

Vérification des suppositions

Ce guide présente la marche à suivre pour vérifier les suppositions des tests statistiques présentés dans le cours.

Les points traités sont les suivants :

- Test t ;
 - [Normalité des résidus](#);
 - [Homogénéité de la variance](#);
- ANOVA et régression;
 - [Normalité des résidus et homogénéité de la variance](#);
 - [Linéarité](#).

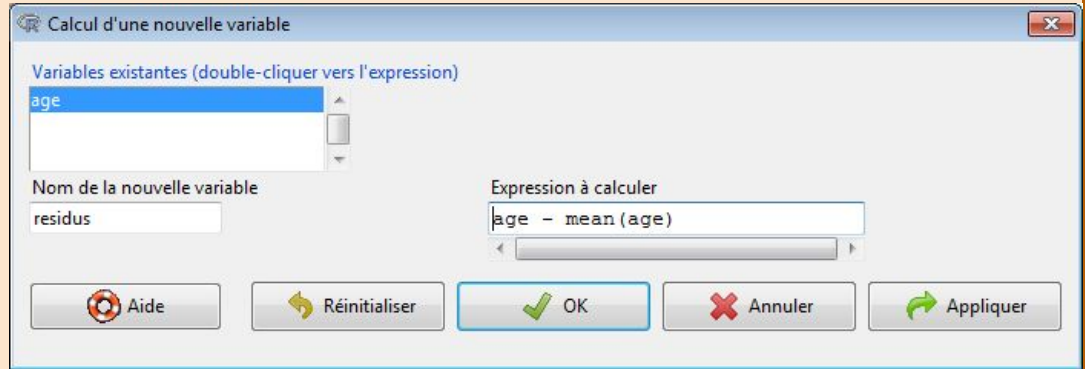
La vérification des suppositions liées aux analyses de fréquences sont présentées dans le guide [Analyses de fréquences](#), puisque ces vérifications sont uniques à ces analyses.

Normalité des résidus - test t

L'Exemple 3.7 (Leçon 3, p. 18) sera utilisé afin de décrire cette vérification.

La vérification de la normalité des résidus débute par le calcul des résidus.

Comme présenté dans le guide *Transformation et manipulation de données*, la création d'une nouvelle variable s'effectue grâce à l'option *Calculer une nouvelle variable* (menu *Données*, sous menu *Gérer les variables du jeu de données actif*).



1) Dans la cellule *Nom de la nouvelle variable*, écrivez le nom *residus*;

2) Saisissez dans la cellule *Expression à calculer* la commande suivante :

```
age - mean (age)
```

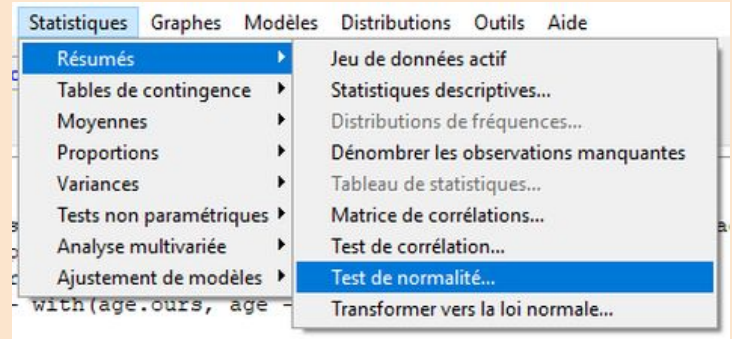
Dans le cas du test t pour deux groupes indépendants, on doit calculer les résidus pour chaque groupe.

On fusionne ensuite les résidus des groupes à l'aide de la procédure expliquée dans le guide *Transformation et manipulation de données*.

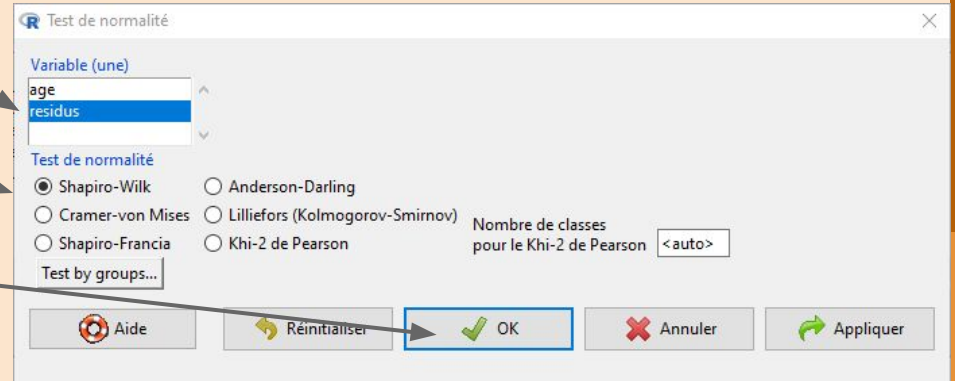
Normalité des résidus - test t (suite)

Les tests de normalité disponible dans R Commander se trouvent à l'option Test de normalité.

3) Dans le menu Statistiques, on choisit l'option Résumés et ensuite l'option Test de normalité...



4) Dans la fenêtre qui apparaît, choisissez la variable à analyser, soit residus;



5) Sélectionnez le test de normalité souhaité (p. ex. : Shapiro-Wilk)

6) Cliquez sur le bouton OK.

Le résultat du test apparaît dans la zone Sortie;

```
> normalityTest(~residus, test="shapiro.test", data=age.ours)

      Shapiro-Wilk normality test

data:  residus
W = 0.93179, p-value = 0.1672
```

Normalité des résidus - test t (fin)

La vérification de la normalité des résidus s'effectue également grâce à l'évaluation d'un graphique quantile-quantile. Ce graphique est créé de la façon suivante :

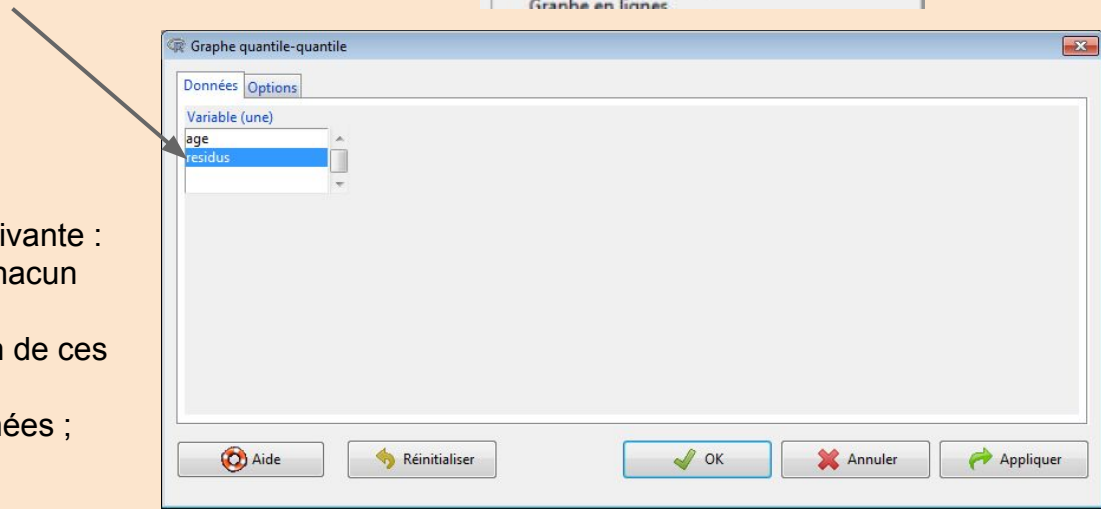
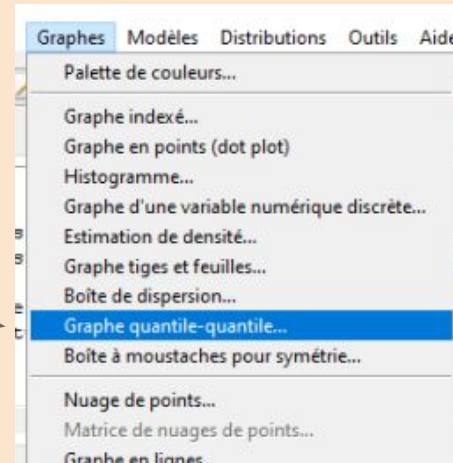
1) Cliquez sur le menu Graphes et ensuite sur l'option Graphique quantile-quantile;

2) À la fenêtre qui apparaît, cliquez sur la variable à étudier (résidus) et appuyez sur le bouton OK.

Un graphique apparaîtra dans une nouvelle fenêtre.

Dans le cas où vous compareriez deux moyennes, la marche à suivre est la suivante :

- Créer un jeu de données pour chacun des traitements ;
- Calculer les résidus pour chacun de ces jeux ;
- Fusionner les deux jeux de données ;
- Vérifier la normalité des résidus.

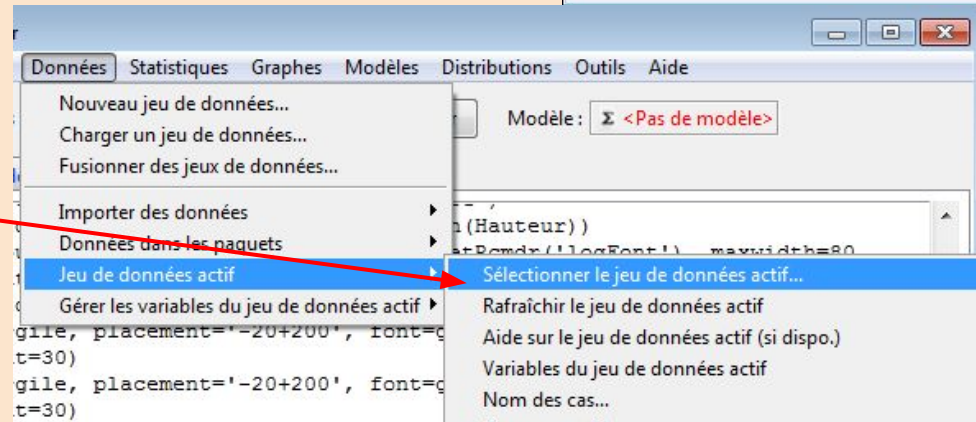
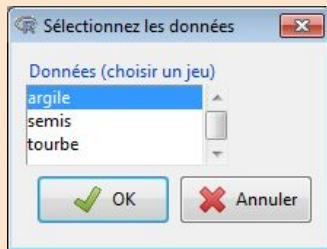
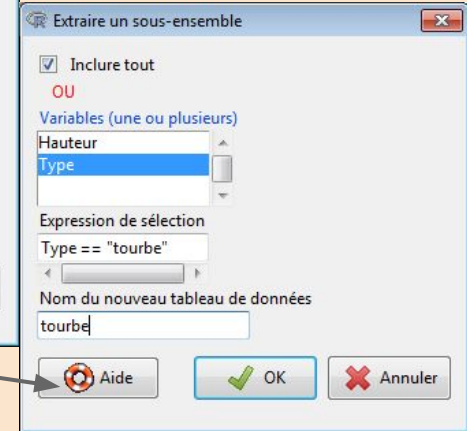
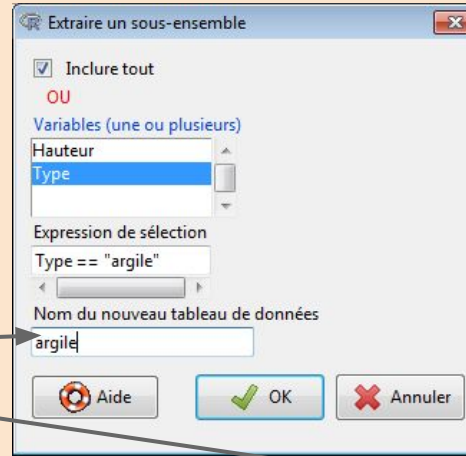


Homogénéité de la variance - test t

On vérifie notamment l'homogénéité des variances à l'aide d'un diagramme de boîtes et moustaches. Cette vérification sera présentée à l'aide des données de l'Exemple 4.1 (Leçon 4, p. 2).

1) Il faut tout d'abord calculer les résidus pour chacun des traitements. En pratique cela signifie que l'on :

- Crée deux jeux de données, soit un pour les semis sur argile et un autre pour les semis sur tourbe;
- Calcule les résidus dans chacun de ces jeux de données. Il est possible de choisir un jeu de données en cliquant sur le nom du jeu de données en cours ou en choisissant l'option Sélectionner le jeu de données actif... et en sélectionnant le jeu avec lequel vous souhaitez travailler;



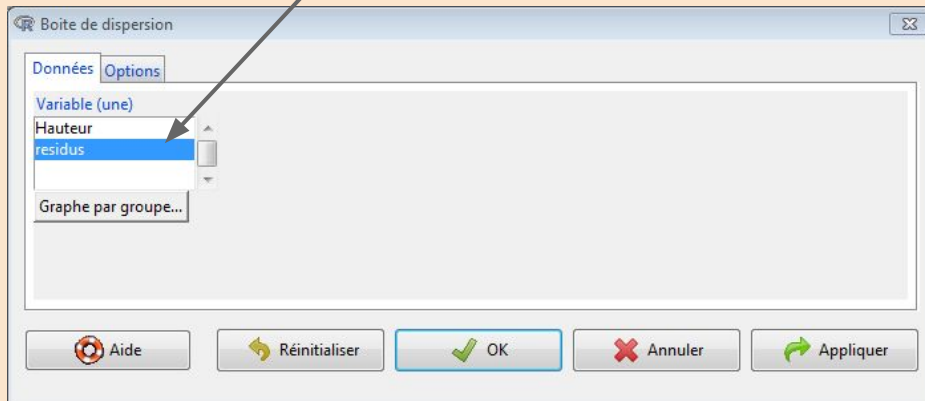
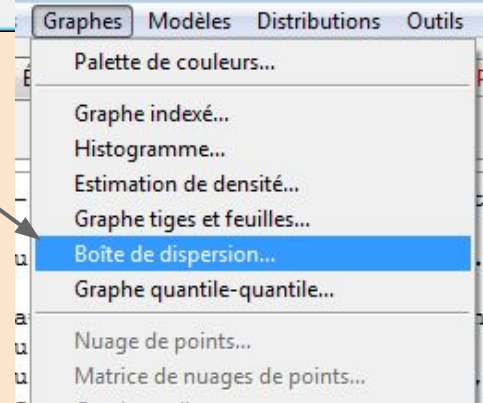
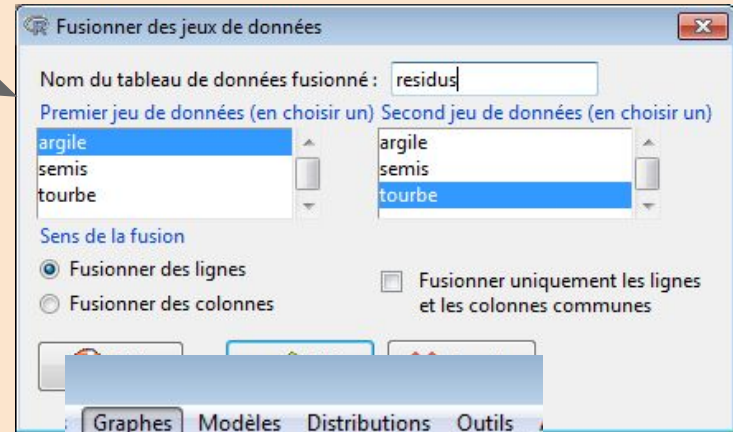
Homogénéité de la variance - test t (suite)

- On fusionne les jeux de données nouvellement créés (ici argile et tourbe). Vous trouverez de plus amples détails sur la marche à suivre dans le guide *Transformation et manipulation de données*;

Les prochaines étapes consistent à créer un diagramme de boîtes et moustaches avec le nouveau jeu de données (c.-à-d. residus).

2) Cliquez sur le menu Graphes et choisissez l'option Boîte de dispersion...;

3) Sélectionnez la variable que vous souhaitez étudier (ici residus);



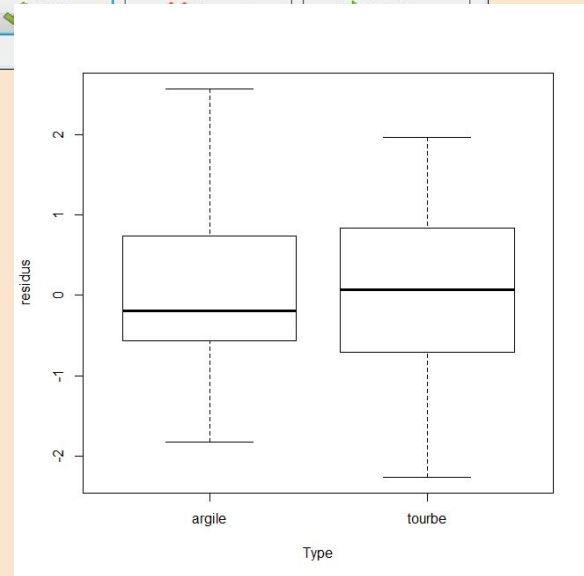
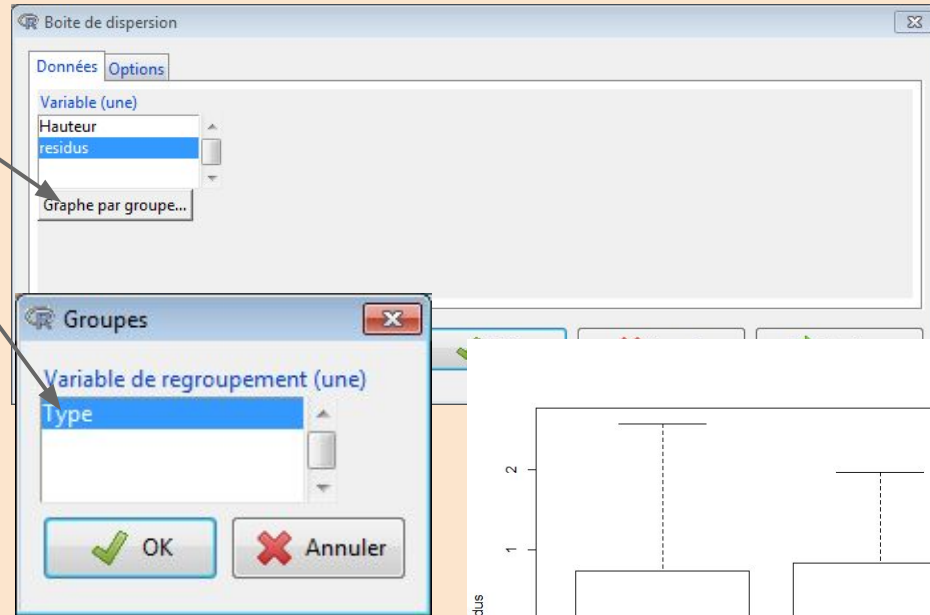
Homogénéité de la variance - test t (suite)

4) Cliquez ensuite sur le bouton Graphe par groupe...;

5) Dans la nouvelle fenêtre qui apparaît, sélectionnez la variable de groupement souhaitée (ici Type) et cliquez sur OK;

6) Cliquez de nouveau sur OK afin de créer le diagramme.

On juge que les variances sont hétérogènes lorsque la hauteur d'une boîte (1^{er} et 3^e quartiles) dépasse de 30% la hauteur de l'autre boîte.



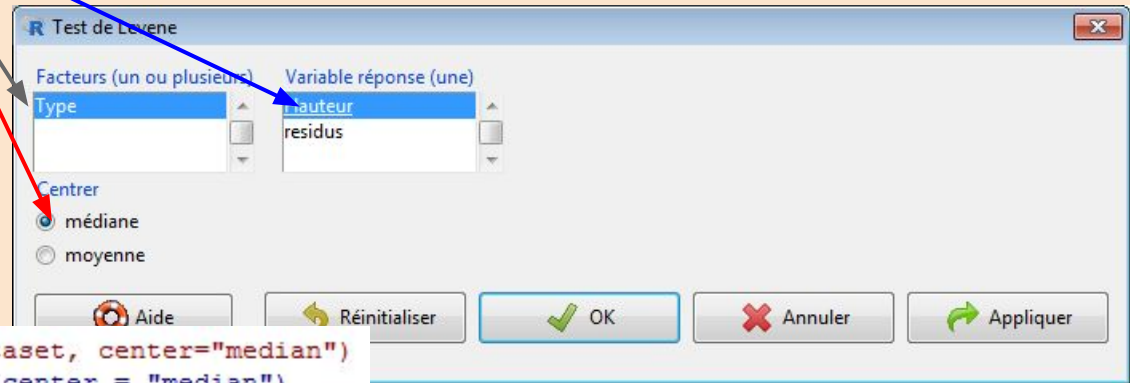
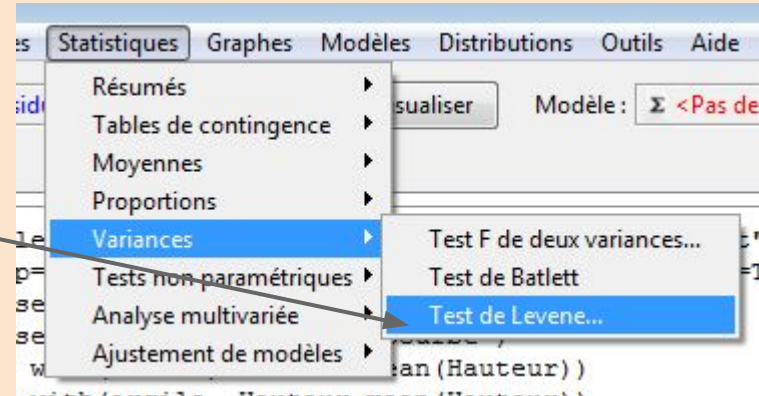
Homogénéité de la variance - test t (fin)

Il est également possible de vérifier l'homogénéité des variances à l'aide de tests formels. L'exemple ci-dessous illustre la réalisation du test de Levene.

1) Dans le menu Statistiques, glissez le curseur sur Variances et sélectionnez l'option Test de Levene...

2) Avant de cliquer sur OK, sélectionnez le ou les facteurs (ici *Type*), la variable réponse (ici *Hauteur*) ainsi que la méthode de centrage (*médiane*).

Le résultat du test est présenté dans la fenêtre *Sortie*



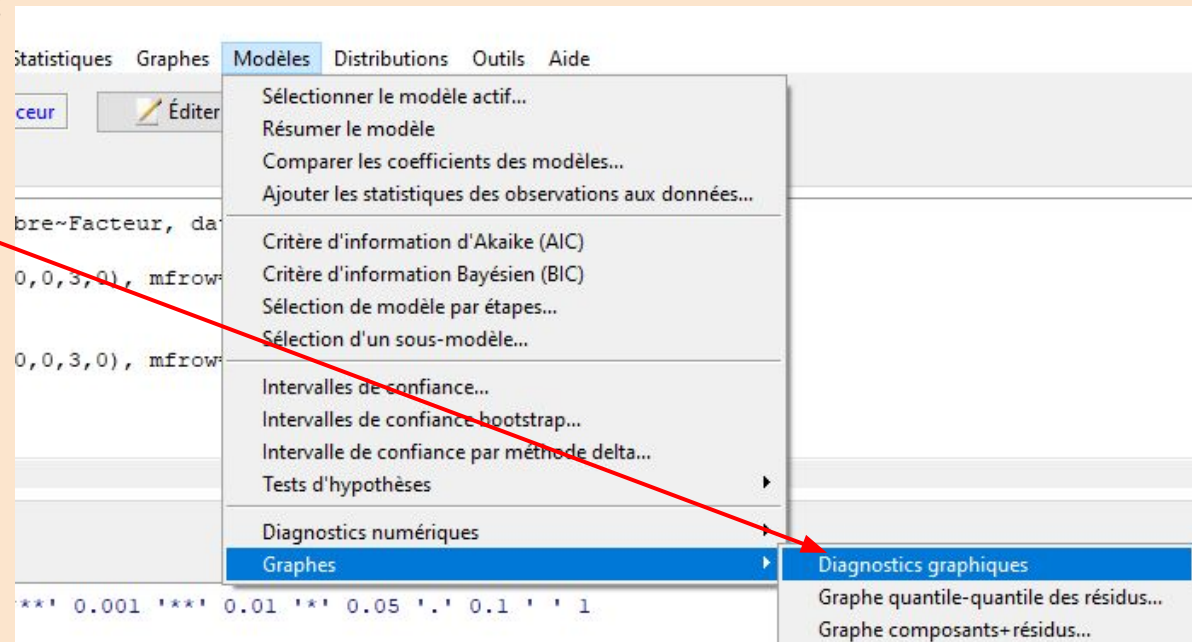
```
> leveneTest(Hauteur ~ Type, data=MergedDataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
  Df F value Pr(>F)
group 1 0.0214 0.8842
  58
```

Normalité des résidus et homogénéité de la variance - ANOVA et régression

La procédure pour vérifier les suppositions des ANOVAS à un critère, à deux critères ou les régressions est toujours la même puisque les fonctions prévues pour ces tests calculent les données nécessaires pour les vérifier.

La procédure qui suit est réalisée grâce à l'Exemple 6.2 des notes de cours (Leçon 6, p. 6 et 15).

Une fois que l'ANOVA ou la régression est réalisée, ouvrez le menu Modèles, sélectionnez Graphes et ensuite l'option Diagnostics graphiques.

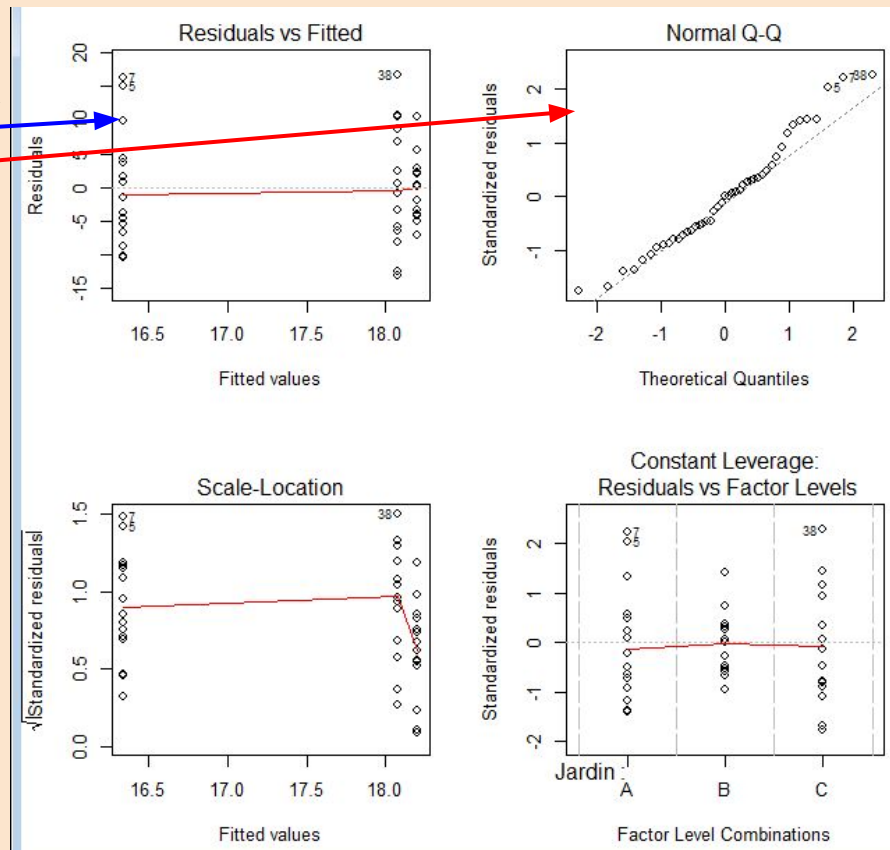


Normalité des résidus & homogénéité de la variance

ANOVA et régression (suite)

Une fois que ces commandes sont soumises, vous obtenez les graphiques suivants :

- Résidus en fonction des valeurs prédites;
- Graphique quantile-quantile.

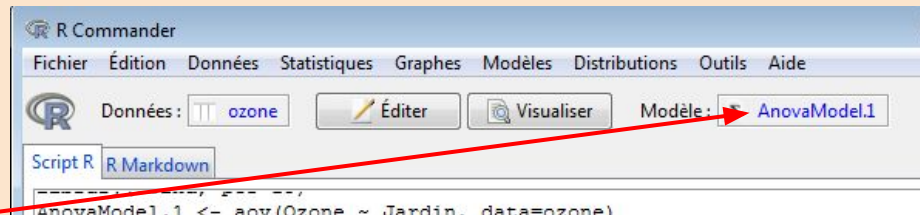


Normalité des résidus & homogénéité de la variance

ANOVA et régression (suite)

La réalisation du test de Shapiro-Wilk doit passer par la soumission d'une commande étant donné que les résidus ne sont pas formellement disponibles dans R Commander.

Lorsque vous réalisez une ANOVA ou une régression, R Commander affiche à la zone *Modèle* l'objet qui contient les données pour réaliser ce test (ici l'objet est AnovaModel.1). Assurez-vous que le nom de cet objet figure dans la commande débutant par `shapiro.test(`.



La commande à soumettre est la suivante :

```
shapiro.test(residuals(AnovaModel.1))
```

où :

- `shapiro.test` : nom de la fonction permettant de réaliser le test de Shapiro-Wilk;
- `residuals` : nom de la fonction permettant d'extraire les résidus de l'objet `AnovaModel.1`.

Les résultats apparaissent dans la zone *Sortie*.

```
> shapiro.test(residuals(AnovaModel.1))  
  
      Shapiro-Wilk normality test  
  
data:  residuals(AnovaModel.1)  
W = 0.9661, p-value = 0.2078
```

Normalité des résidus & homogénéité de la variance

ANOVA et régression (fin)

La vérification formelle de l'homogénéité de la variance sera illustrée grâce au test de Levene.

À l'instar du test précédent, il est également nécessaire d'utiliser une commande. Cette commande est la suivante :

```
leveneTest(Ozone ~ Jardin, data=ozone)
```

où `leveneTest` est le nom de la fonction permettant de réaliser le test de Levene.

Le résultat du test est présenté dans la fenêtre *Sortie*.

```
> leveneTest(Ozone ~ Jardin, data=ozone)
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 2    3.584 0.03655 *
  42
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Dans le cas d'une ANOVA à deux critères comprenant une interaction, la commande prendra la forme suivante :

```
leveneTest(Ozone ~ Facteur1*Facteur2, data=ozone)
```

Ici, `Facteur1` et `Facteur2` sont les facteurs étudiés.

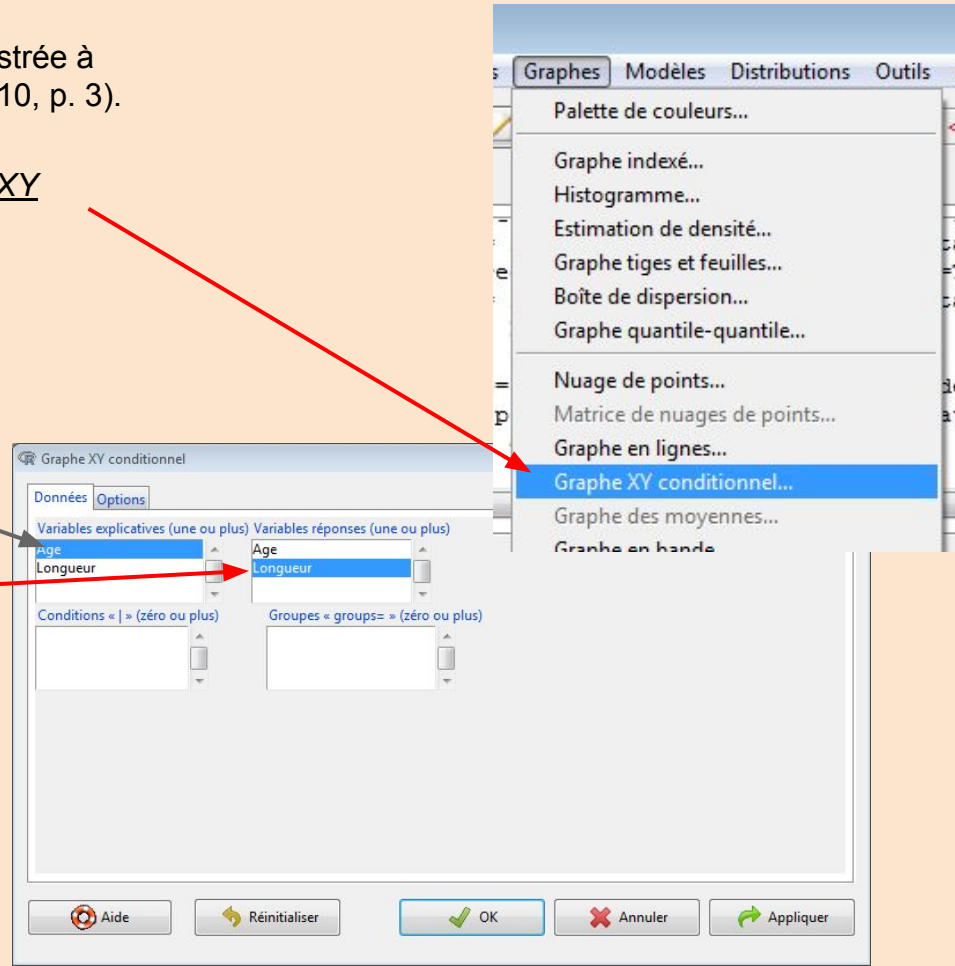
Linéarité

La vérification de la supposition de linéarité sera illustrée à l'aide de l'Exemple 11.1 des notes de cours (Leçon 10, p. 3).

Dans le menu Graphes, choisissez l'option Grappe XY conditionnel...

À la fenêtre qui apparaît, choisissez tout d'abord la variable indépendante (ou "variables explicatives" dans l'interface). Cette variable est Age.

Sélectionnez ensuite la variable dépendante (ou "variables réponses" dans l'interface). À l'exemple 11.1, cette variable est Longueur.



Linéarité (fin)

Le résultat de ces démarches est le suivant :

