

ANOVA et comparaisons multiples des moyennes

Ce guide présente la marche à suivre pour réaliser les ANOVAS prévues à votre formation.

Les points traités sont les suivants :

- [ANOVA à un critère](#) (leçon 6);
- ANOVA à deux critères;
 - [Nombre de répétitions par groupe](#)
 - [Avec un terme d'interaction](#) (leçon 7)
 - [Sans terme d'interaction](#) (leçon 8)
- [Calcul et tri des moyennes](#).

Les procédures pour effectuer les comparaisons multiples des moyennes sont décrites pour chacune des ANOVAS. Prenez note que ces procédures diffèrent selon les ANOVAS.

La vérification des suppositions de ces tests est présentée dans le guide *Vérification des suppositions*.

ANOVA à un critère

La présentation de ce test se base sur l'Exemple 6.2 des notes de cours (Leçon 6, p. 6 et 15). Vous pouvez accéder aux données à partir de ce [lien](#).

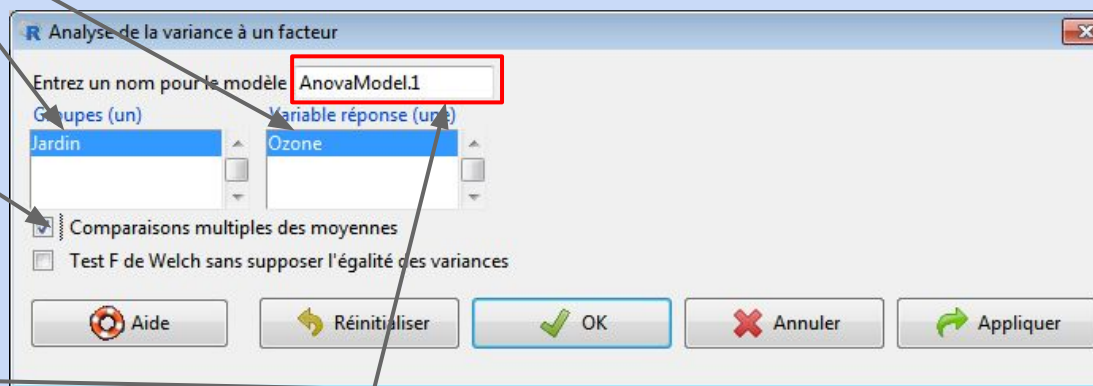
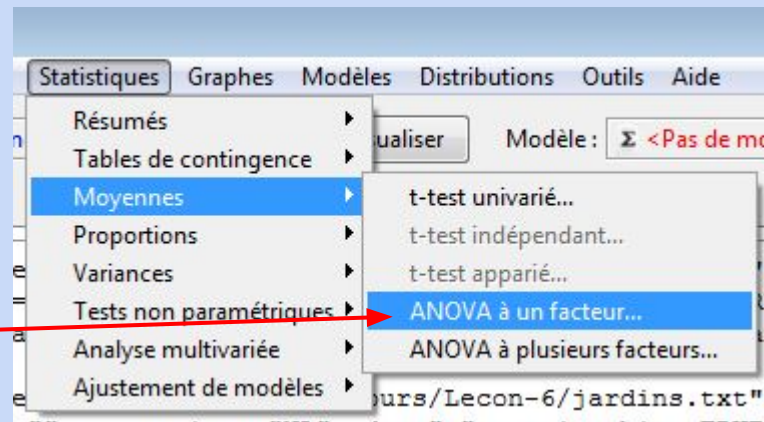
Voici les étapes à suivre pour réaliser une ANOVA à un critère :

1) Cliquez sur le menu Statistiques. Choisissez tout d'abord l'option Moyennes et ensuite l'option ANOVA à un facteur...;

2) Choisissez la variable réponse (ici Ozone) et le groupe (ici Jardin);

3) Cochez la case Comparaison multiple des moyennes. Le test effectué joue le même rôle que le test de Tukey présenté dans les notes de cours.

R Commander sauvegarde les résultats de l'ANOVA dans un objet. Le nom par défaut de l'objet débute toujours par `AnovaModel` (ici `AnovaModel.1`). Il est possible de modifier ce nom. Vous devrez cependant vous en souvenir lors des étapes ultérieures de l'analyse.



ANOVA à un critère (suite)

Vous trouverez les résultats de ce test dans la zone Sortie.

Vous trouverez plusieurs résultats. Afin de trouver les résultats de l'ANOVA, cherchez vers le haut les résultats sous la commande `summary` (ici `summary(AnovaModel.1)`)

Les résultats des comparaisons multiples se trouvent à quelques lignes en dessous de ceux de l'ANOVA.

```
> AnovaModel.1 <- aov(Ozone ~ Jardin, data=ozone)
> summary(AnovaModel.1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Jardin	2	32.4	16.22	0.278	0.759
Residuals	42	2451.5	58.37		

```
Simultaneous Tests for General Linear Hypotheses

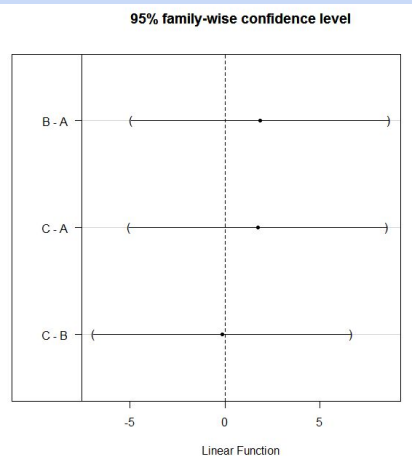
Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = Ozone ~ Jardin, data = ozone)

Linear Hypotheses:
```

	Estimate	Std. Error	t value	Pr(> t)
B - A == 0	1.8593	2.7897	0.667	0.784
C - A == 0	1.7360	2.7897	0.622	0.809
C - B == 0	-0.1233	2.7897	-0.044	0.999

(Adjusted p values reported -- single-step method)



Vous remarquerez dans une nouvelle fenêtre que R Commander a créé un graphique permettant de comparer les intervalles de confiance des comparaisons multiples.

Ce graphique ne doit pas être utilisé pour présenter les résultats des tests étant donné qu'il ne permet pas de comparer les moyennes entre elles.

Nombre de répétitions par groupe

Avant de procéder à l'ANOVA deux critères, on doit tout d'abord déterminer le nombre de répétitions pour chaque combinaison de traitement. Ce nombre permet d'établir si l'on réalise une ANOVA avec ou sans terme d'interaction.

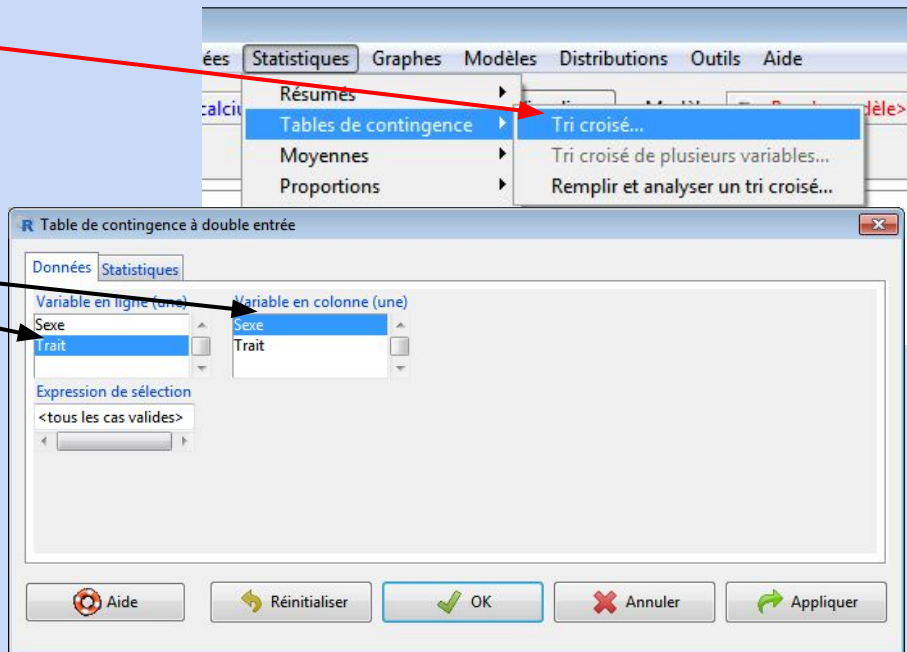
Le calcul du nombre de répétitions est illustré à l'aide de l'Exemple 7.1 des notes de cours (Leçon 7, p. 7). Les données sont accessibles à partir de ce [lien](#).

1) Cliquez sur le menu Statistiques. Choisissez tout d'abord l'option Tables de contingences et ensuite l'option Tri croisé...;

2) Choisissez ensuite les traitements pour lesquels vous souhaitez déterminer le nombre de répétitions (ici *Trait* et *Sexe*);

Le résultat apparaît dans la fenêtre *Sortie*.

```
Frequency table:
      Sexe
Trait  f m
horm   5 5
sans_horm 5 5
```



ANOVA à deux critères avec un terme d'interaction

Ce test sera expliqué grâce à l'Exemple 7.1 des notes de cours (Leçon 7, p. 7). Les données sont accessibles à partir de ce [lien](#).

R Commander ne permet pas de réaliser les ANOVAS à deux critères comme prévu dans le cours. Il est donc nécessaire de soumettre les commandes qui suivent afin de réaliser cette analyse :

```
aov2 <- aov(Concentration~Sexe+Trait+Sexe:Trait,data=calcium)
summary(aov2)
```

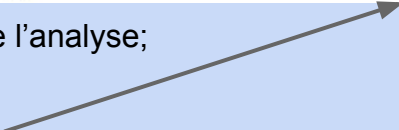
où :

- `aov` : nom de la fonction permettant de réaliser l'ANOVA;
- `Concentration` : nom de la variable réponse;
- `Sexe` et `Trait` : nom des facteurs retenus pour l'analyse;
- `Sexe:Trait` : terme d'interaction des facteurs retenus lors de l'analyse;
- `calcium` : nom du jeu de données;
- `summary` : commande permettant d'afficher les résultats de l'ANOVA.

```
> summary(aov2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sexe	1	70.3	70.3	3.071	0.0989 .
Trait	1	1386.1	1386.1	60.534	7.94e-07 ***
Sexe:Trait	1	4.9	4.9	0.214	0.6499
Residuals	16	366.4	22.9		

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



À l'instar des autres analyses, les résultats apparaissent dans la zone Sortie.

ANOVA à deux critères avec un terme d'interaction (suite)

Des commandes doivent également être utilisées pour réaliser le test de Tukey. Les commandes à soumettre sont les suivantes :

```
trait.mult <- TukeyHSD(aov2, which = "Trait")
trait.mult
```


où :

- TukeyHSD : nom de la fonction permettant de réaliser les comparaisons;
- which : option permettant d'identifier le facteur pour lequel la comparaison doit être faite (ici Trait).

```
> sexe.mult
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Concentration ~ Sexe + Trait + Sexe:Trait, data = calcium)

$Trait
              diff              lwr              upr p adj
sans_horm-horm -16.65 -21.18662 -12.11338 8e-07
```



Les résultats de cette comparaison sont disponibles dans la zone Sortie.

Dans l'éventualité où les deux facteurs avaient été significatifs, les commandes auraient été les suivantes :

```
sexe.trait.mult <- TukeyHSD(aov2, which = "Sexe:Trait")
sexe.trait.mult
```

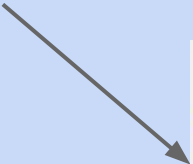
R Commander peut dessiner un graphique des moyennes avec intervalle de confiance. La marche à suivre est décrite dans le guide *Les graphiques avec R Commander*.

ANOVA à deux critères sans terme d'interaction

La présentation de ce test est illustrée à l'aide de l'Exemple 8.4 des notes de cours (Leçon 8, p. 11). Vous pouvez accéder aux données de cet exemple à partir de ce [lien](#).

À l'instar de l'ANOVA précédente, nous devons soumettre deux commandes afin de réaliser cette analyse :

```
bloc.aov <- aov(log.Masse ~ Diete + Bloc, data = cochons)
summary(bloc.aov)
```



```
> summary(bloc.aov)
              Df Sum Sq Mean Sq F value    Pr(>F)
Diete          3  2.2515   0.7505   49.363 5.03e-07 ***
Bloc           4  0.0960   0.0240    1.579   0.243
Residuals     12  0.1824   0.0152
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pour information :

- La variable `Masse` a subi une transformation logarithmique avant l'analyse;
- Les variables `Diete` et `Bloc` ont préalablement été converties en facteurs à l'aide de l'option Convertir des variables numériques en facteurs... (veuillez consulter le guide Transformation et manipulation de données pour de plus amples détails).

Enfin, on effectue le test de Tukey de la même façon qu'avec l'ANOVA à deux critères avec un terme d'interaction. Il en est de même pour le graphique des moyennes avec intervalle de confiance.

Trier les moyennes

Les moyennes présentées dans les notes de cours sont le plus souvent triées. Une première marche à suivre pour trier les moyennes est illustrée à l'aide de l'Exemple 7.1 des notes de cours (Leçon 7, p. 7).

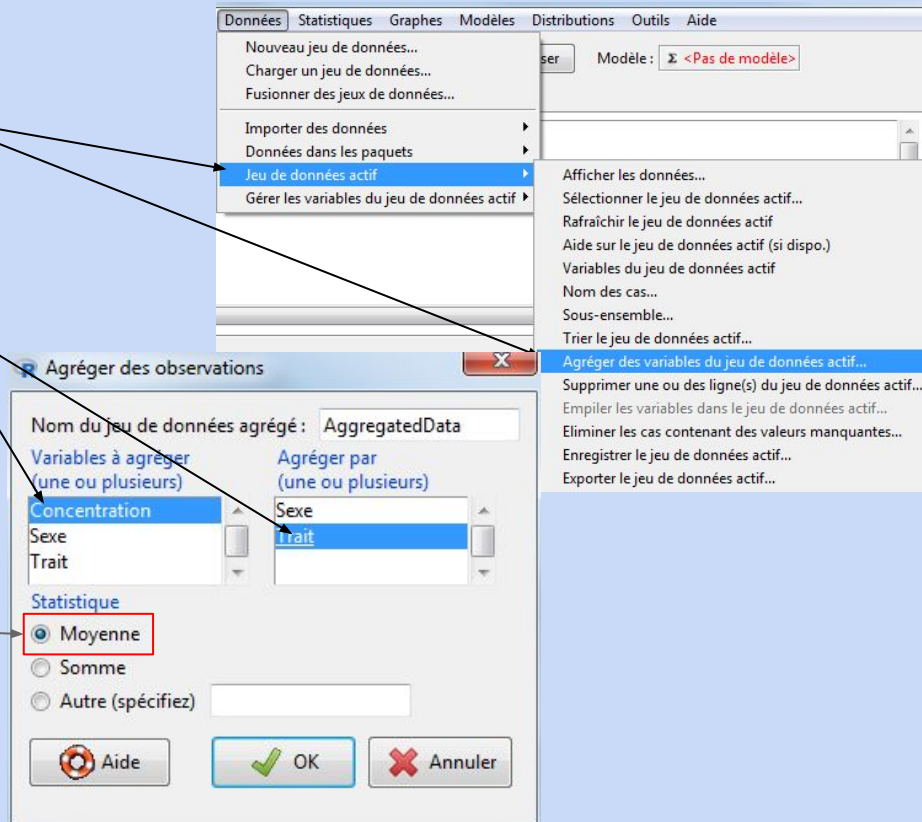
Les données sont accessibles à partir de ce [lien](#).

1) Cliquez sur le menu Données. Choisissez tout d'abord l'option Jeu de données actif et ensuite l'option Agréger des variables du jeu de données actif...;

2) Sélectionnez ensuite la variable à agréger (ici Concentration) et comment cette variable doit être agrégée (zone *Agréger par*). On sélectionne la variable Trait étant donné qu'elle a un effet significatif sur Concentration.

Veuillez ne pas modifier le nom du jeu de données agrégé.

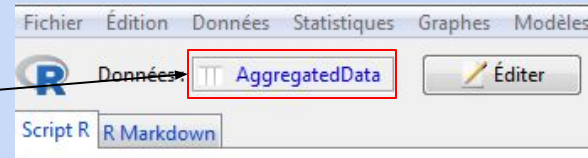
Assurez-vous que la statistique *Moyenne* est sélectionnée et cliquez sur OK.



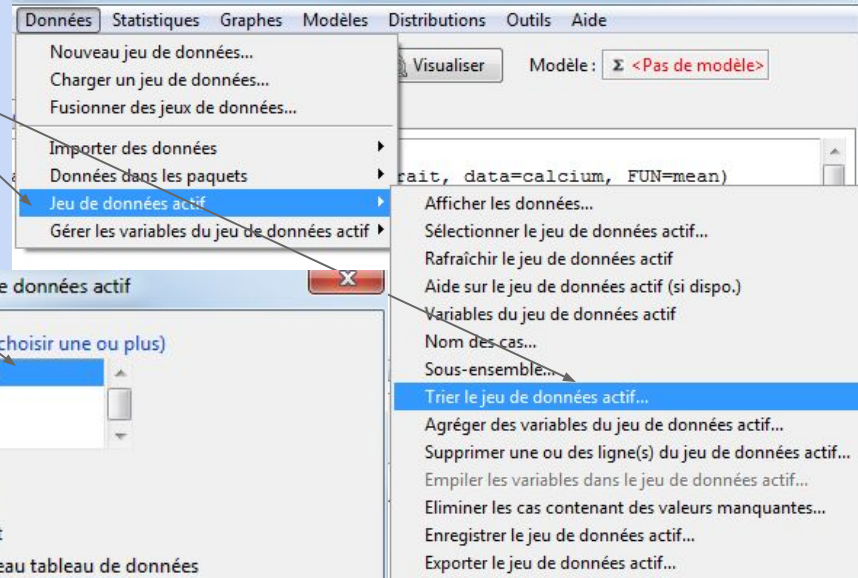
Trier les moyennes (suite)

On doit maintenant trier les moyennes. Pour ce faire :

1) Assurez-vous que le jeu de données actif est celui créé à la diapositive précédente (ici `AggregatedData`)

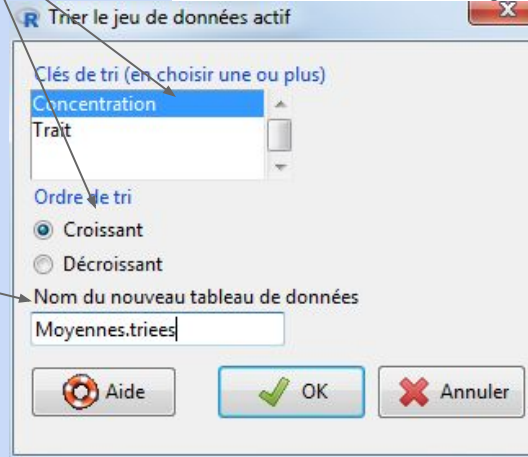


2) Cliquez sur le menu Données. Choisissez tout d'abord l'option Jeu de données actif et ensuite l'option Trier le jeu de données actif...;



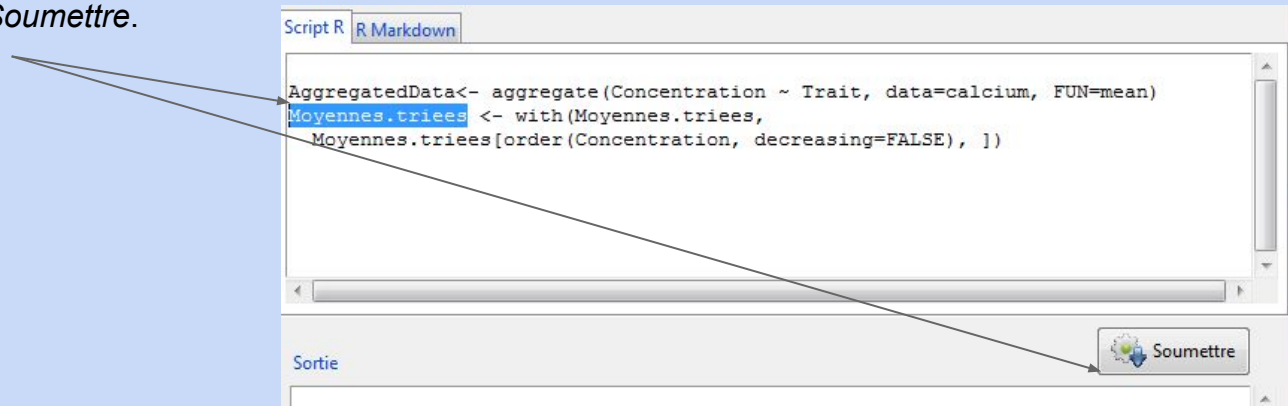
3) Sélectionnez la variable dont vous avez calculé les moyennes (ici `Concentration`) ainsi que l'ordre où vous souhaitez les trier;

4) Donnez un nom au nouveau jeu de données. Saisissez ce nom dans la cellule *Nom du nouveau tableau de données* (p. ex. : `Moyennes.triees`). Cliquez sur **OK** une fois ces étapes terminées.



Trier les moyennes (suite)

La dernière étape consiste à faire apparaître les résultats dans la zone *Sortie*. Pour ce faire, surlignez le nom du nouveau jeu de données et cliquez sur le bouton *Soumettre*.



Les résultats apparaîtront dans la zone *Sortie*.

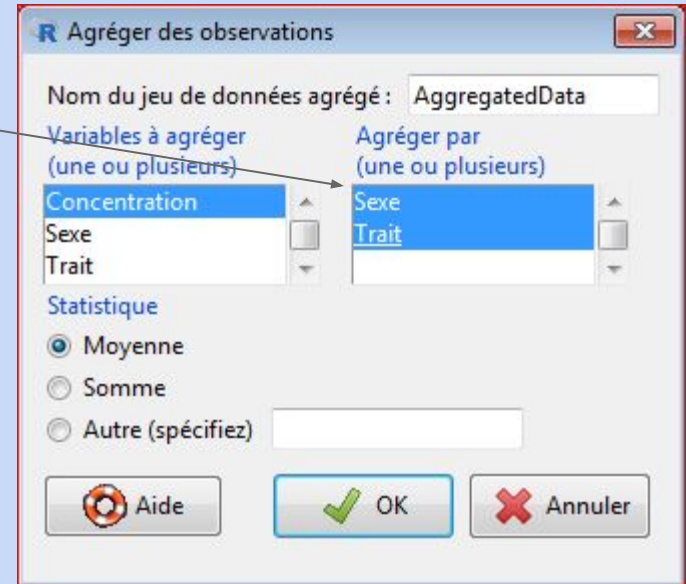
```
> Moyennes.triees
  Trait Concentration
2 sans_horm      13.50
1   horm       30.15
```

Trier les moyennes (suite)

Il est possible d'agréger une variable à partir de plusieurs facteurs. Il s'agit tout simplement de sélectionner les facteur qui nous intéressent (ici `Sexe` et `Trait`).

Une fois les moyennes triées, on obtient le résultat ci-dessous.

	Sexe	Trait	Concentration
4	m	sans_horm	12.12
3	f	sans_horm	14.88
2	m	horm	27.78
1	f	horm	32.52



Trier les moyennes (suite)

Il est parfois impossible de sélectionner l'option Agréger des variables du jeu de données actif....

On peut contourner cette difficulté en soumettant les commandes suivantes :

```
moy <- tapply(X=calcium$Concentration, INDEX=calcium$Trait, FUN=mean)
sort(moy)
```

où :

- `calcium$Concentration` : nom de la variable dont on souhaite calculer la moyenne;
- `calcium$Trait` : variable par laquelle on agrège les moyennes.

Les moyennes sont alors triées en ordre croissant et s'affichent dans la zone *Sortie*.

```
> sort(moy)
sans_horm    horm
    13.50    30.15
```