

# Régression linéaire simple et corrélation

Ce guide présente la marche à suivre pour réaliser des régressions linéaires simples et des corrélations.

Les points traités sont les suivants :

- Régression linéaire simple;
  - Résidus de Student
  - Intervalle de confiance des coefficients de régression
  - Intervalle de confiance d'une valeur prédite
  - Intervalle de confiance d'une droite
- Corrélation.

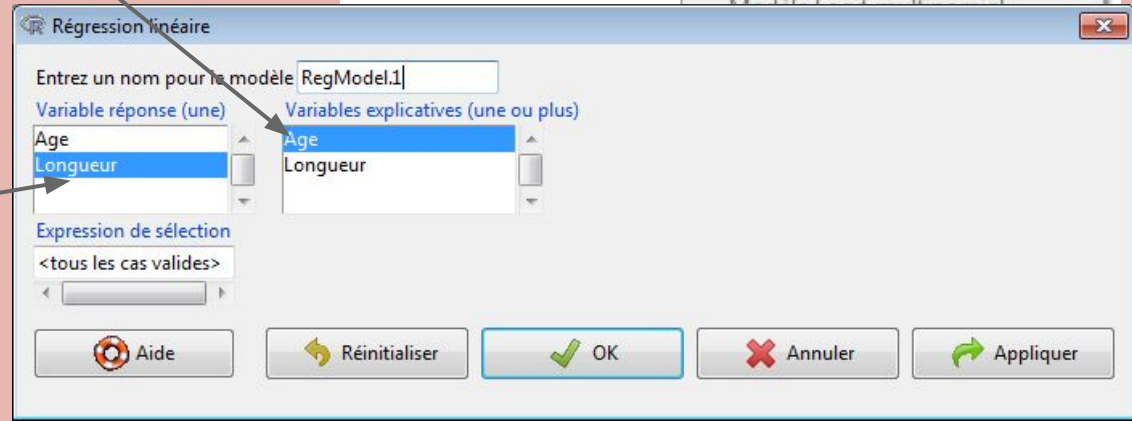
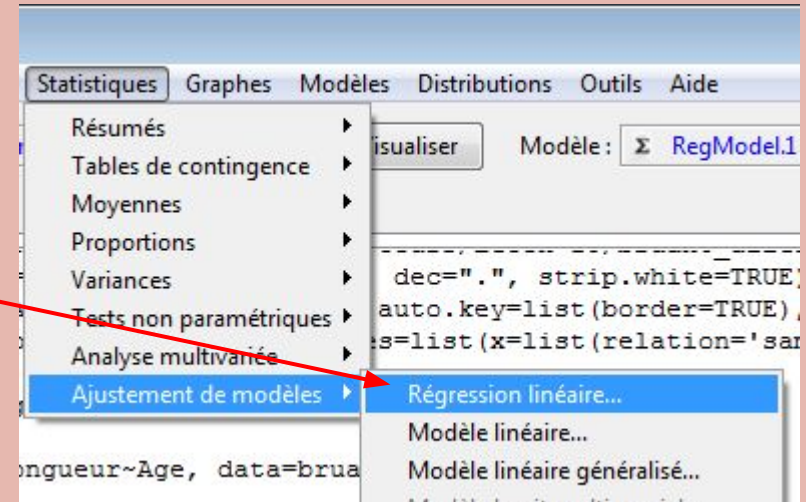
# Régression linéaire simple

La marche à suivre pour réaliser une régression linéaire simple est illustrée à l'aide de l'Exemple 11.1 des notes de cours (Leçon 10, p. 3).

Dans le menu Statistiques, choisissez l'option Ajustement de modèles et ensuite l'option Régression linéaire...

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

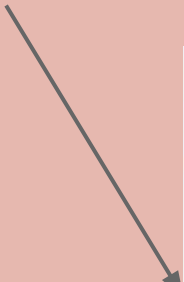
Sélectionnez ensuite la variable dépendante (ou “variable réponse” dans l'interface). À l'Exemple 11.1, cette variable est Longueur.



# Régression linéaire simple (suite)

Après avoir cliqué sur le bouton OK, les résultats de ce test apparaîtront dans la zone

Sortie



```
Call:
lm(formula = Longueur ~ Age, data = bruants)

Residuals:
    Min       1Q   Median       3Q      Max
-0.91092 -0.25558 -0.03573  0.27915  0.83946

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.4804     0.3584   4.130  0.00167 **
Age           0.1950     0.0327   5.965 9.39e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5293 on 11 degrees of freedom
Multiple R-squared:  0.7638, Adjusted R-squared:  0.7424
F-statistic: 35.58 on 1 and 11 DF,  p-value: 9.388e-05
```

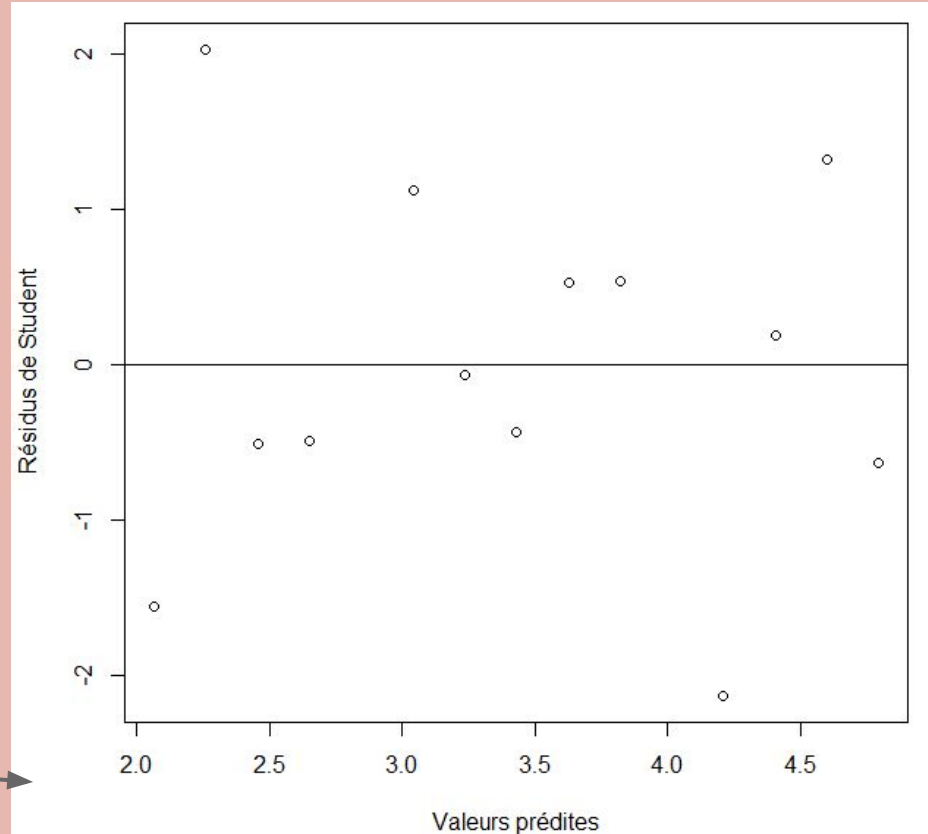
# Résidus de Student

Les notes de cours prévoient l'utilisation des résidus de Student afin d'identifier les valeurs extrêmes. Puisque ces résidus ne sont pas calculés par R Commander, on doit soumettre les commandes qui suivent afin d'obtenir le graphique souhaité :

```
plot(rstudent(RegModel.1) ~  
fitted(RegModel.1), ylab = "Résidus de  
Student", xlab = "Valeurs prédites")  
abline(a=0, b=0)
```

où :

- `rstudent` : fonction permettant de calculer les résidus de Student;
- `fitted` : fonction permettant d'extraire les valeurs prédites de l'objet `RegModel.1`;
- `abline` : fonction permettant d'ajouter une ligne droite. Ici, R ajoute une ligne horizontale à la valeur zéro des résidus.



# Intervalles de confiance des coefficients de régression

Il peut parfois être nécessaire de connaître les intervalles de confiance de l'estimation de l'ordonnée à l'origine et de la pente. Puisque R Commander ne calcule pas cet intervalle, nous devons soumettre la commande qui suit :

```
confint(RegModel.1)
```

où `confint` est la fonction permettant de calculer les intervalles de confiance recherchés.

La colonne 2.5% correspond à la limite inférieure de l'intervalle de confiance alors que la colonne 97.5% donne la valeur supérieure de cet intervalle.

La ligne (Intercept) donne les valeurs pour l'ordonnée à l'origine. La ligne Age quant à elle présente les valeurs pour la pente.

```
> confint(RegModel.1)
              2.5 %      97.5 %
(Intercept) 0.6915240 2.2692511
Age          0.1230712 0.2670051
```

# Intervalle de confiance d'une valeur prédite

À l'instar diapositives précédentes, R Commander ne calcule pas l'intervalle de confiance d'une valeur prédite. La commande à soumettre dans ce cas est présentée ci-dessous.

```
predict (RegModel.1,newdata=data.frame (Age=13) ,  
interval = "confidence")
```

où :

- `predict` : fonction permettant de réaliser la prédiction;
- `newdata=data.frame` : option permettant de soumettre une valeur de variable indépendante;
- `interval` : option indiquant à R qu'il doit calculer l'intervalle de confiance.

Cette commande mène aux résultats suivants :

où :

- `fit` est la valeur estimée;
- `lwr` est la borne inférieure de l'intervalle de confiance;
- `upr` est la borne supérieure de l'intervalle de confiance.

```
> predict (RegModel.1,newdata=data.frame (Age=13) ,interval = "confidence")  
      fit      lwr      upr  
1 4.015884 3.627303 4.404464
```

Ces bornes sont pour un  $\alpha = 0,05$

# Intervalle de confiance d'une droite de régression

R Commander ne permet pas de tracer une régression avec son intervalle de confiance. Il est cependant possible de contourner cette limitation à l'aide de quelques commandes. L'Exemple 11.4 des notes de cours (Leçon 10, p. 25) sera utilisé pour illustrer l'utilisation de ces commandes.

Après avoir importé le fichier `chocolats.txt`, disponible sur le site Web du cours, veuillez soumettre les commandes qui suivent :

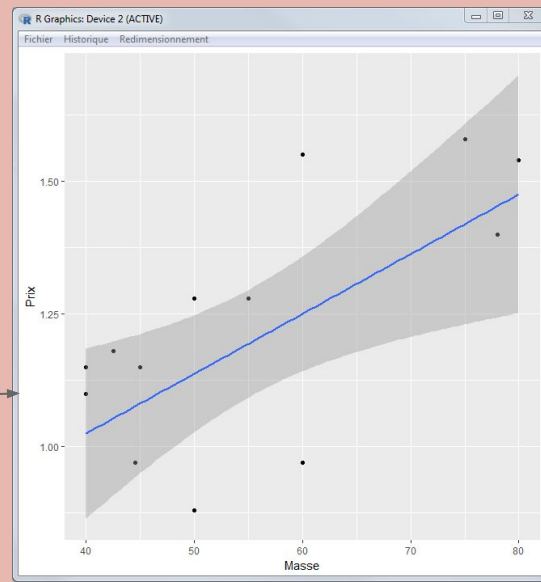
```
library(ggplot2)
ggplot(choco, aes(x=Masse, y=Prix)) + geom_point() + geom_smooth(method=lm)
```

où :

- `library(ggplot2)` : rend disponible la librairie `ggplot2` ;
- `choco` : nom de l'objet contenant les données du fichier `chocolats.txt` ;
- `Masse` : nom de la variable indépendante ;
- `Prix` : nom de la variable dépendante.

Les autres options s'appliquent à tous les jeux de données.

Le graphique apparaît dans une nouvelle fenêtre.



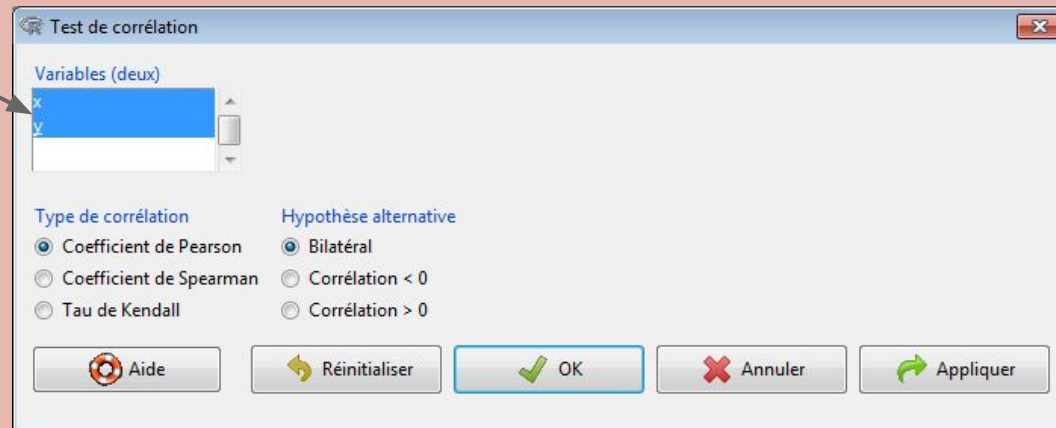
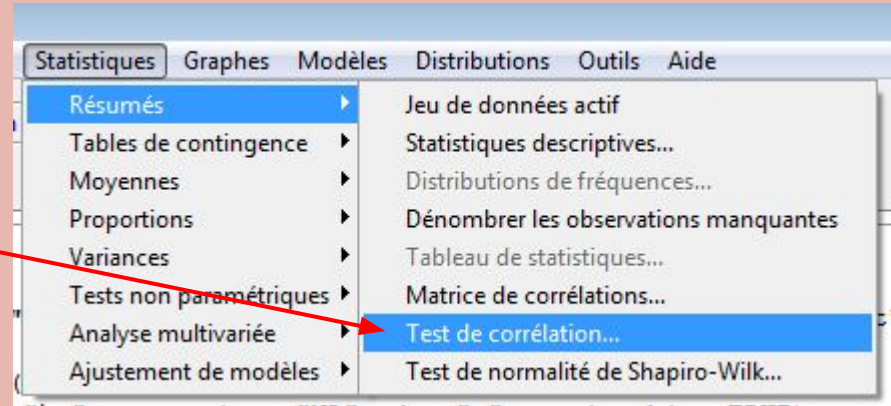


# Corrélation

La marche à suivre pour réaliser la corrélation avec R Commander sera illustrée à l'aide des données utilisées dans [l'aide de cette fonction](#).

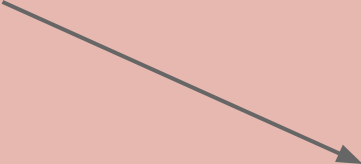
Dans le menu Statistiques, choisissez l'option Résumé et ensuite l'option Test de corrélation...

À la fenêtre qui apparaît, choisissez les variables à analyser (ici  $x$  et  $y$ ).



# Corrélation (suite)

Une fois que vous aurez cliqué sur OK, vous trouverez les résultats dans la fenêtre Sortie.



```
Pearson's product-moment correlation
```

```
data:  x and y
t = 1.8411, df = 7, p-value = 0.1082
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1497426  0.8955795
sample estimates:
      cor
0.5711816
```