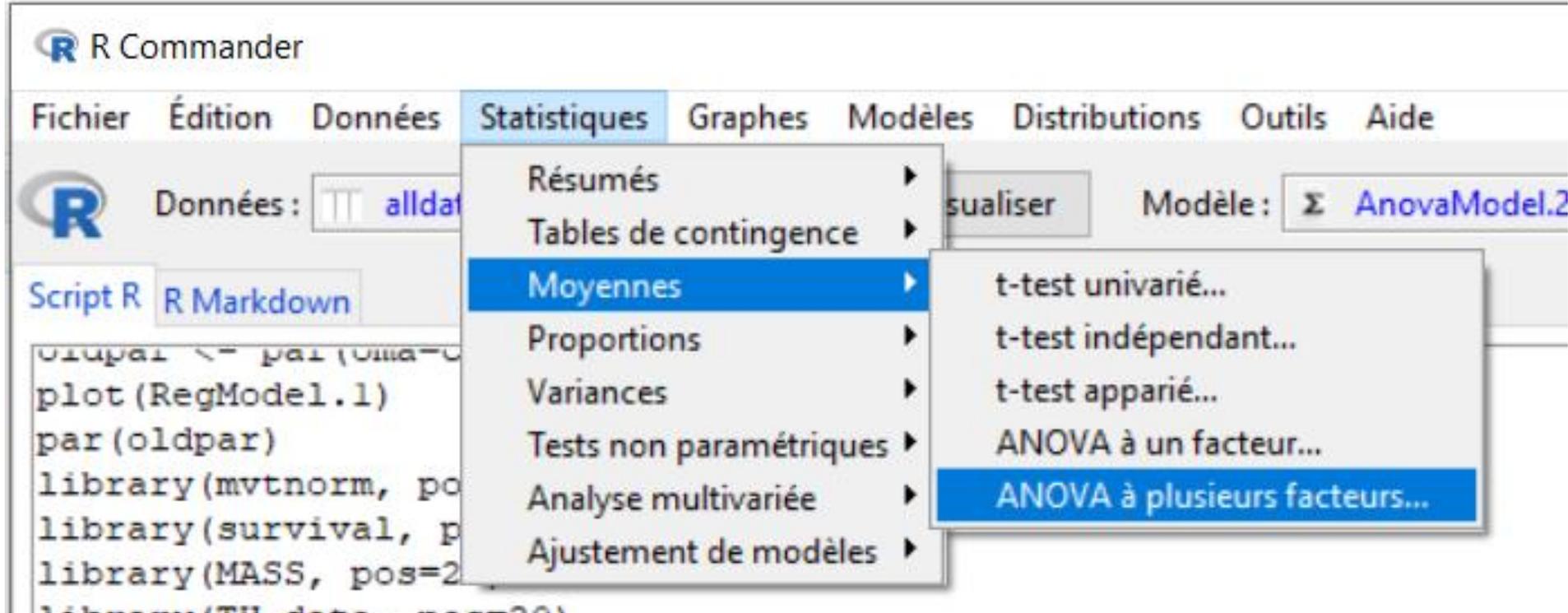
	Estimate	lwr	upr
fish - coc == 0	-1.82625	-2.40335	-1.24915
lin - coc == 0	-1.69500	-2.27210	-1.11790
ref - coc == 0	-1.66375	-2.24085	-1.08665
sun - coc == 0	-1.75000	-2.32710	-1.17290
lin - fish == 0	0.13125	-0.44585	0.70835
ref - fish == 0	0.16250	-0.41460	0.73960
sun - fish == 0	0.07625	-0.50085	0.65335
ref - lin == 0	0.03125	-0.54585	0.60835
sun - lin == 0	-0.05500	-0.63210	0.52210
sun - ref == 0	-0.08625	-0.66335	0.49085

coc fish lin ref sun  
 "b" "a" "a" "a" "a"

## 95% family-wise confidence level



# ANOVA à plusieurs facteurs



Analyse de la variance à plusieurs facteurs

Entrez un nom pour le modèle AnovaModel.4

Facteurs (un ou plusieurs)

Variable réponse (une)

C14.0.bin  
diet  
genotype

ACAT1  
ACAT2  
ACBP  
ACC1  
ACC2  
ACOTH



```
Rcmdr> AnovaModel.3 <- lm(C14.0 ~ diet*genotype, data=alldata, contrasts=list(diet
Rcmdr+   ="contr.sum", genotype ="contr.sum"))
```

```
Rcmdr> Anova(AnovaModel.3)
Anova Table (Type II tests)
```

```
Response: C14.0
```

	Sum Sq	Df	F value	Pr(>F)	
diet	19.3595	4	179.921	< 2.2e-16	***
genotype	1.9892	1	73.947	0.000000001358	***
diet:genotype	2.8398	4	26.392	0.000000001895	***
Residuals	0.8070	30			

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

```
Rcmdr> with(alldata, (tapply(C14.0, list(diet, genotype), mean, na.rm=TRUE)))
```

```
Rcmdr+   # means
      ppar  wt
coc  2.9025 1.3975
fish 0.3975 0.2500
lin  0.5925 0.3175
ref  0.5900 0.3825
sun  0.4475 0.3525
```

```
Rcmdr> with(alldata, (tapply(C14.0, list(diet, genotype), sd, na.rm=TRUE)))
```

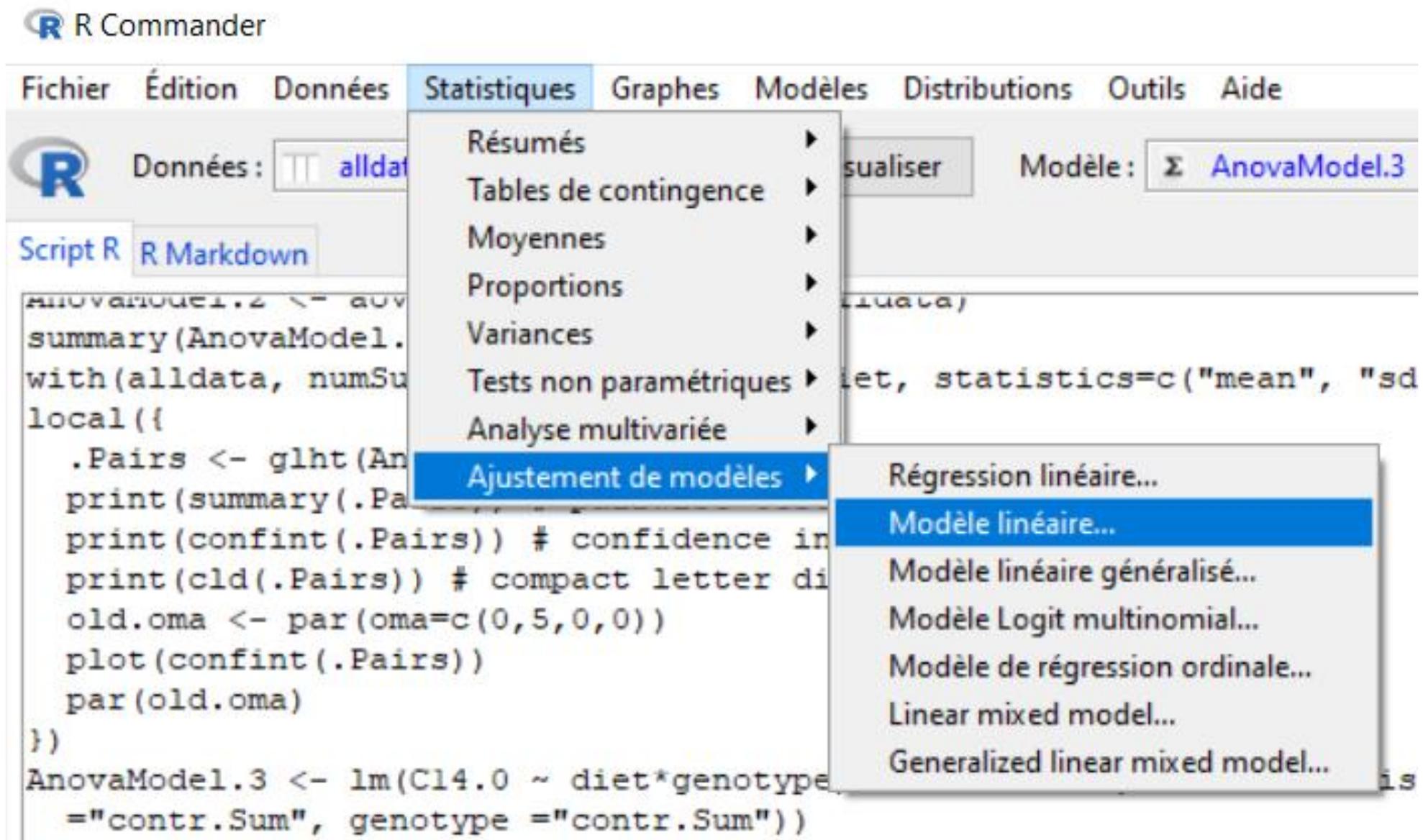
```
Rcmdr+   # std. deviations
      ppar      wt
coc 0.45080484 0.21328776
fish 0.03304038 0.03829708
lin  0.06946222 0.05188127
ref  0.05033223 0.04193249
sun  0.07274384 0.02500000
```

```
Rcmdr> xtabs(~ diet + genotype, data=alldata) # counts
```

diet	genotype	
	ppar	wt
coc	4	4
fish	4	4
lin	4	4
ref	4	4
sun	4	4

## **Partie 5.3 : modèle linéaire**

# Modèle linéaire



R Commander

Fichier Édition Données **Statistiques** Graphes Modèles Distributions Outils Aide

Données:  Actualiser Modèle:

Script R R Markdown

```
AnovaModel.2 <- anova(
summary(AnovaModel.
with(alldata, numSu
local({
.Pairs <- glht(An
print(summary(.Pa
print(confint(.Pairs)) # confidence in
print(cld(.Pairs)) # compact letter di
old.oma <- par(oma=c(0,5,0,0))
plot(confint(.Pairs))
par(old.oma)
})
AnovaModel.3 <- lm(C14.0 ~ diet*genotype
="contr.Sum", genotype = "contr.Sum"))
```

- Résumés
- Tables de contingence
- Moyennes
- Proportions
- Variances
- Tests non paramétriques
- Analyse multivariée
- Ajustement de modèles**
  - Régression linéaire...
  - Modèle linéaire...**
  - Modèle linéaire généralisé...
  - Modèle Logit multinomial...
  - Modèle de régression ordinale...
  - Linear mixed model...
  - Generalized linear mixed model...

Entrez un nom pour le modèle 

Variables (double-clic envoie vers la formule)

ACAT1  
ACAT2  
ACBP  
ACC1  
ACC2  
ACOTH

Formule du modèle

Opérateurs (un clic envoie vers la formule) :

+ \* : / %in% - ^ ( )

Splines/Polynômes :  
(sélectionnez une variable et cliquez)

B-spline spline naturelle polynôme orthogonal polynôme brut

DL pour les splines :   
DL pour les polynômes : 

~

< > < >

 Aide sur les formules de modèle

Expression de sélection

Hauteur

< >



Aide



Réinitialiser



OK



Annuler



Appliquer

```
Rcmdr> LinearModel1.4 <- lm(C14.0 ~ diet, data=alldata)
```

```
Rcmdr> summary(LinearModel1.4)
```

```
Call:
```

```
lm(formula = C14.0 ~ diet, data = alldata)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-0.9400	-0.1053	-0.0200	0.1138	1.0900

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	2.1500	0.1419	15.154	< 2e-16	***
diet[T.fish]	-1.8262	0.2006	-9.102	9.37e-11	***
diet[T.lin]	-1.6950	0.2006	-8.448	5.75e-10	***
diet[T.ref]	-1.6637	0.2006	-8.292	8.93e-10	***
diet[T.sun]	-1.7500	0.2006	-8.722	2.67e-10	***

```
---
```

```
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.4013 on 35 degrees of freedom
```

```
Multiple R-squared:  0.7745,    Adjusted R-squared:  0.7488
```

```
F-statistic: 30.06 on 4 and 35 DF,  p-value: 6.956e-11
```

Entrez un nom pour le modèle 

Variables (double-clic envoie vers la formule)

ACAT1  
ACAT2  
ACBP  
ACC1  
ACC2  
ACOTH

Formule du modèle

Opérateurs (un clic envoie vers la formule) :

+ \* : / %in% - ^ ( )

Splines/Polynômes :

(sélectionnez une variable et cliquez)

B-spline

spline  
naturellepolynôme  
orthogonalpolynôme  
brut

DL pour les splines :

DL pour les polynômes :

 ~ 

 Aide sur  
les formules de modèle

Expression de sélection

Hauteur



Aide



Réinitialiser



OK



Annuler



Appliquer

```
Rcmdr> LinearModel.5 <- lm(C14.0 ~ diet +C16.1n.7, data=alldata)
```

```
Rcmdr> summary(LinearModel.5)
```

```
Call:
```

```
lm(formula = C14.0 ~ diet + C16.1n.7, data = alldata)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-0.41505	-0.10462	0.01557	0.08673	0.48804

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.297422	0.201319	-1.477	0.1488
diet[T.fish]	0.016597	0.167706	0.099	0.9217
diet[T.lin]	-0.246454	0.141853	-1.737	0.0914 .
diet[T.ref]	-0.125270	0.147577	-0.849	0.4019
diet[T.sun]	0.000353	0.161480	0.002	0.9983
C16.1n.7	0.256040	0.020106	12.734	1.71e-14 ***

```
---
```

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

```
Residual standard error: 0.1695 on 34 degrees of freedom
```

```
Multiple R-squared:  0.9609,    Adjusted R-squared:  0.9552
```

```
F-statistic: 167.2 on 5 and 34 DF,  p-value: < 2.2e-16
```

# Modèle linéaire généralisé

The image shows the R Commander interface. The menu bar includes 'Fichier', 'Édition', 'Données', 'Statistiques', 'Graphes', 'Modèles', 'Distributions', 'Outils', and 'Aide'. The 'Statistiques' menu is open, showing options like 'Résumés', 'Tables de contingence', 'Moyennes', 'Proportions', 'Variances', 'Tests non paramétriques', 'Analyse multivariée', and 'Ajustement de modèles'. The 'Ajustement de modèles' option is selected, opening a sub-menu with options: 'Régression linéaire...', 'Modèle linéaire...', 'Modèle linéaire généralisé...', 'Modèle Logit multinomial...', 'Modèle de régression ordinale...', 'Linear mixed model...', and 'Generalized linear mixed model...'. The 'Modèle linéaire généralisé...' option is highlighted. In the background, the R script editor shows code for loading libraries, creating a GLM model, and summarizing it. The 'Données' field shows 'alldata' and the 'Modèle' field shows 'GLM.1'. A 'Soumettre' button is visible at the bottom left.

```
rm(MergedDataset)
library(abind, pos=
library(e1071, pos=
numSummary(alldata[
  "quantiles"), quantiles=c(0,.25,.5,.75
GLM.1 <- glm(diet ~ C16.ln.7, family=bin
summary(GLM.1)
exp(coef(GLM.1)) # Exponentiated coeffi
```

Entrez un nom pour le modèle 

Variables (double-clic envoie vers la formule)

ACAT1  
ACAT2  
ACBP  
ACC1  
ACC2  
ACOTH

Formule du modèle

Opérateurs (un clic envoie vers la formule) :

+ \* : / | %in% - ^ ( )

Splines/Polynômes :

(sélectionnez une variable et cliquez)

B-spline spline naturelle polynôme orthogonal polynôme brut

DL pour les splines : DL pour les polynômes :  ~ 

Aide sur les formules de modèle

Expression de sélection

&lt;tous les cas valides&gt;

Hauteur

&lt;aucune variable sélec &gt;

Famille (double-clic pour sélectionner)

gaussian  
binomial  
poisson  
Gamma  
inverse.gaussian  
quasibinomial  
quasipoisson

Fonction de lien

logit  
probit  
cloglog



Aide



Réinitialiser



OK



Annuler



Appliquer

```
Rcmdr> GLM.1 <- glm(diet ~ C16.1n.7, family=binomial(logit), data=alldata)
RcmdrMsg: [10] AVIS: warning: glm.fit: algorithm did not converge
RcmdrMsg+ warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
Rcmdr> summary(GLM.1)
```

```
Call:
glm(formula = diet ~ C16.1n.7, family = binomial(logit), data = alldata)
```

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-0.000045132	0.000000021	0.000000021	0.000000021	0.000050477

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	271.30	167066.78	0.002	0.999
C16.1n.7	-46.28	28645.62	-0.002	0.999

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 4.0032e+01 on 39 degrees of freedom
Residual deviance: 4.5848e-09 on 38 degrees of freedom
AIC: 4
```

```
Number of Fisher Scoring iterations: 25
```

```
Rcmdr> exp(coef(GLM.1)) # Exponentiated coefficients ("odds ratios")
(Intercept) C16.1n.7
6.669068e+117 7.986535e-21
```