

Régression Linéaire Simple (séance 7)

Données :

Éditer

Visualiser

Modèle : Script R

Empty script editor area.

Sortie

Soumettre

Empty output console area.

Messages

```
[2] AVIS: The Windows version of the R Commander works best under  
RGui with the single-document interface (SDI); see ?Commander.
```

R Commander

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

Données Visualiser Modèle : Σ <Pas de modèle>

Script R R Markd

Importer des données

- depuis un fichier texte, le presse-papier ou une URL...
- depuis des données SPSS..
- depuis un fichier SAS xport...
- depuis un fichier SAS b7dat...
- depuis des données Minitab...
- depuis des données STATA...
- depuis un fichier Excel...

Soumettre

R Lire des données depuis un fichier, le presse-papi...

Nom du tableau de données : polriv

Noms de variables dans le fichier :

Indicateur de données manquantes : NA

Emplacement du fichier de données

- Système de fichiers local
- Presse-papier
- Lien internet (URL)

Séparateur de champs

- Espaces
- Virgules [,]
- Semicolons [;]
- Tabulations
- Autre Spécifiez : ;

Séparateur décimal

- Point [.]
- Virgule [,]

Aide OK Annuler

Organiser Nouveau dossier

Nom	Modifié le	Type	Taille
analyses_eau2.csv	14/01/2019 22:32	Fichier CSV Micro...	1 Ko

Nom du fichier : analyses_eau2.csv

Tous les fichiers (*.*)

Ouvrir Annuler

Messages

```
[2] AVIS: The Windows version of the R Commander works best under RGui with the single-document interface (SDI); see ?Commander.
```


Données:

Éditer

Visualiser

Modèle: Script R

```
polriv <- read.table("C:/Users/Pascal/Desktop/seances_5_a_8_2020/data/analyses_eau2.csv",  
  header = TRUE, sep = ";", na.strings = "NA", dec = ",", strip.white = TRUE)
```

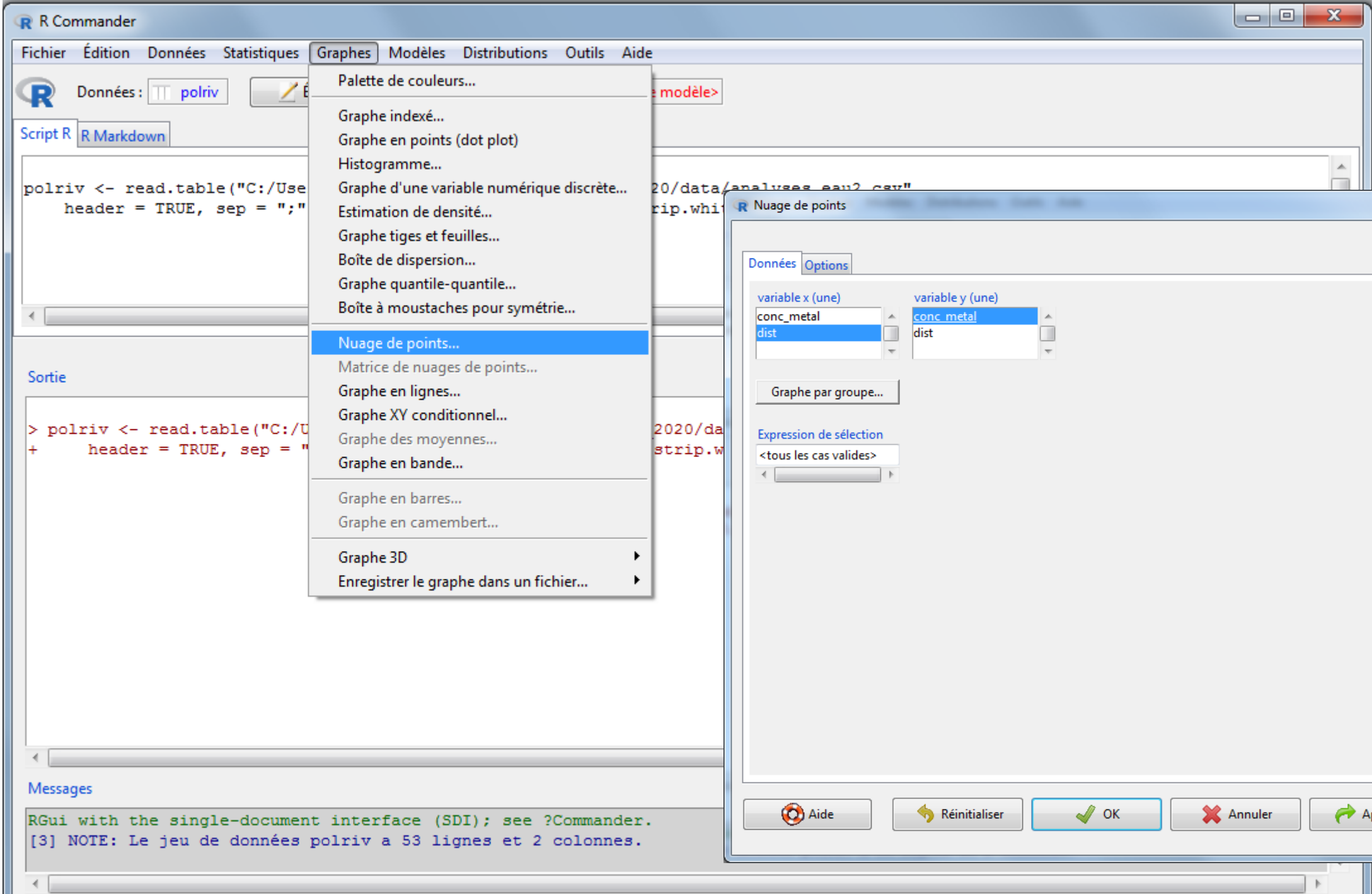
Sortie

Soumettre

```
> polriv <- read.table("C:/Users/Pascal/Desktop/seances_5_a_8_2020/data/analyses_eau2.csv",  
+   header = TRUE, sep = ";", na.strings = "NA", dec = ",", strip.white = TRUE)
```

Messages

```
RGui with the single-document interface (SDI); see ?Commander.  
[3] NOTE: Le jeu de données polriv a 53 lignes et 2 colonnes.
```



R Commander

Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide

Données: polriv

Script R R Markdown

```
polriv <- read.table("C:/Use  
header = TRUE, sep = ";"
```

Sortie

```
> polriv <- read.table("C:/U  
+ header = TRUE, sep = "
```

Messages

```
RGui with the single-document interface (SDI); see ?Commander.  
[3] NOTE: Le jeu de données polriv a 53 lignes et 2 colonnes.
```

- 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...

Nuage de points

Données Options

variable x (une)	variable y (une)
conc_metal	conc_metal
dist	dist

Graphe par groupe...

Expression de sélection

<tous les cas valides>

Aide Réinitialiser OK Annuler Appliquer

R Commander

Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide

Données: polriv

Script R R Markdown

```
polriv <- read.table(  
  header = TRUE, sep
```

Sortie

```
> polriv <- read.tabl  
+   header = TRUE,
```

Messages

```
RGui with the single-  
[3] NOTE: Le jeu de d
```

Nuage de points

Données Options

Options du graphe

- Décalages aléatoires x
- Décalages aléatoires y
- Axe X logarithmique
- Axe Y logarithmique
- Boite à dispersion marginales
- Ligne des moindres carrés
- Courbe de lissage
- Afficher l'étendue

Fenêtre de lissage: 50

- Graphe d'ellipses de concentration

Niveaux de concentration: .5, .9

Identifier des points

- Automatiquement
- Avec la souris de manière interactive
- Ne pas identifier

Nombre de points à identifier: 2

Étiquettes et points du graphe

libellé de l'axe X: <auto>

libellé de l'axe Y: <auto>

Titre du graphe: <auto>

Caractères à utiliser: <auto>

Taille de point: 1.0

Taille du texte des axes: 1.0

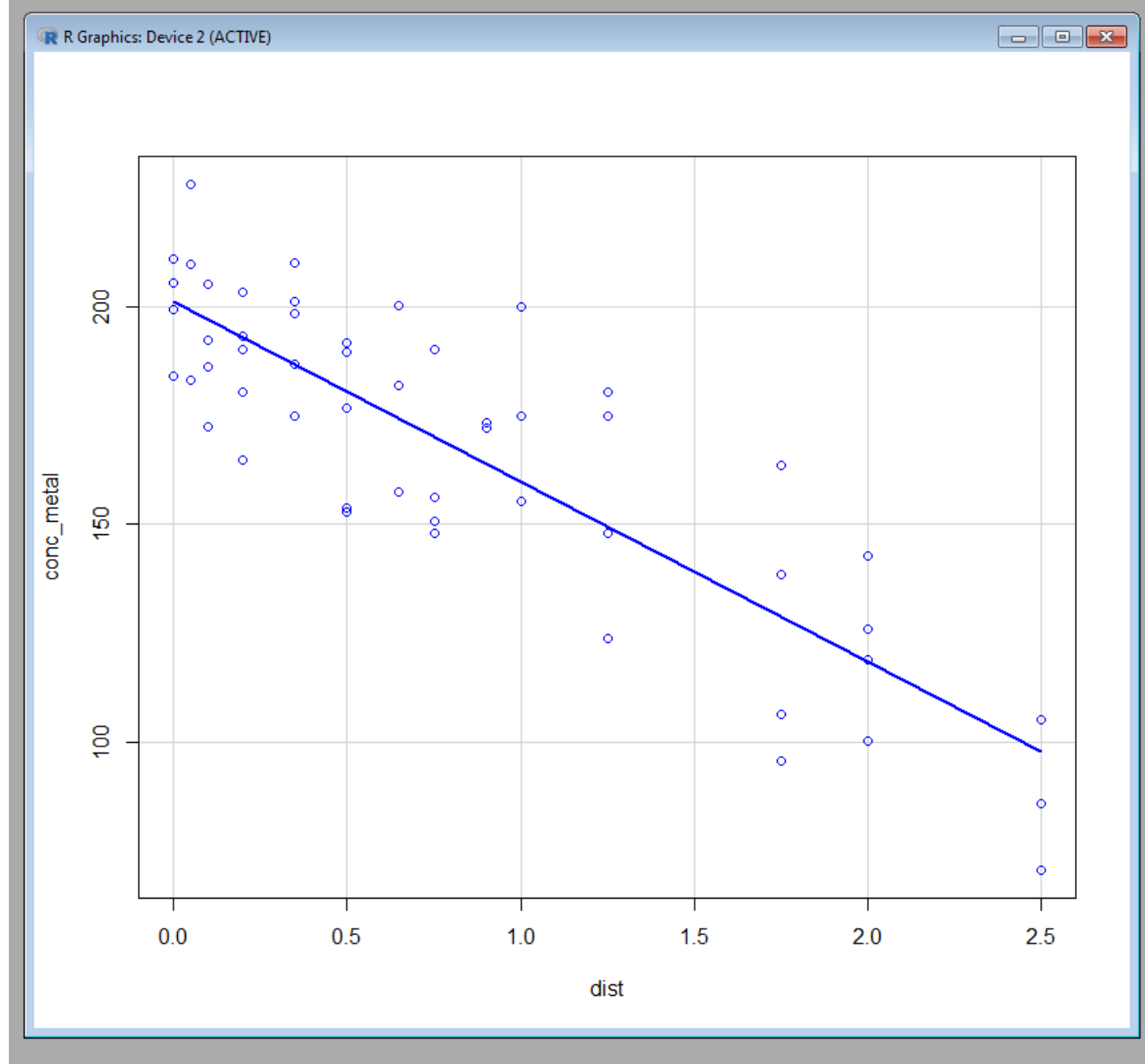
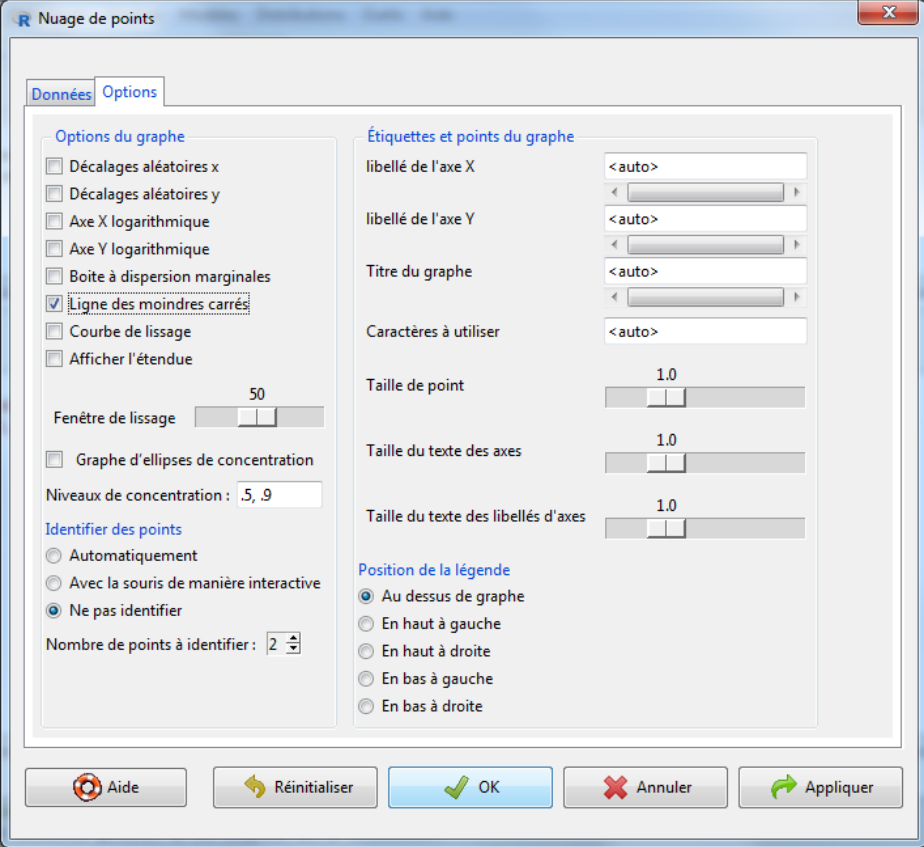
Taille du texte des libellés d'axes: 1.0

Position de la légende

- Au dessus de graphe
- En haut à gauche
- En haut à droite
- En bas à gauche
- En bas à droite

Aide Réinitialiser OK Annuler Appliquer

Soumettre



R Commander

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

Données: [polriv] Modèle: Σ <Pas de modèle>

Script R R Markdown

```
polriv <- read.table("C:/Users/Pascal/Desktop/seances_5_a_8_2020/data/analyses_eau2.csv",
  header = TRUE,
  sep = ";", na.strings = "NA", dec = ",", strip.white = TRUE)
scatterplot(conc_metal ~ dist, regLine = TRUE, smooth = FALSE, boxplots = FALSE, data = polriv)
```

Sortie

```
> polriv <- read.table("C:/Users/Pascal/Desktop/seances_5_a_8_2020/data/analyses_eau2.csv",
+   header = TRUE, sep = ";", na.strings = "NA", dec = ",", strip.white = TRUE)
> scatterplot(conc_metal ~ dist, regLine = TRUE, smooth = FALSE, boxplots = FALSE, data = polriv)
```

Messages

```
RGui with the single-document interface (SDI); see ?Commander.
[3] NOTE: Le jeu de données polriv a 53 lignes et 2 colonnes.
```

Statistiques menu items:
Résumés
Tables de contingence
Moyennes
Proportions
Variances
Tests non paramétriques
Analyse multivariée
Ajustement de modèles

Ajustement de modèles sub-menu items:
Régression linéaire...
Modèle linéaire...
Modèle linéaire généralisé...
Modèle Logit multinomial...
Modèle de régression ordinale...
Linear mixed model...
Generalized linear mixed model...

Soumettre

R Régression linéaire

Entrez un nom pour le modèle MODEL1

Variable réponse (une): conc_metal, dist

Variabes explicatives (une ou plus): conc_metal, dist

Expression de sélection: <tous les cas valides>

Aide Réinitialiser OK Annuler Appliquer

Données:

Éditer

Visualiser

Modèle: Script R

```
polriv <- read.table("C:/Users/Pascal/Desktop/seances_5_a_8_2020/data/analyses_eau2.csv",
  header = TRUE, sep = ";", na.strings = "NA", dec = ",", strip.white = TRUE)
scatterplot(conc_metal ~ dist, regLine = TRUE, smooth = FALSE, boxplots = FALSE, data = polriv)
MODEL1 <- lm(conc_metal ~ dist, data = polriv)
summary(MODEL1)
```

Sortie

Soumettre

Residuals:

Min	1Q	Median	3Q	Max
-33.016	-15.964	0.243	10.536	40.148

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	201.234	3.824	52.62	< 2e-16 ***
dist	-41.382	3.494	-11.85	2.95e-16 ***

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

Residual standard error: 18.66 on 51 degrees of freedom
Multiple R-squared: 0.7334, Adjusted R-squared: 0.7282
F-statistic: 140.3 on 1 and 51 DF, p-value: 2.95e-16

Messages

```
RGui with the single-document interface (SDI); see ?Commander.  
[3] NOTE: Le jeu de données polriv a 53 lignes et 2 colonnes.
```

Régression linéaire : quelques repères

```
> summary(model1)
```

```
Call:
```

```
lm(formula = conc_metal ~ dist)
```

Residuals:

Min	1Q	Median	3Q	Max
-33.016	-15.964	0.243	10.536	40.148

Coefficients:

	Estimate	<u>Std. Error</u>	t value	Pr(> t)
<u>(Intercept)</u>	201.234	3.824	52.62	< 2e-16 ***
<u>dist</u>	-41.382	3.494	-11.85	2.95e-16 ***

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

Residual standard error: 18.66 on 51 degrees of freedom

Multiple R-squared: 0.7334, Adjusted R-squared: 0.7282

F-statistic: 140.3 on 1 and 51 DF, p-value: 2.95e-16

Régression linéaire simple

VD: [metz] conc-metzl (QUANTI continue)

Variable zlietoir (µg/L) (QUANTI continue)

VI: Variable fixée (indépendante): dist

dist → distance en km

conc-metzl

→ concentration de Hg (mercure) en µg/L

en metz

(distance en km
à 250 m de pollution
sur une portion de
la rivière)

noms de variables

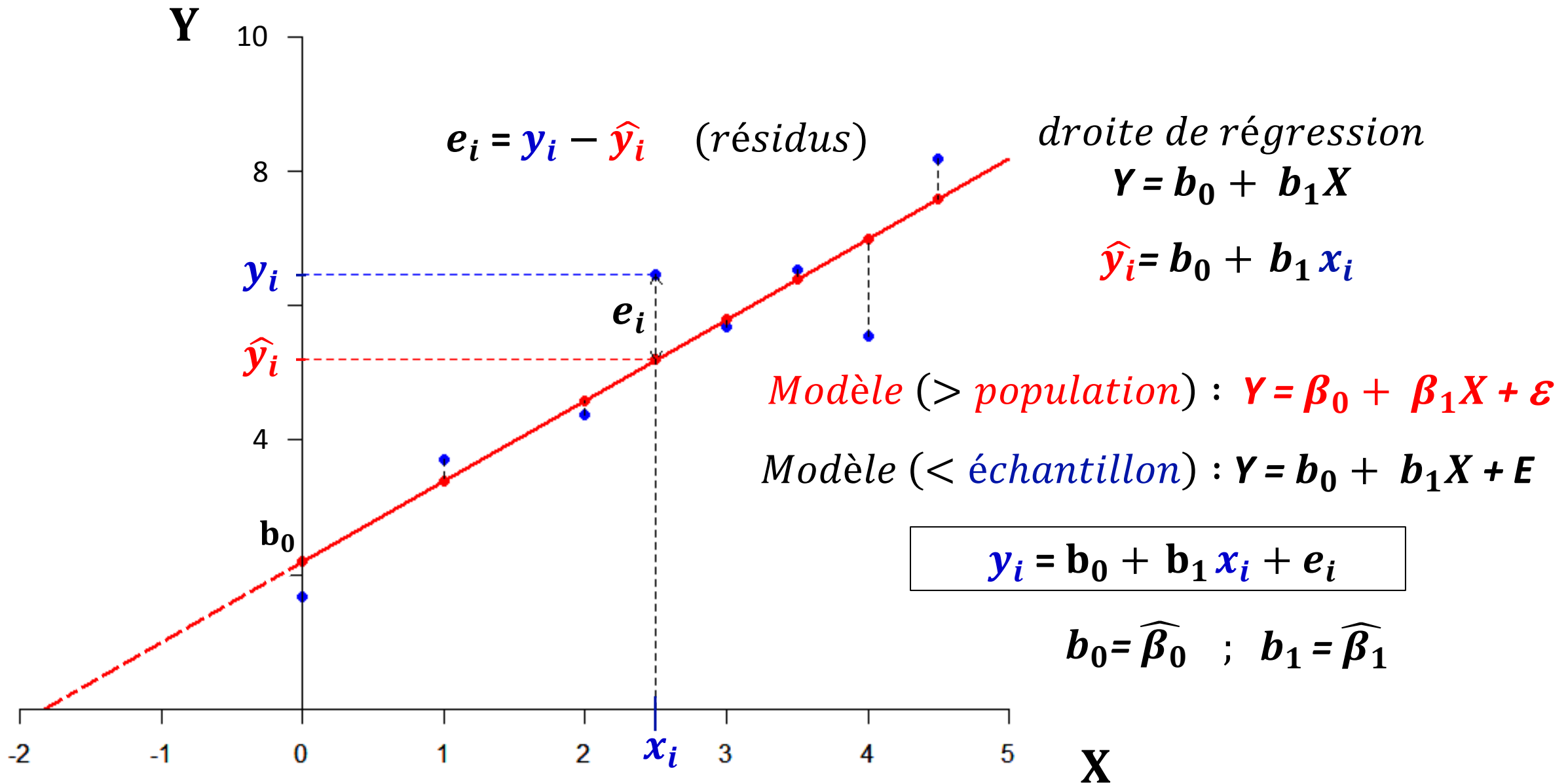
CSV → fichiers texts → séparateur de champs
(d'information) = ";"

$$\text{conc-metzl}_i = b_0 + b_1 \times \text{dist}_i + e_i$$

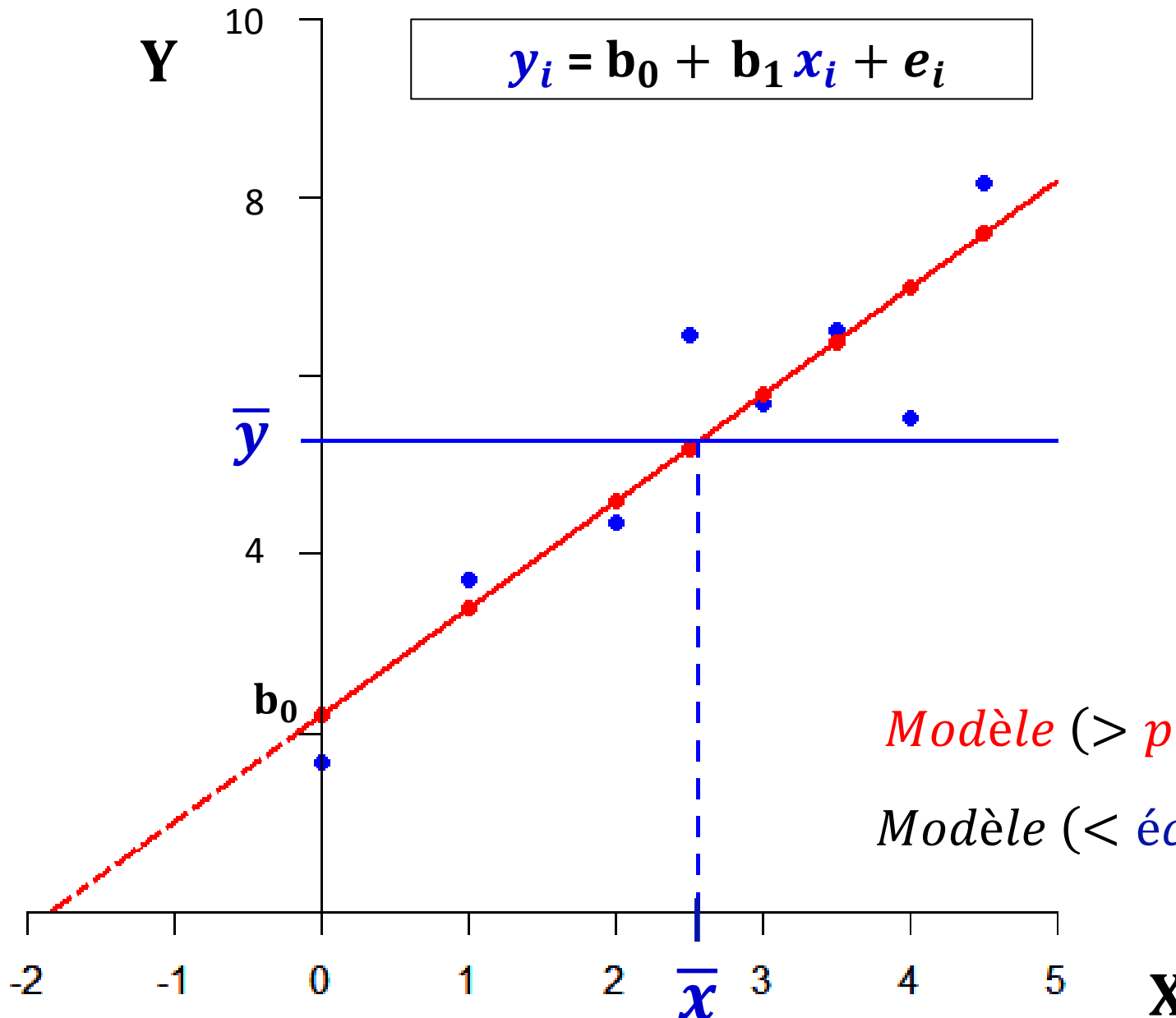
$$b_1 = -41,382$$

$$b_0 = 201,234$$

Régression linéaire : quelques repères



Régression linéaire : quelques repères



droite de régression

$$Y = b_0 + b_1 X$$

$$\hat{y}_i = b_0 + b_1 x_i$$

Tests sur les coefficients b :

(tests de Student)

$$H_0 : \beta_1 = 0 ; H_1 : \beta_1 \neq 0$$

$$H_0 : \beta_0 = 0 ; H_1 : \beta_0 \neq 0$$

Modèle (> population) : $Y = \beta_0 + \beta_1 X + \varepsilon$

Modèle (< échantillon) : $Y = b_0 + b_1 X + E$

$$b_0 = \hat{\beta}_0 ; b_1 = \hat{\beta}_1$$

R Commander

Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide

Données : polriv Éditer

Script R R Markdown

```
polriv <- read.table("C:/Users/Pasc...
  header = TRUE, sep = ";", na.st...
scatterplot(conc_metal ~ dist, regL...
MODEL1 <- lm(conc_metal ~ dist, dat...
summary(MODEL1)
```

Sortie

```
Residuals:
  Min      1Q  Median      3Q      Max
-33.016 -15.964  0.243  10.536  40.148

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   201.234     3.824    52.62 < 2e-16 ***
dist          -41.382     3.494   -11.85 2.95e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.66 on 51 degrees of freedom
Multiple R-squared:  0.7334, Adjusted R-squared:  0.7282
F-statistic: 140.3 on 1 and 51 DF, p-value: 2.95e-16
```

Messages

```
RGui with the single-document interface (SDI); see ?Commander.
[3] NOTE: Le jeu de données polriv a 53 lignes et 2 colonnes.
```

Modèles

- 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

Soumettre

R Ajouter les statistiques des observations au jeu de données

- Valeurs ajustées
- Résidus
- Résidus studentisés
- Valeurs estimées (hat-values)
- Distances de Cook
- Indices des observations

Aide Réinitialiser OK Annuler

R Commander

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

Données: Éditer

Script R R Markdown

```

polriv <- read.table("C:/Users/Pasc...
  header = TRUE, sep = ";", na.st...
scatterplot(conc_metal ~ dist, regL...
MODEL1 <- lm(conc_metal ~ dist, dat...
summary(MODEL1)

```

Sortie

```

Residuals:
  Min      1Q  Median      3Q      Max
-33.016 -15.964   0.243  10.536  40.148

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   201.234     3.824   52.62 < 2e-16 ***
dist          -41.382     3.494  -11.85 2.95e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.66 on 51 degrees of freedom
Multiple R-squared:  0.7334, Adjusted R-squared:  0.7282
F-statistic: 140.3 on 1 and 51 DF,  p-value: 2.95e-16

```

Messages

```

RGui with the single-document interface (SDI); see ?Commander.
[3] NOTE: Le jeu de données polriv a 53 lignes et 2 colonnes.

```

Modèles menu:

- 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...
- Intervalle de confiance...
- Intervalle de confiance bootstrap...
- Intervalle de confiance par méthode delta...
- Tests d'hypothèses
- Diagnostiques numériques
- Graphes**
 - Diagnostiques 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...

Soumettre

R Graphe quantile-quantile pour les résidus

Simuler l'enveloppe de confiance

Identifier des points

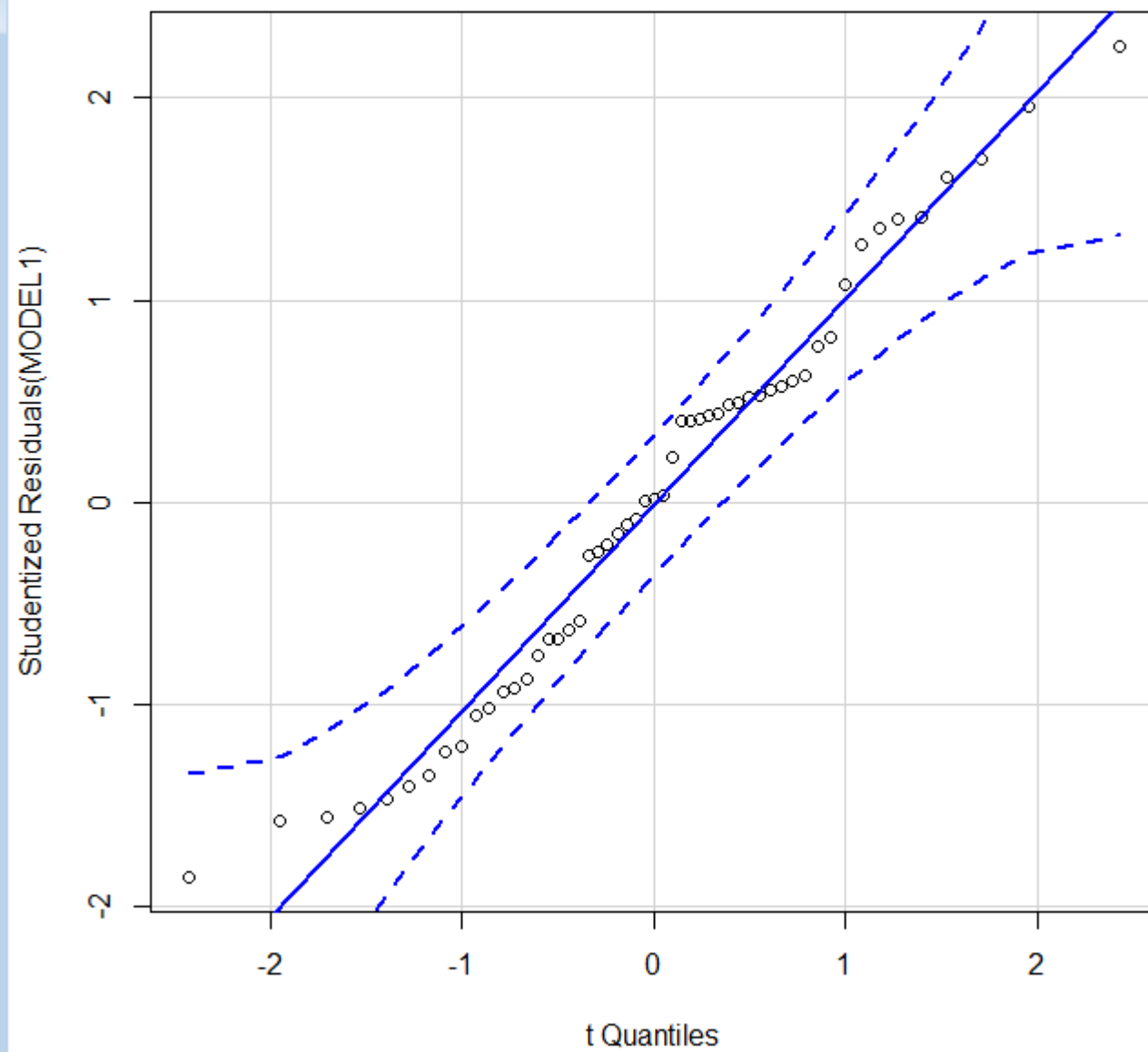
Automatiquement

Avec la souris de manière interactive

Ne pas identifier

Nombre de points à identifier:

Aide Réinitialiser OK Annuler Appliquer



R Commander

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

Données: polriv

Script R R Markdown

```
scatterplot(conc_metal ~ dist)
MODEL1 <- lm(conc_metal ~ dist)
summary(MODEL1)
polriv <- within(polriv, {
  residuals.MODEL1 <- residuals(MODEL1)
})
```

boxplots = FALSE, data = polriv

Soumettre

Sortie

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	201.234	3.123	64.46	0.000000e+00
dist	-41.382	3.123	-13.25	0.000000e+00

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

Residual standard error: 18.12 on 51 degrees of freedom
Multiple R-squared: 0.7334, Adjusted R-squared: 0.7211
F-statistic: 140.3 on 1 and 51 DF, p-value: 2.95e-16

```
> polriv <- within(polriv, {
+   residuals.MODEL1 <- residuals(MODEL1)
+ })
```

Messages

```
[3] NOTE: Le jeu de données polriv a 53 lignes et 2 colonnes.
[4] NOTE: Le jeu de données polriv a 53 lignes et 3 colonnes.
```

Graphes menu:

- 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...

Histogramme dialog:

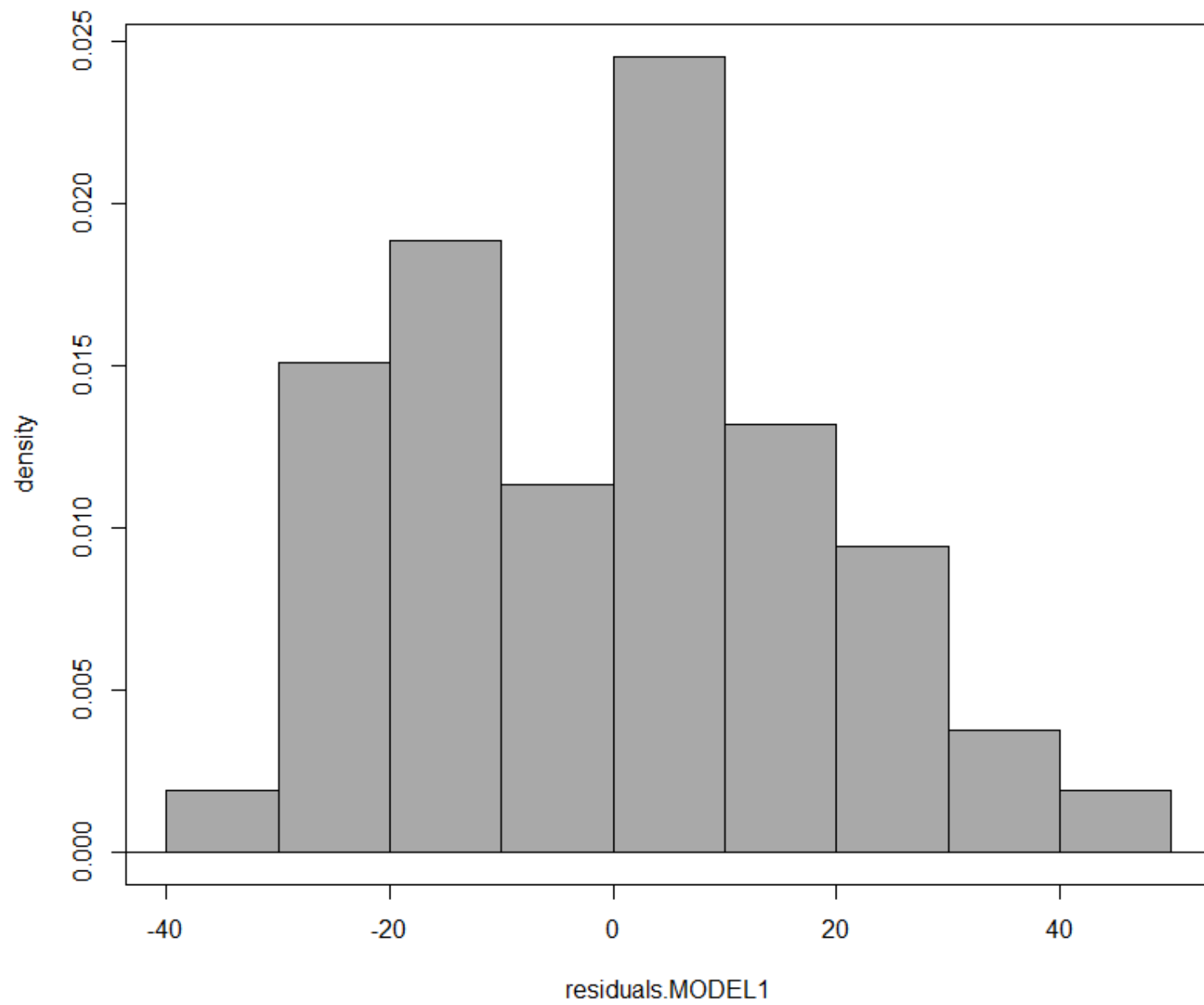
Données Options

Variable (une)

- conc_metal
- dist
- residuals.MODEL1**

Graphe par groupe...

Aide Réinitialiser OK Annuler Appliquer



R Commander

Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide

Données: polriv

Script R R Markdown

```
residuals.MODEL1
})
with(polriv, Hist(residuals.MODEL1, scale = "density", col = "darkgray"))
qqPlot(MODEL1, simulate = TRUE, id = list(method = "shapiro-wilk"))
qqPlot(MODEL1, simulate = FALSE, id = list(method = "shapiro-wilk"))
qqPlot(MODEL1, simulate = FALSE, id = list(method = "shapiro-wilk"))
```

Statistiques

- 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

 Soumettre

Sortie

```
Residual standard error: 18.66 on 51 degrees of freedom
Multiple R-squared: 0.7334, Adjusted R-squared: 0.7282
F-statistic: 140.3 on 1 and 51 DF, p-value: 2.95e-16
```

```
> polriv <- within(polriv, {
+   residuals.MODEL1 <- residuals(MODEL1)
+ })
> with(polriv, Hist(residuals.MODEL1, scale = "density", col = "darkgray"))
> qqPlot(MODEL1, simulate = TRUE, id = list(method = "shapiro-wilk"))
> qqPlot(MODEL1, simulate = FALSE, id = list(method = "shapiro-wilk"))
> qqPlot(MODEL1, simulate = FALSE, id = list(method = "shapiro-wilk"))
```

Messages

```
[3] NOTE: Le jeu de données polriv a 53 lignes et 2 colonnes.
[4] NOTE: Le jeu de données polriv a 53 lignes et 3 colonnes.
```

R Test de normalité

Variable (une)






conc_metal
dist
residuals.MODEL1

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: <auto>

Test by groups...

 Aide  Réinitialiser  OK  Annuler  Appliquer

R Commander

Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide

Données: polriv Éditer Visualiser Modèle: MODEL1

Script R R Markdown

```
})  
with(polriv, Hist(residuals.MODEL1, scale = "density", breaks = "Sturges", col = "darkgray"))  
qqPlot(MODEL1, simulate = TRUE, id = list(method = "y", n = 0))  
qqPlot(MODEL1, simulate = FALSE, id = list(method = "y", n = 0))  
qqPlot(MODEL1, simulate = FALSE, id = list(method = "y", n = 0))  
normalityTest(~residuals.MODEL1, test = "shapiro.test", data = polriv)
```

Sortie

Soumettre

```
> with(polriv, Hist(residuals.MODEL1, scale = "density", breaks = "Sturges", col = "darkgray"))  
  
> qqPlot(MODEL1, simulate = TRUE, id = list(method = "y", n = 0))  
  
> qqPlot(MODEL1, simulate = FALSE, id = list(method = "y", n = 0))  
  
> qqPlot(MODEL1, simulate = FALSE, id = list(method = "y", n = 0))  
  
> normalityTest(~residuals.MODEL1, test = "shapiro.test", data = polriv)  
  
      Shapiro-Wilk normality test  
  
data:  residuals.MODEL1  
W = 0.97065, p-value = 0.2152
```

Messages

```
[3] NOTE: Le jeu de données polriv a 53 lignes et 2 colonnes.  
[4] NOTE: Le jeu de données polriv a 53 lignes et 3 colonnes.
```

Shapiro-Wilk normality test

data: residuals.MODEL1

W = 0.97065, p-value = 0.2152

CONCLUSION