DÉPARTEMENT DE MATHÉMATIQUES
ET DE GÉNIE INDUSTRIEL

MTH2302D - PROBABILITÉS ET STATISTIQUE

Devoir - Hiver 2023

Date de remise : 18 avril avant 23h59 (dans Moodle)

Veuillez remplir le tableau suivant et joindre cette page à votre rapport.

Identification de l'étudiant(e) 1	
Nom : Roy	**Prénom :** Sébastien
Groupe : 03	**Matricule :** 2146331

Identification de l'étudiant(e) 2	
Nom : Roux	**Prénom :** Julien
Groupe : 03	**Matricule :** 2060886

Placer les deux fichiers `DevoirDH23.csv` et `charger.R` dans le répertoire de travail de `R`.
En utilisant votre **matricule**, exécuter ensuite (dans cet ordre) les deux commandes suivantes dans `R` pour générer votre ensemble de données personnalisées 'mondata' :

```
source('charger.R')
mondata <- charger(matricule)
```

Question	Note
a)	**/4**
b)	**/7**
c)	**/12**
d)	**/5**
Présentation	**/2**
TOTAL	**/30**

Mardi le 18 avril 2023

Devoir MTH2302D

Julien Roux 2060886 - Sébastien Roy 2146331

Option génerale pour les graphiques

Phase 1

On charge les données depuis le csv en fonction du matricule.

A data.frame: 5 × 4

	Sales	Price	Advertising	Region
	<dbl>	<int>	<int>	<int>
25	5.58	148	10	1
181	12.61	104	10	0
202	9.48	132	10	0
129	5.87	109	0	1
289	3.02	90	11	0

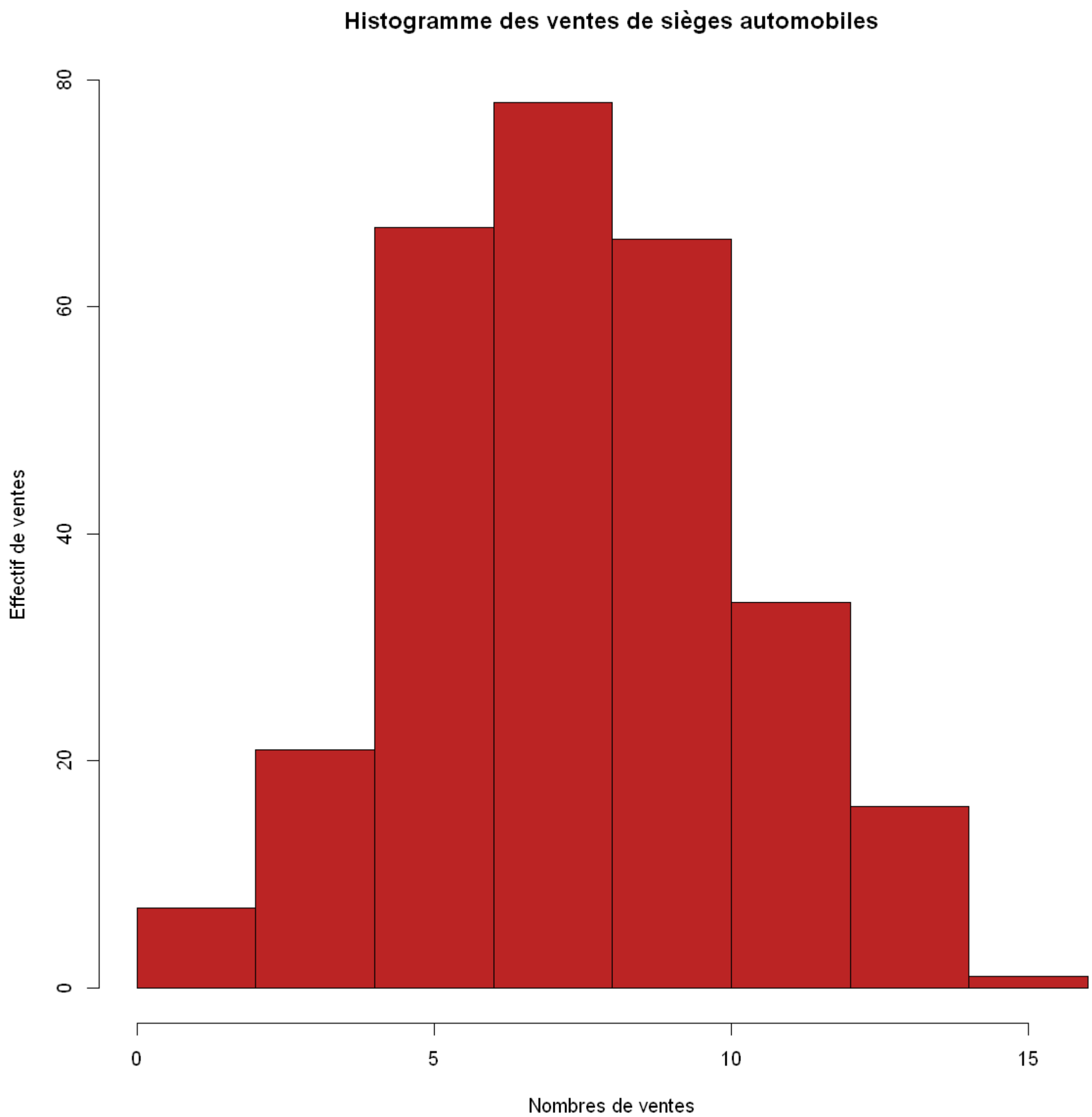
On charge toutes les données depuis le csv.

A data.frame: 5 × 5

	NUM	Sales	Price	Advertising	Region
	<int>	<dbl>	<int>	<int>	<int>
1	142	5.40	163	13	0
2	104	7.99	99	0	1
3	103	4.21	137	14	0
4	274	4.34	111	0	0
5	286	6.42	126	5	1

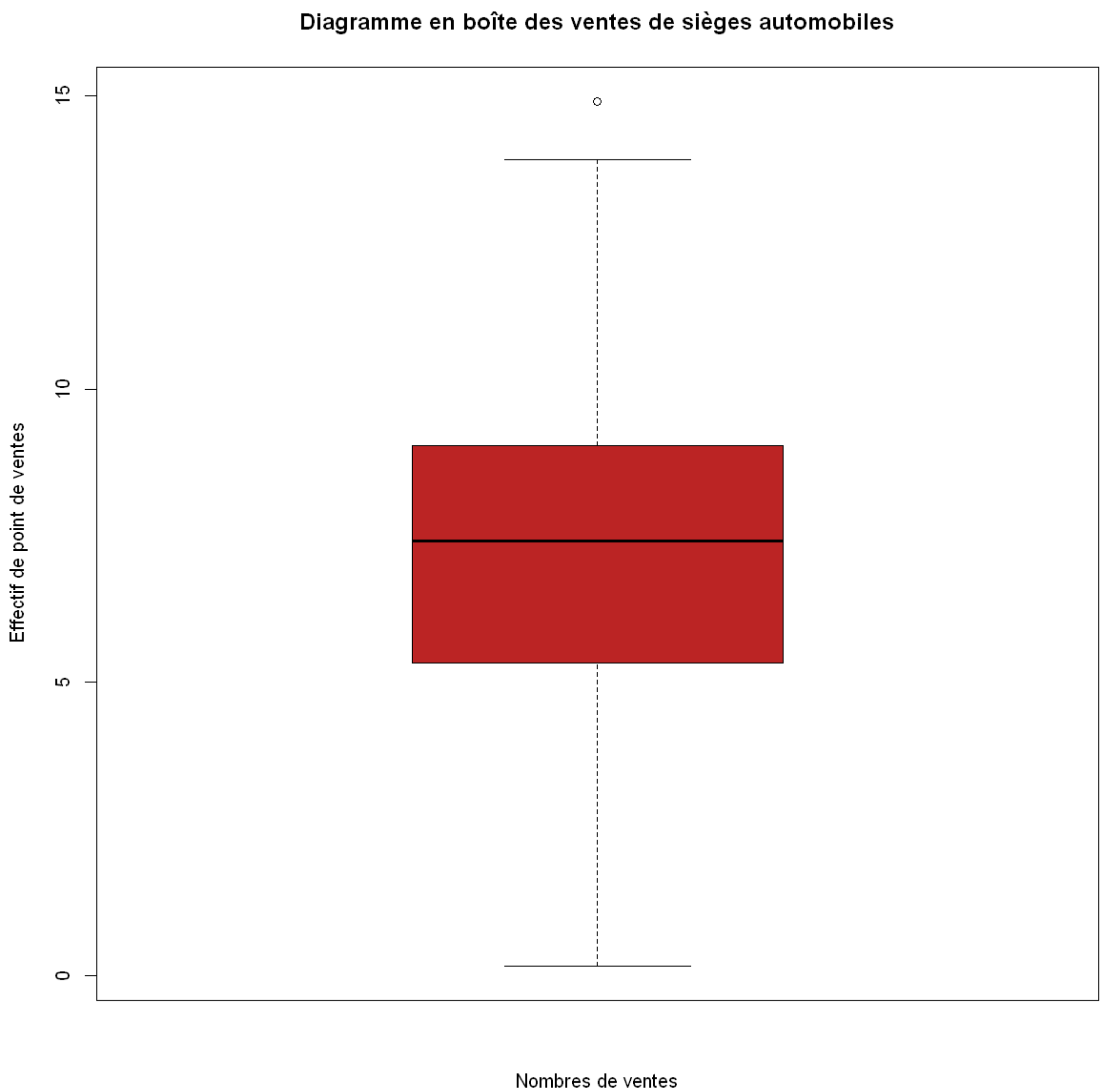
a)

Histogramme du nombre de ventes de siège automobiles



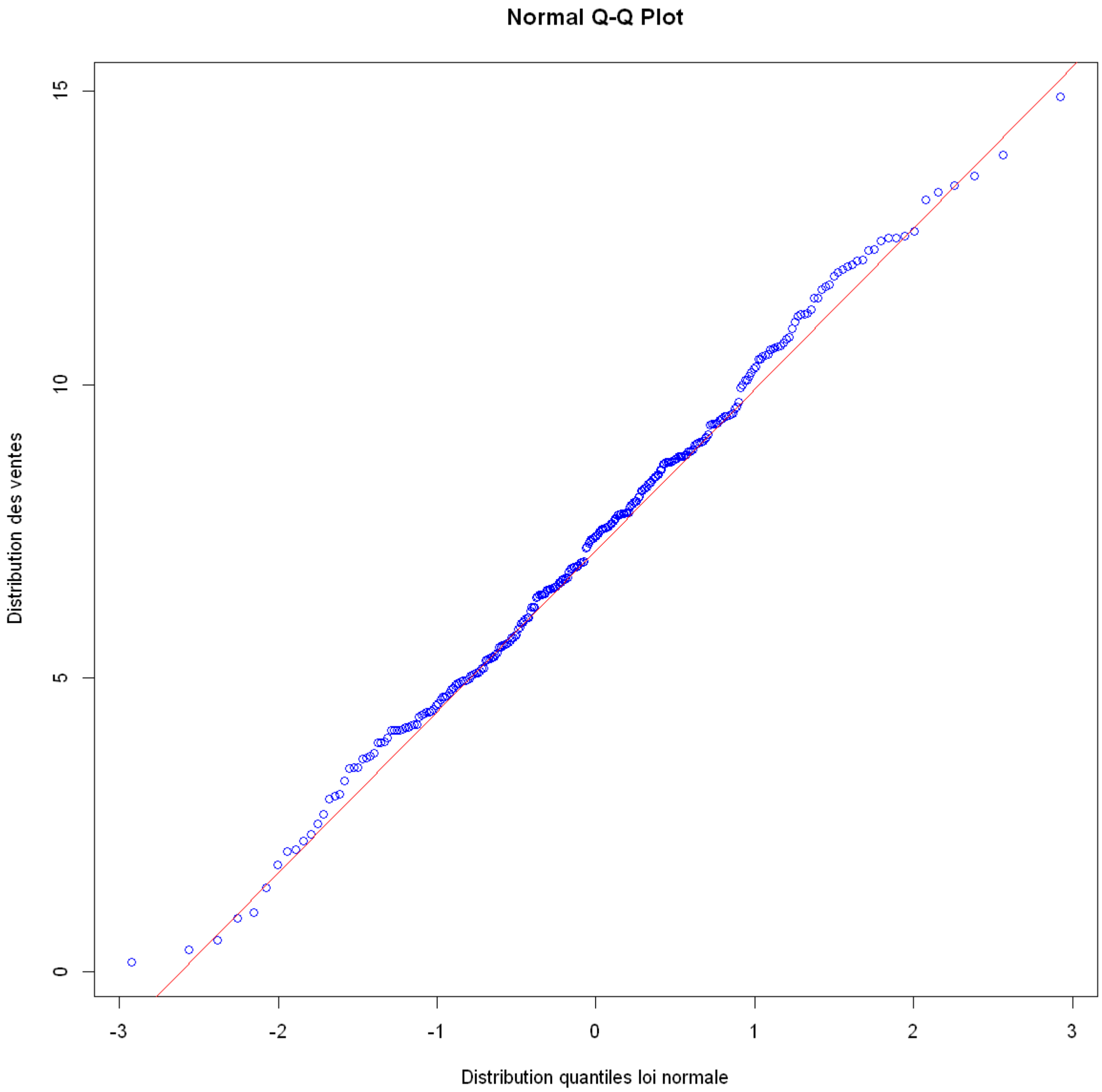
Ce graphique montre que la majorité des points de vente ont vendu entre 5 et 10 sièges automobiles.

Box plot du nombre de ventes de siège automobiles



Ce graphique montre que la moiter des points de vente ont vendu entre 5 et 10 sièges automobiles et que la médiane est d'environ 7,5.

Droite de Henry du nombre de ventes de siège automobiles



Les données de ventes semblent suivre une droite de Henry. On peut alors dire que la loi normale est une
bonne approximation de la distribution des ventes de sièges automobiles.

Test de normalité du nombre de ventes de siège automobiles

H_0: Les données suivent une loi normale
H_1: Les données ne suivent pas une loi normale

```
        Shapiro-Wilk normality test

data:  Ventes
W = 0.99571, p-value = 0.6102
```

Comme W = 0.99568 n'est pas petit et que p-value = 0.6051 n'est pas petit, on ne rejette pas H_0.
Ainsi le test de normalité montre que les données suivent une loi normale.

Tableau de statistiques descriptives du nombre de ventes de siège automobiles

Utilisation d'une librarie de R pour le calcule de l'intervalle de confiance à 95% pour la moyenne.

```
La marge d'erreur est  0.3186124
```

Caclule de l'intervalle de confiance à 95% pour la moyenne détailé.

```
La moyenne des ventes est : 7.346793
La variance des ventes est : 7.663499
Taille de la population : n = 290
```

Calcul intervale de confiance à 95% (1 - α = 0.95) pour la moyenne de la population.
$\bar{X}$ = 7.343 et σ^2 = 7.71, de plus Ventes ~ N(μ, σ^2).
On a donc: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ ~ N(0, 1) soit $\mu \in \bar{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

Calcul $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

```
La marge d'erreur est : 0.3186124
```

On à donc: $\mu \in \bar{X} \pm 0.3195956$ soit $\mu \in [7.023404, 7.662596]$ à 95%.

Tableau de statistiques :

A matrix: 1 × 6 of type chr

1er Quartile	Médiane	Moyenne	Écart type	3e Quartile	Intervale de confiance
5.3225	7.415	7.34679310344828	2.76830261510739	9.025	[7.023404, 7.662596]

b)

On extrait les données de ventes de sièges automobiles pour chaque région

12.61 · 9.48 · 3.02 · 11.19 · 7.64

5.58 · 5.87 · 9.5 · 7.23 · 4.42

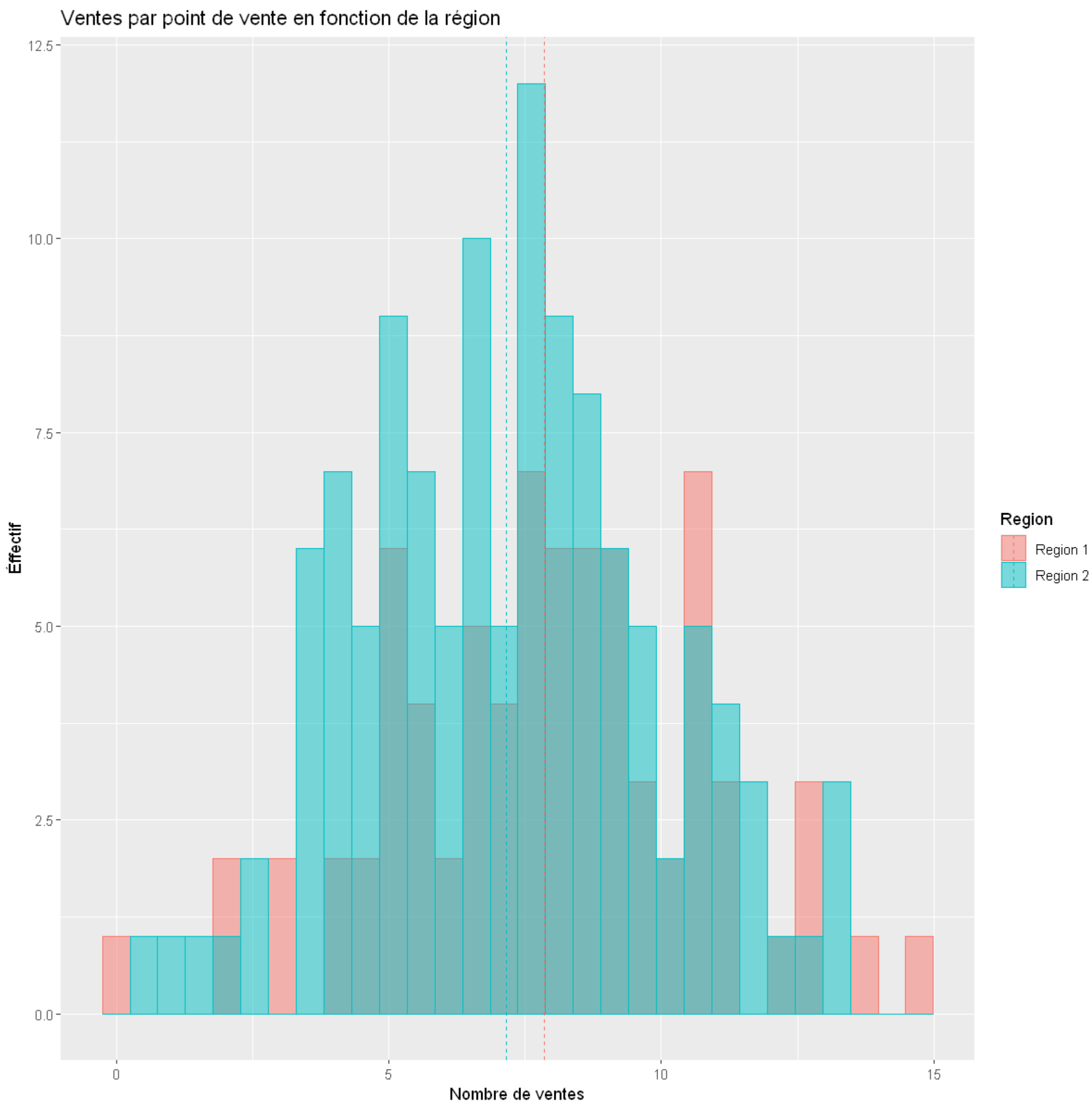
On crée un dataframe avec les données de ventes de sièges avec une colonne pour différencier les régions.

A data.frame: 5 × 2

	Region	Ventes
	<chr>	<dbl>
1	Region 1	12.61
2	Region 1	9.48
3	Region 1	3.02
4	Region 1	11.19
5	Region 1	7.64

Histogramme du nombre de ventes de siège automobiles par point de vente en fonction de la région

```
Warning message:
"le package 'ggplot2' a été compilé avec la version R 4.2.3"
```



Box plot du nombre de ventes de siège automobiles en fonction de la région

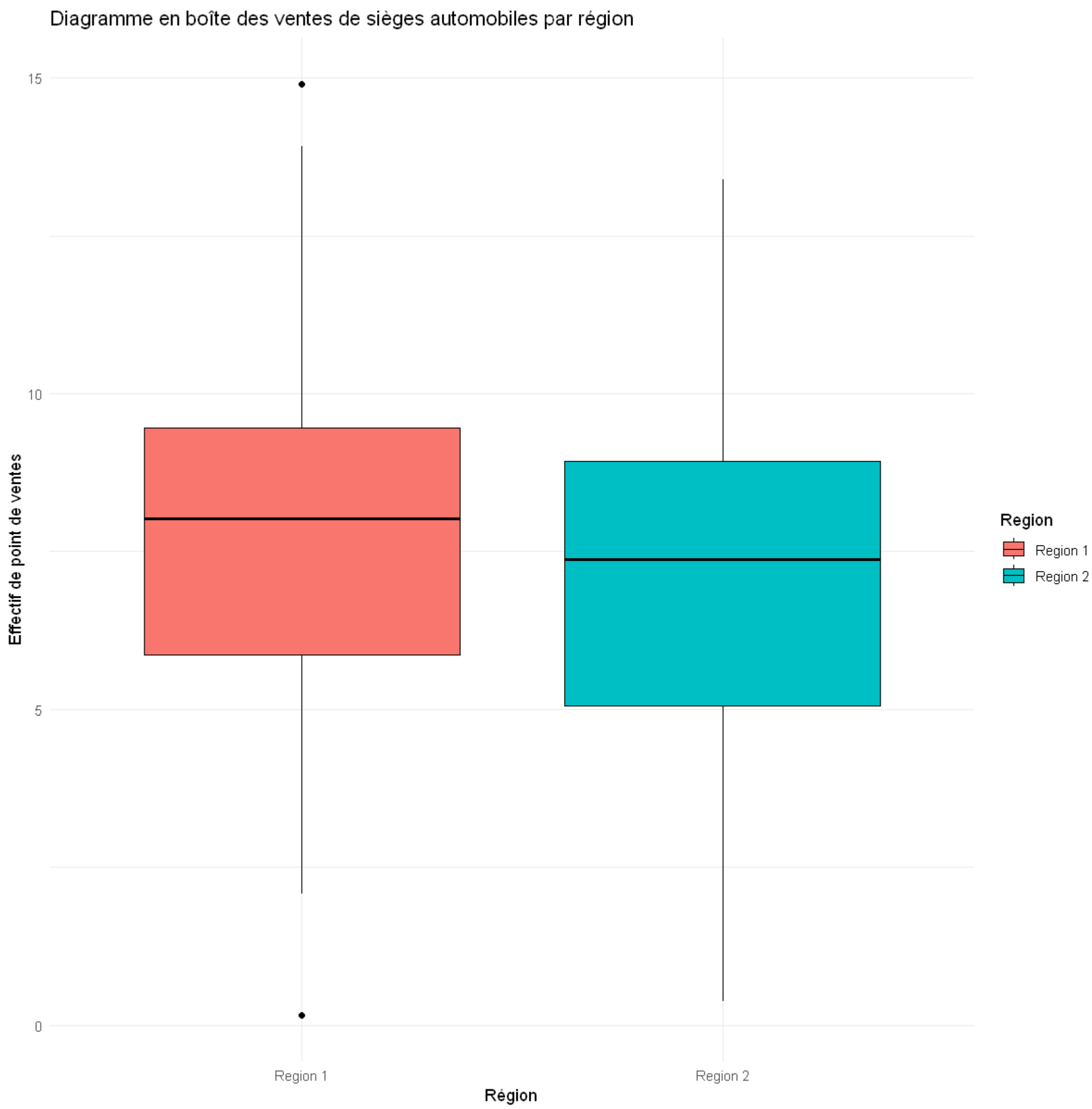


Tableau de statistiques descriptives du nombre de ventes pas point de vente en fonction de la région

A matrix: 2 × 7 of type chr

Région	1er Quartile	Médiane	Moyenne	Écart type	3e Quartile	Intervalle de confiance
0	5.8675	8.02	7.85736842105263	2.82292726009536	9.4575	[7.223, 8.492]
1	5.06	7.37	7.14941176470588	2.74392558520608	8.935	[6.656, 7.642]

Ainsi on peut voir les moyennes de ventes en fonction de la région ne diffère pas significativement. La région 1 a tout de même une moyenne un peu plus élevé.

Test d'hypotheses sur la variance des deux régions

$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 \neq \sigma_2^2$

```
	F test to compare two variances

data:  Ventes by Region
F = 1.0215, num df = 117, denom df = 171, p-value = 0.8919
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7353614 1.4340909
sample estimates:
ratio of variances 
          1.021504 
La p-value est : 0.891923
```

Comme p-value = 0.891923 > $\alpha = 0.05$, on ne rejette pas H_0 donc le test sur les variances montre que les deux régions ont des variances que ne diffèrent pas significativement au seuil α = 5%.

Démarche détaillée :

$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 \neq \sigma_2^2$
$f_0 : \frac{s_1^2}{s_2^2}$

On rejete H_0 si $F_{1-\frac{\alpha}{2}}, n_1 - 1, n_2 - 1 > f_0$ ou $F_{\frac{\alpha}{2}}, n_1 - 1, n_2 - 1 < f_0$

1.05841190653207

$F_{1-\frac{\alpha}{2}}, n_1 - 1, n_2 - 1$ = 0.71 < f_0 = 1.021, On ne peut conclure pour l'insant.

0.656037994453261

$F_{\frac{\alpha}{2}}, n_1 - 1, n_2 - 1$ = 1.389 > f_0 = 1.021, Les critères de rejet ne sont pas respectés donc H_0 est accepté.

1.49542836708511

Test d'hypotheses sur l'égalité des moyennes des deux régions

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

```
	Welch Two Sample t-test

data:  Ventes by Region
t = 1.146, df = 249.83, p-value = 0.2529
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to
0
95 percent confidence interval:
 -0.2728957  1.0324287
sample estimates:
mean in group 0 mean in group 1 
       7.572034        7.192267 
```

Comme p-value = 0.2529 > $\alpha = 0.05$, on ne rejette pas H_0 donc le test sur les moyennes montre que les deux
régions ont des moyennes que ne diffèrent pas significativement au seuil α = 5%.

Phase 2 : Recherche du meilleur modèle

c)

Modèle 1 --- $Y = \beta_0 + \beta_1 X_1 + \epsilon$

Tableau coefficient de regression

A matrix: 2 × 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	13.99532506	0.990598399	14.128152	1.389760e-31
price	-0.05719668	0.008480086	-6.744823	1.735095e-10

Tableau analyse de variance

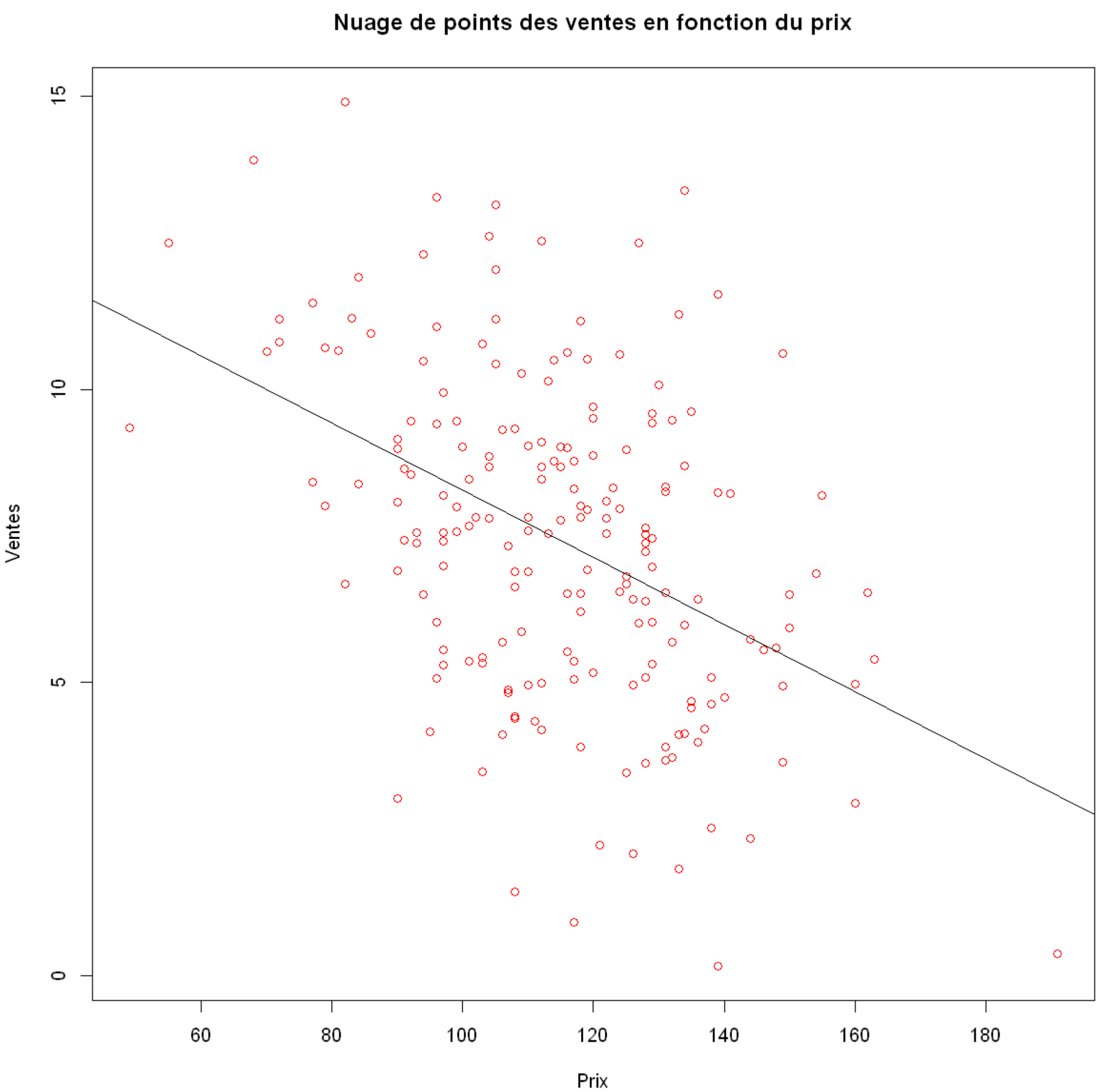
A anova: 2 × 5

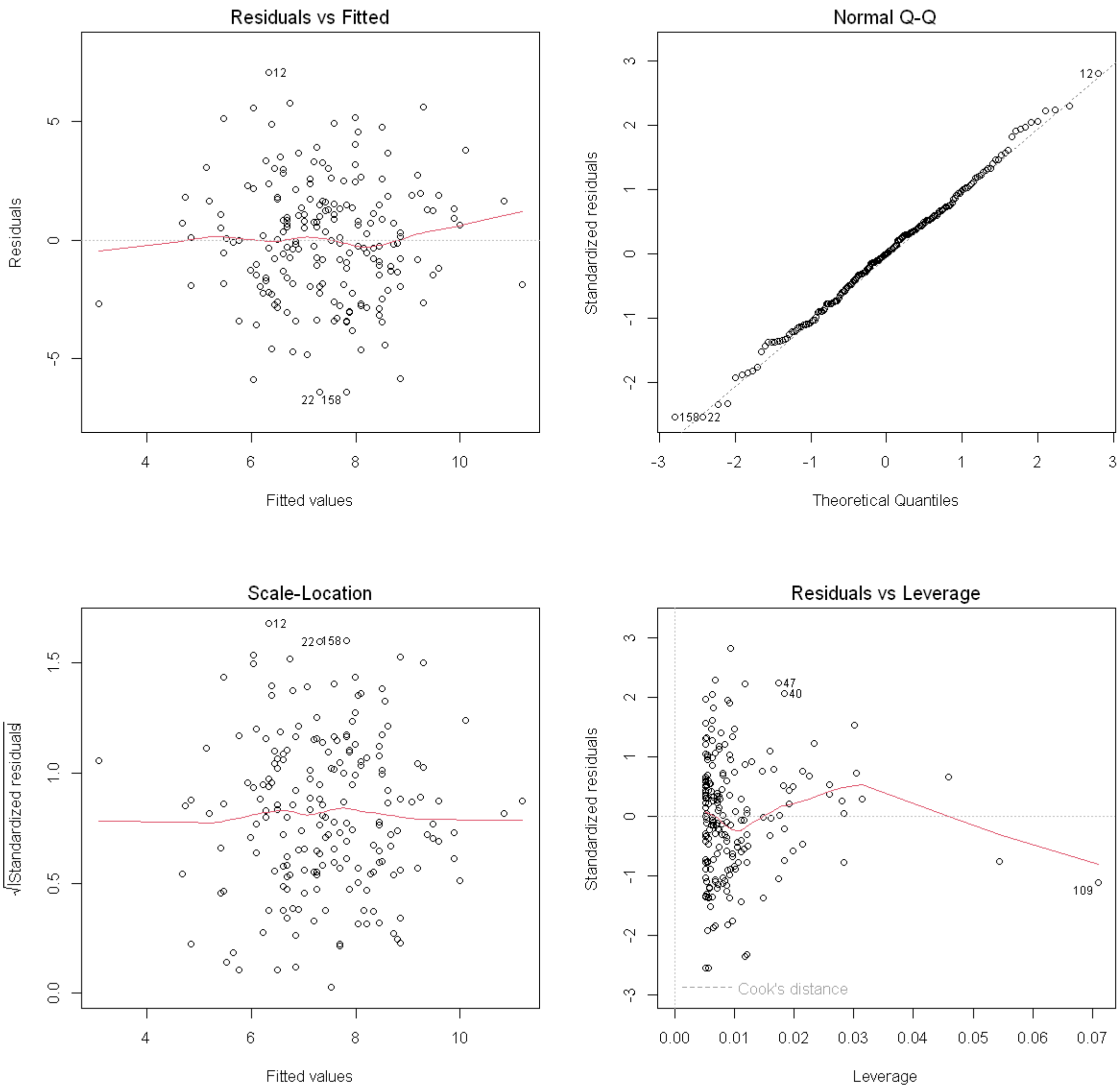
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
price	1	287.9098	287.909849	45.49264	1.735095e-10
Residuals	193	1221.4416	6.328713	NA	NA

Tester la signification du modèle

```
Call:
lm(formula = sales ~ price)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.3981 -1.8418 -0.0021  1.5515  7.0590 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 13.99532    0.99060  14.128  < 2e-16 ***
price       -0.05720    0.00848  -6.745 1.74e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.516 on 193 degrees of freedom
Multiple R-squared:  0.1908,    Adjusted R-squared:  0.1866 
F-statistic: 45.49 on 1 and 193 DF,  p-value: 1.735e-10
```

Nuage de point du modèle 1





Test significativité du modèle 1 :

Hypothèse $H_0 : \beta_1 = 0$ et $H_1 : \beta_1 \neq 0$

On a p-value = 1.735e-10 < $\alpha = 0.05$ On rejete H_0

On à donc que le modèle est significatif au seuil $\alpha = 5\%$

Évaluation validité du modèle 1 :

Le modèle n'est pas très valide car R^2 = 0.1908 or plus R^2 est proche de 1, plus la variabilité des valeurs est expliqué par le modèle.

Analyse des résidus du modèle 1 :

- Les résidus suivent la droite de normalité, l'hypothèse de normalité est donc respectée.
- Les résidus sont répartis de façon homogène autour de 0 dans l'intervalle [4, 10] et on a une homoscédasticité des valeurs qui est valide.
- le modèle à trois point atipiques qui peuvent fausser le modèle. On peut les supprimer pour améliorer le modèle.

Intervale de confiance β_0 et β_1 :

A matrix: 2 × 2 of type dbl

	2.5 %	97.5 %
(Intercept)	12.04153646	15.94911366
price	-0.07392222	-0.04047114

À 95% on a $\beta_0 \in [12.12653560, 15.0994715]$ et $\beta_1 \in [-0.06686684, -0.0416163]$.

Modèle 2 --- $Y = \beta_0 X_1^{\beta_1} e^{\epsilon}$

Équation transformée : $ln(Y) = ln(\beta_0) + \beta_1 ln(X_1) + \epsilon$

Tableau coefficient de regression

A matrix: 2 × 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.027436	0.8776213	8.007368	1.076999e-13
log(mondata$Price)	-1.084843	0.1855614	-5.846276	2.112914e-08

Tableau analyse de variance

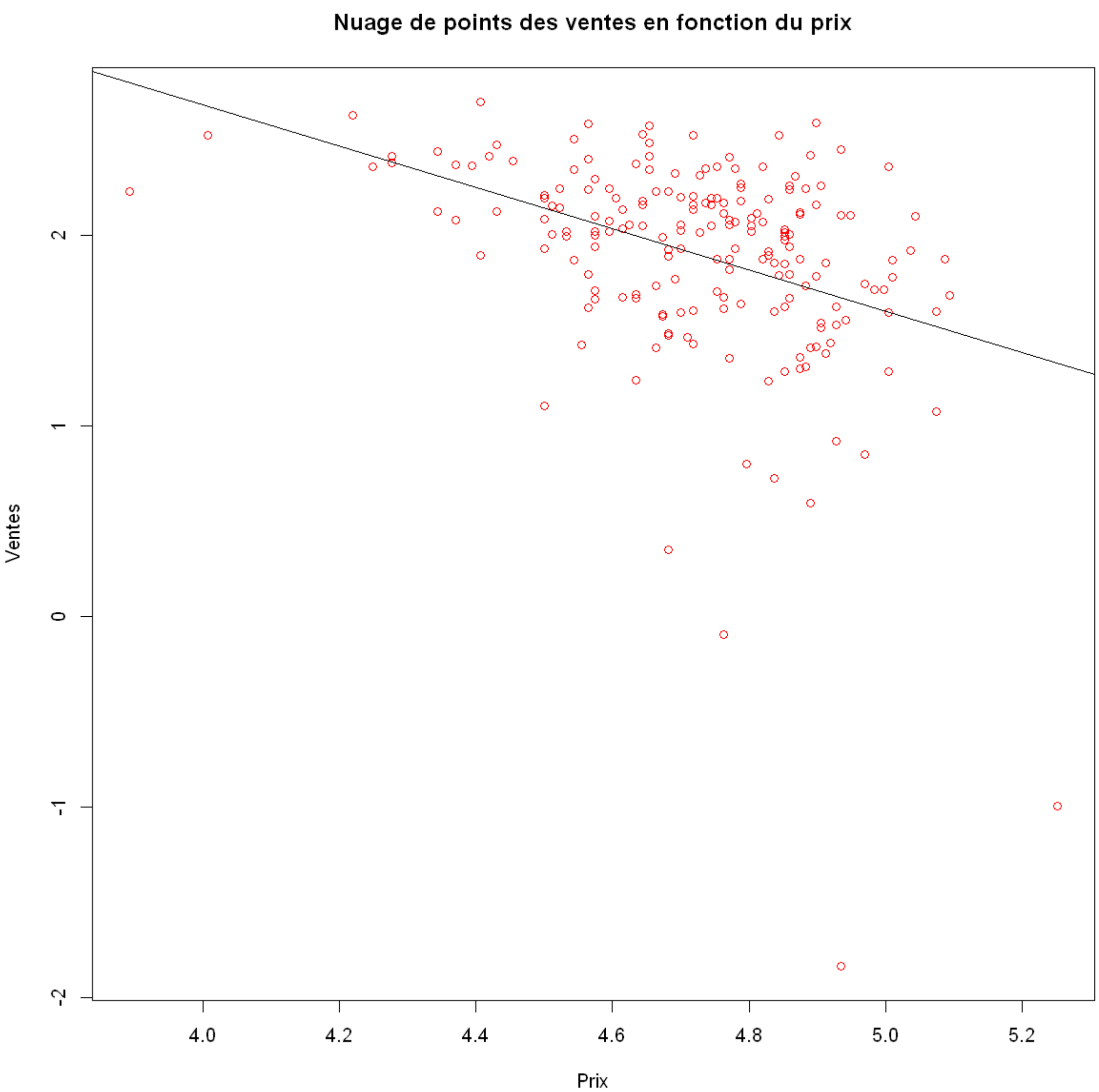
A anova: 2 × 5

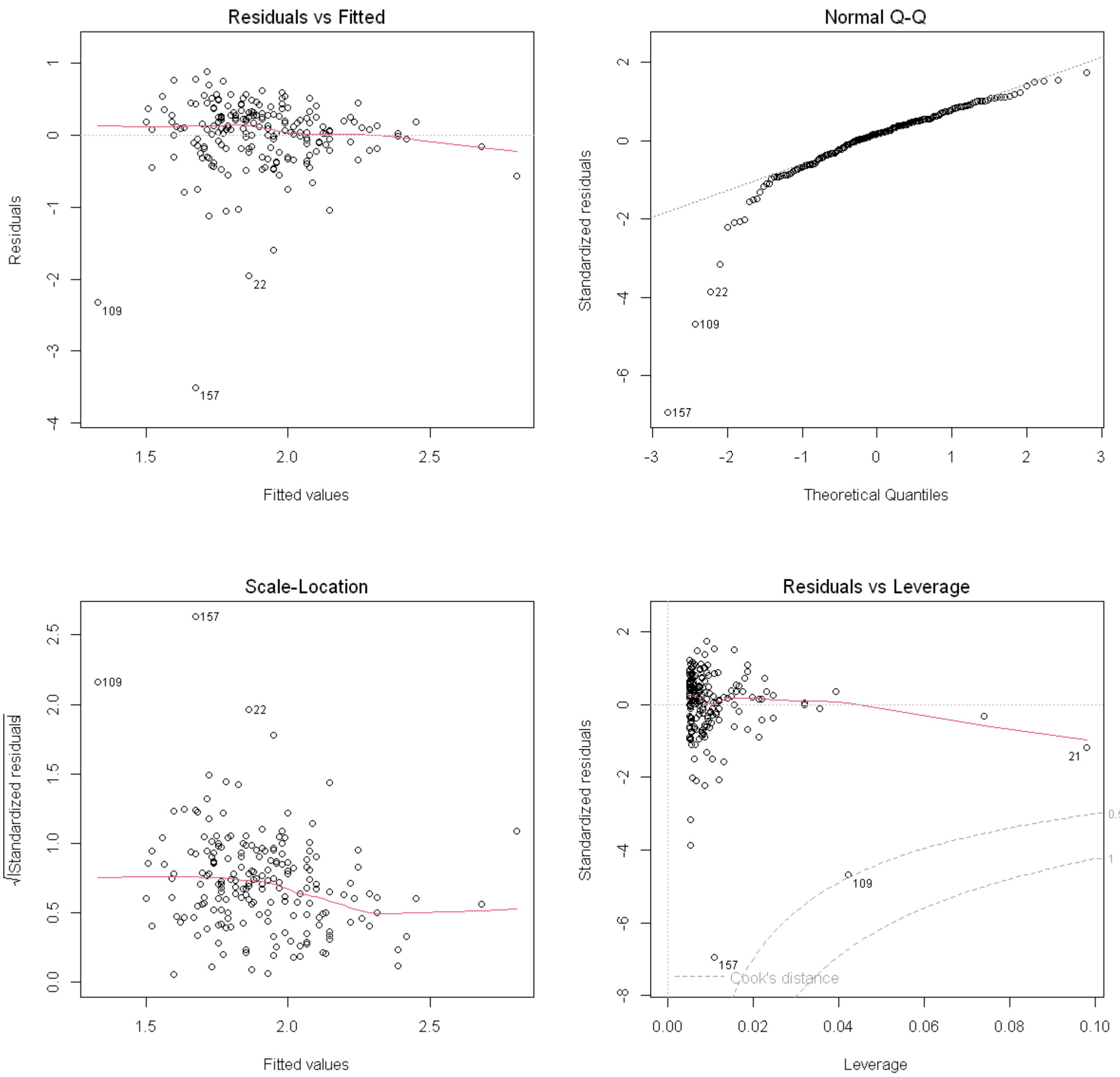
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
log(mondata$Price)	1	8.807513	8.8075133	34.17895	2.112914e-08
Residuals	193	49.733834	0.2576883	NA	NA

Tester la signification du modèle

```
Call:
lm(formula = log(mondata$Sales) ~ log(mondata$Price))

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5069 -0.1872  0.0839  0.2781  0.8805 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)          7.0274     0.8776   8.007 1.08e-13 ***
log(mondata$Price)  -1.0848     0.1856  -5.846 2.11e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5076 on 193 degrees of freedom
Multiple R-squared:  0.1504,    Adjusted R-squared:  0.146 
F-statistic: 34.18 on 1 and 193 DF,  p-value: 2.113e-08
```

Nuage de point du modèle 2





Test significativité du modèle 1 :

Hypothèse $H_0 : \beta_1 = 0$ et $H_1 : \beta_1 \neq 0$

On a p-value = 2.113e-08 < $\alpha = 0.05$ On rejete H_0

On à donc que le modèle est significatif au seuil $\alpha = 5\%$

Évaluation validité du modèle 1 :

Le modèle n'est pas très valide car R^2 = 0.1504 où plus R^2 est proche de 1,
plus la variabilité des valeurs est expliqué par le modèle.

Analyse des résidus du modèle 1 :

- Les résidus suivent assez bien la droite de normalité, l'hypothèse de normalité est donc respectée.
- Les résidus sont répartis de façon homogène autour de la droite dans l'intervalle [1.5, 2.5]
- L'homoscédasticité des valeurs n'est très bonne car on observe une forme d'entonnoire vers la fin de l'intervalle

- le modèle à trois point atipiques qui peuvent fausser le modèle. On peut les supprimer pour améliorer le modèle.

Intervale de confiance β_0 et β_1 :

A matrix: 2 × 2 of type dbl

	2.5 %	97.5 %
(Intercept)	5.296476	8.7583965
log(mondata$Price)	-1.450832	-0.7188545

À 95% on a $\beta_0 \in [\mathbf{5.085692, 7.5340537}]$ et $\beta_1 \in [\mathbf{-1.192916, -0.6755866}]$.

Modèle 3 --- $Y = \beta_0 e^{\beta_1 X_1 + \epsilon}$

Équation transformée : $ln(Y) = ln(\beta_0) + \beta_1 X_1 + \epsilon$

Tableau coefficient de regression

A matrix: 2 × 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.11012244	0.197983079	15.709032	2.328086e-36
mondata$Price	-0.01052611	0.001694848	-6.210652	3.172007e-09

Tableau analyse de variance

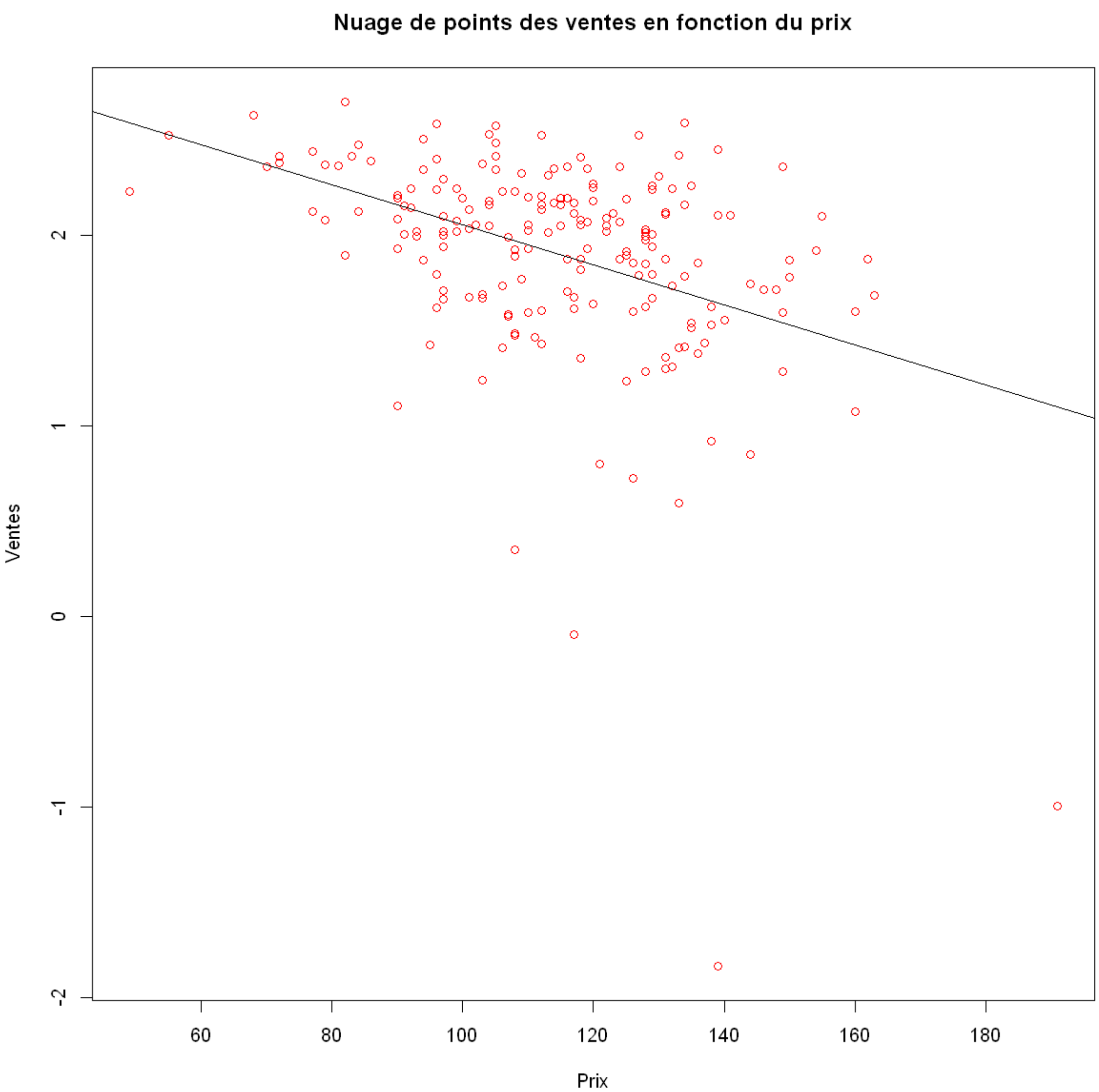
A anova: 2 × 5

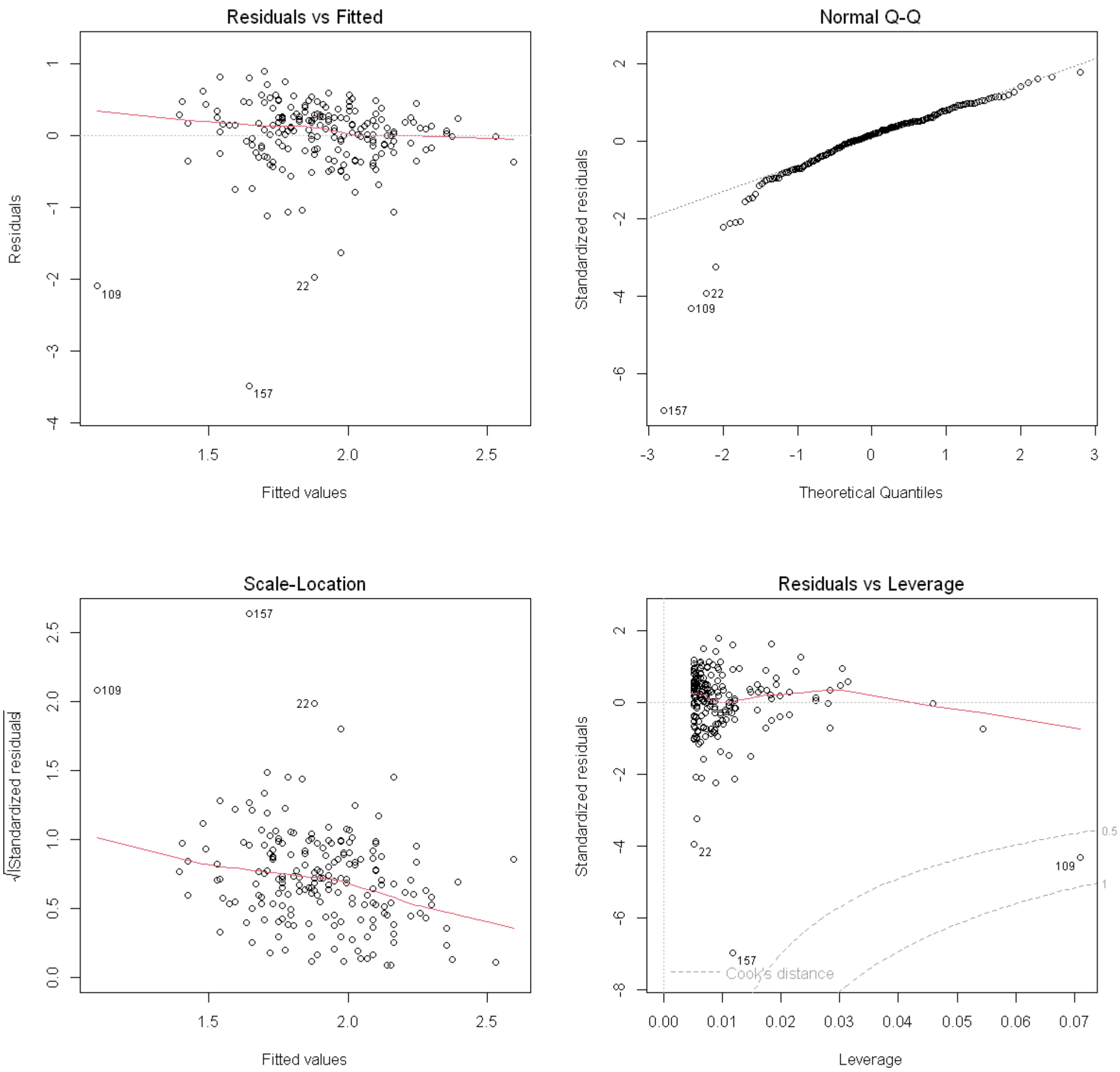
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
mondata$Price	1	9.751034	9.7510336	38.57219	3.172007e-09
Residuals	193	48.790314	0.2527996	NA	NA

Tester la signification du modèle

```
Call:
lm(formula = log(mondata$Sales) ~ mondata$Price)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.4796 -0.1954  0.0888  0.2671  0.8949 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    3.110122   0.197983  15.709  < 2e-16 ***
mondata$Price -0.010526   0.001695  -6.211 3.17e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5028 on 193 degrees of freedom
Multiple R-squared:  0.1666,	Adjusted R-squared:  0.1622 
F-statistic: 38.57 on 1 and 193 DF,  p-value: 3.172e-09
```

Nuage de point du modèle 3





Test significativité du modèle 3 :

Hypothèse $H_0 : \beta_1 = 0$ et $H_1 : \beta_1 \neq 0$

On a p-value = 3.172e-09 < $\alpha = 0.05$ On rejete H_0

On à donc que le modèle est significatif au seuil $\alpha = 5\%$

Évaluation validité du modèle 3 :

Le modèle 3 est plus représentatif que le modèle 2 mais moins que le 1, car on y retrouve R^2 = 0.1666. On sait aussi que plus R^2 est proche de 1, plus la variabilité des valeurs est expliqué par le modèle. Cependant cette valeur est assez éloignée de 1, ce qui montre que le modèle n'est pas très valide.

Analyse des résidus du modèle 3 :

- Les résidus suivent assez bien la droite de normalité, l'hypothèse de normalité est donc respectée.
- Les résidus sont répartis de façon homogène autour de la droite.
- L'homoscédasticité des valeurs est assez bonne, on observe un léger retrésissement vers la fin de l'intervalle.

- Quelques point atipiques qui peuvent fausser le modèle. On peut les supprimer pour améliorer le modèle.

Intervale de confiance β_0 et β_1 :

A matrix: 2 × 2 of type dbl

	2.5 %	97.5 %
(Intercept)	2.71963414	3.500610737
mondata$Price	-0.01386891	-0.007183307

À 95% on a $\beta_0 \in [2.75376911, 3.331969180]$ et $\beta_1 \in [-0.01240794, -0.007497019]$.

Modèle 4 --- $Y = \beta_0 + \beta_1 X_2 + \epsilon$

Tableau coefficient de regression

A matrix: 2 × 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.5075945	0.26827938	24.256782	1.636539e-60
mondata$Advertising	0.1409127	0.02920756	4.824527	2.842981e-06

Tableau analyse de variance

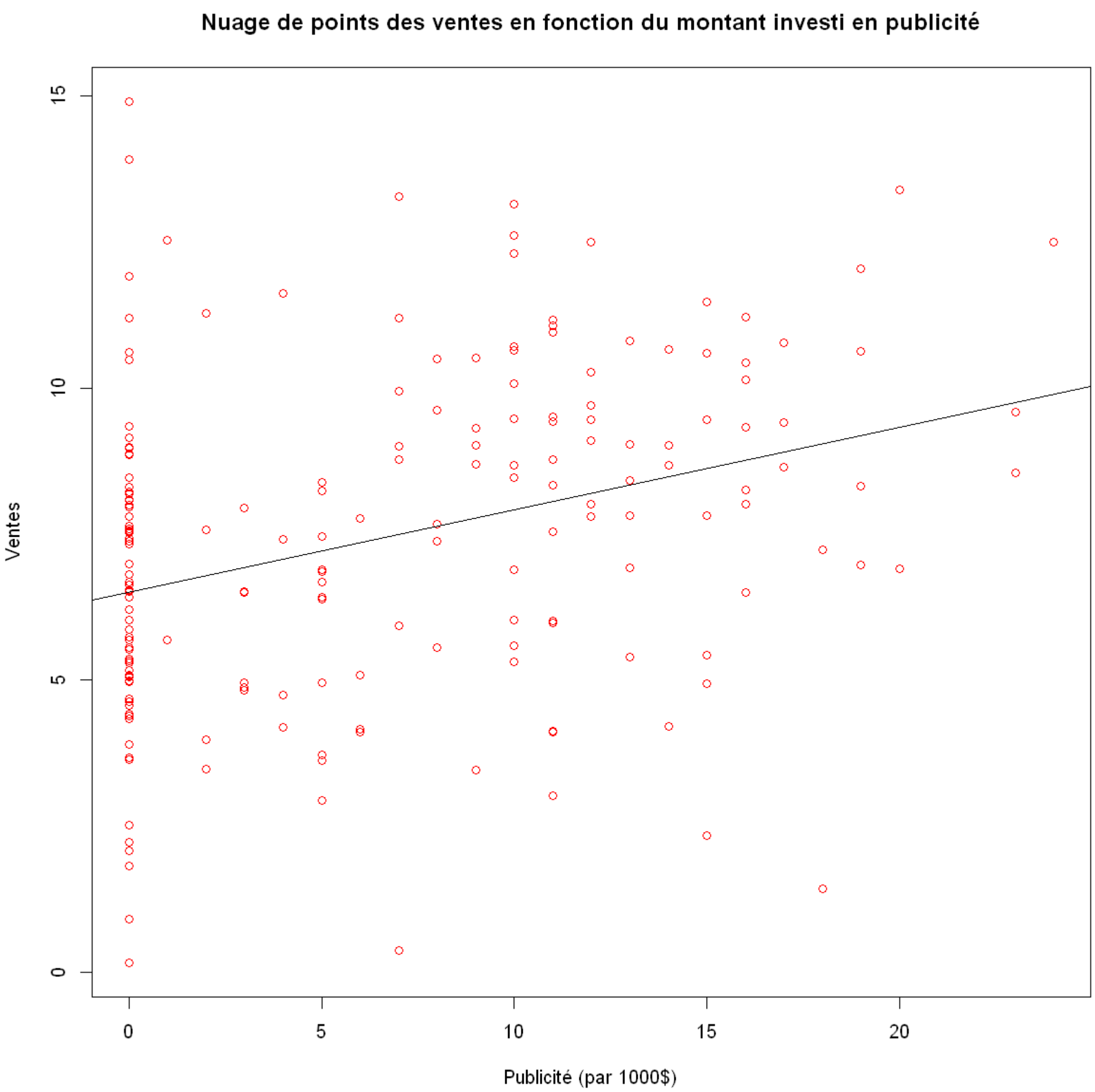
A anova: 2 × 5

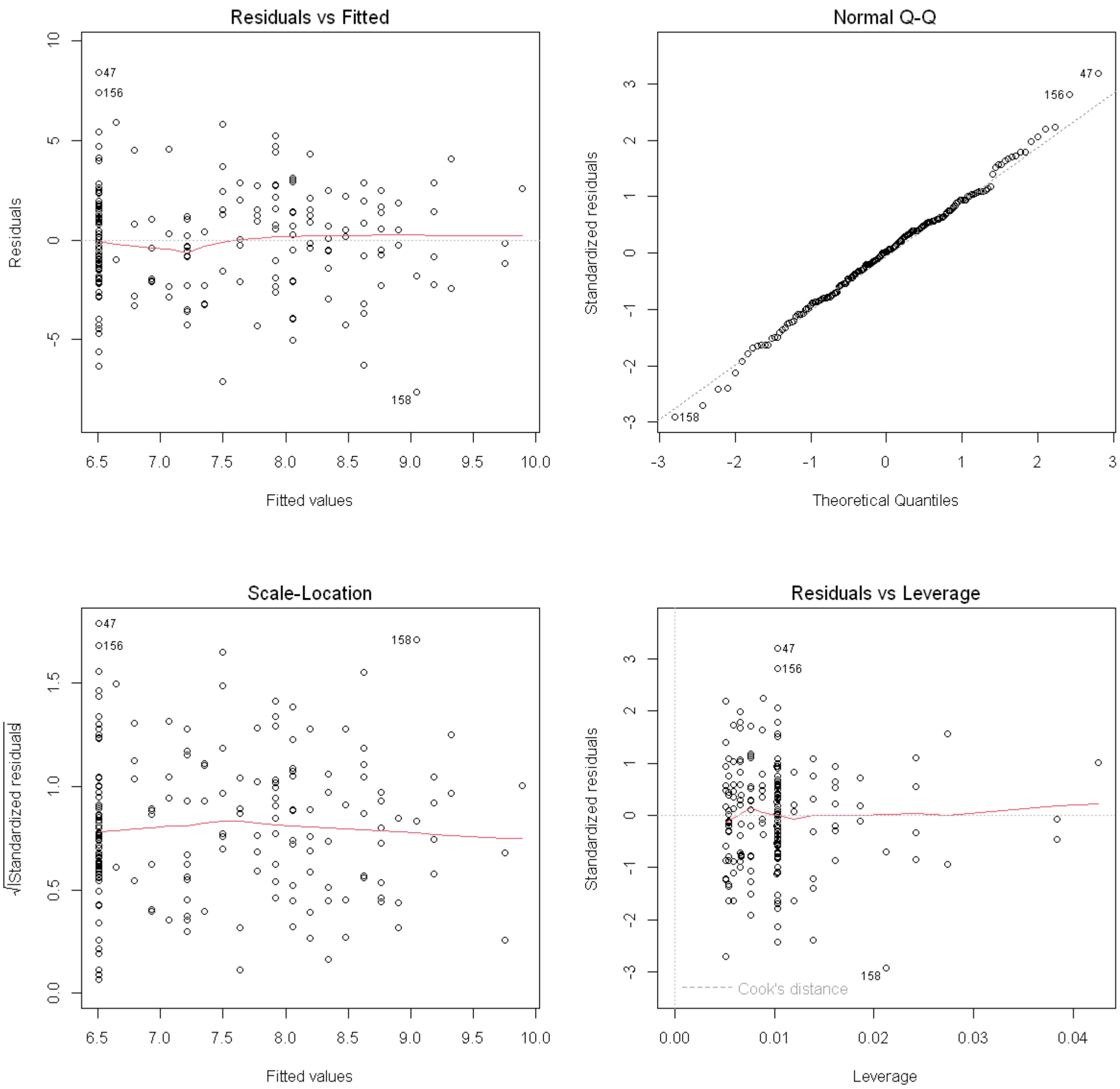
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
mondata$Advertising	1	162.4394	162.439409	23.27606	2.842981e-06
Residuals	193	1346.9120	6.978819	NA	NA

Tester la signification du modèle

```
Call:
lm(formula = mondata$Sales ~ mondata$Advertising)

Residuals:
    Min      1Q  Median      3Q     Max
-7.6240 -1.8572  0.0351  1.5628  8.3924

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          6.50759    0.26828  24.257  < 2e-16 ***
mondata$Advertising  0.14091    0.02921   4.825 2.84e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.642 on 193 degrees of freedom
Multiple R-squared:  0.1076,    Adjusted R-squared:  0.103
F-statistic: 23.28 on 1 and 193 DF,  p-value: 2.843e-06
```

Nuage de point du modèle 4





Test significativité du modèle 4 :

Hypothèse $H_0 : \beta_1 = 0$ et $H_1 : \beta_1 \neq 0$

On a p-value = 2.843e-06 < $\alpha = 0.05$ On rejete H_0

On à donc que le modèle est significatif au seuil $\alpha = 5\%$

Évaluation validité du modèle 4 :

Le modèle 4 pas très valide par rar rapport aux autres modèles, car on y retrouve R^2 = 0.1076.

On sait aussi que plus R^2 est proche de 0, moins la variabilité des valeurs est pas expliqué par le modèle.

Analyse des résidus du modèle 4 :

- Les résidus suivent bien la droite de normalité, l'hypothèse de normalité est donc respectée.
- Les résidus ne sont pas répartis de façon homogène autour de la droite,

en effet on observe une forte concentration de points au début.

- L'homoscédasticité des valeurs est plutôt bonne.

- Deux points atipiques peuvent fausser le modèle. On pourrait les supprimer pour améliorer les résultats.

Intervale de confiance β_0 et β_1 :

A matrix: 2 × 2 of type dbl

	2.5 %	97.5 %
(Intercept)	5.97845859	7.0367304
mondata$Advertising	0.08330566	0.1985196

À 95% on a $\beta_0 \in [6.07514639, 6.9209938]$ et $\beta_1 \in [0.09092425, 0.1873456]$.

Modèle 5 --- $Y = \beta_0(8 + X_2)^{\beta_1}e^{\epsilon}$

Équation transformée : $ln(Y) = ln(\beta_0) + \beta_1 ln(8 + X_2) + \epsilon$

Tableau coefficient de regression

A matrix: 2 × 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.0880022	0.22274236	4.884577	2.170394e-06
log(8 + mondata$Advertising)	0.3156951	0.08521537	3.704673	2.763175e-04

Tableau analyse de variance

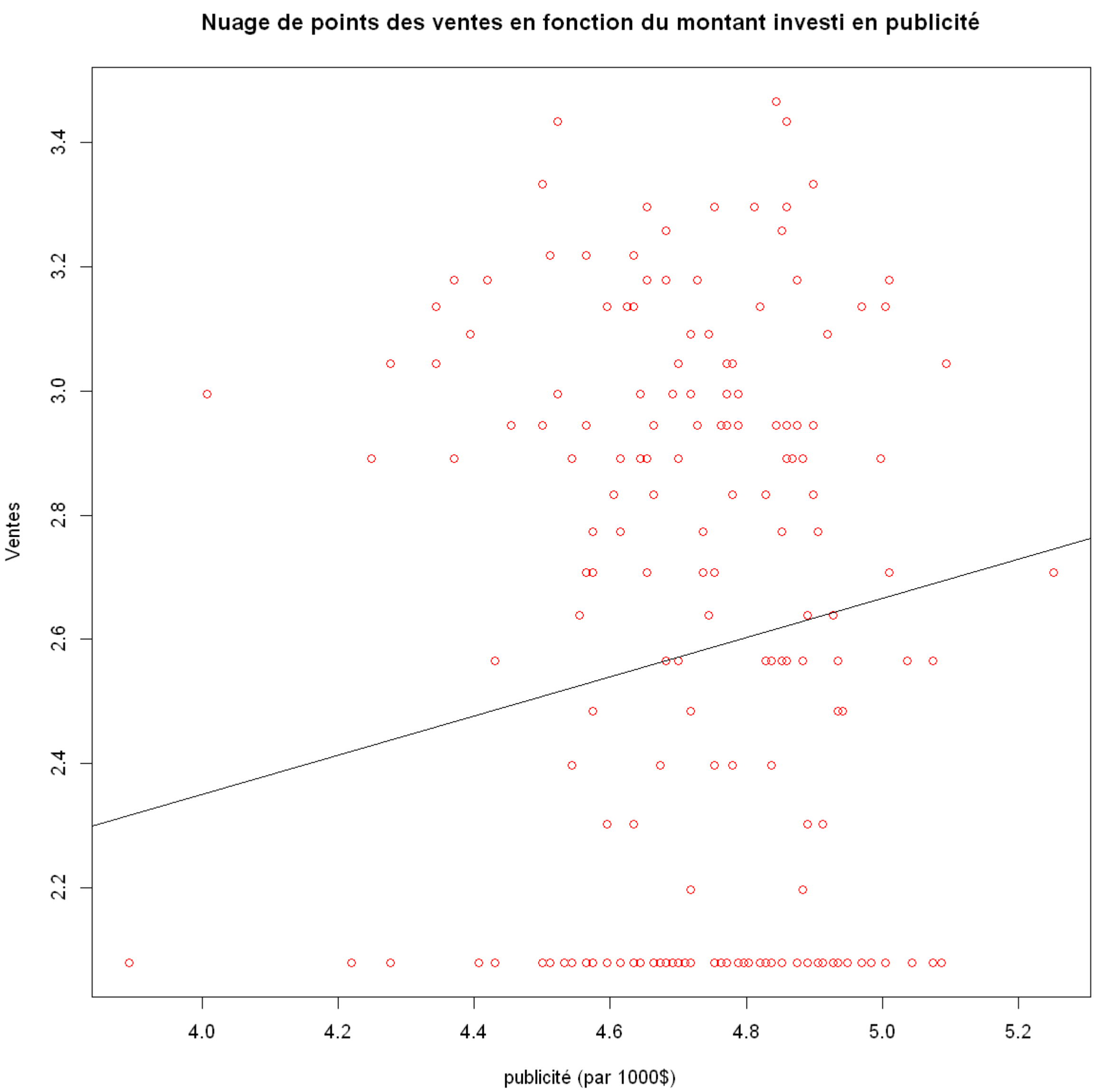
A anova: 2 × 5

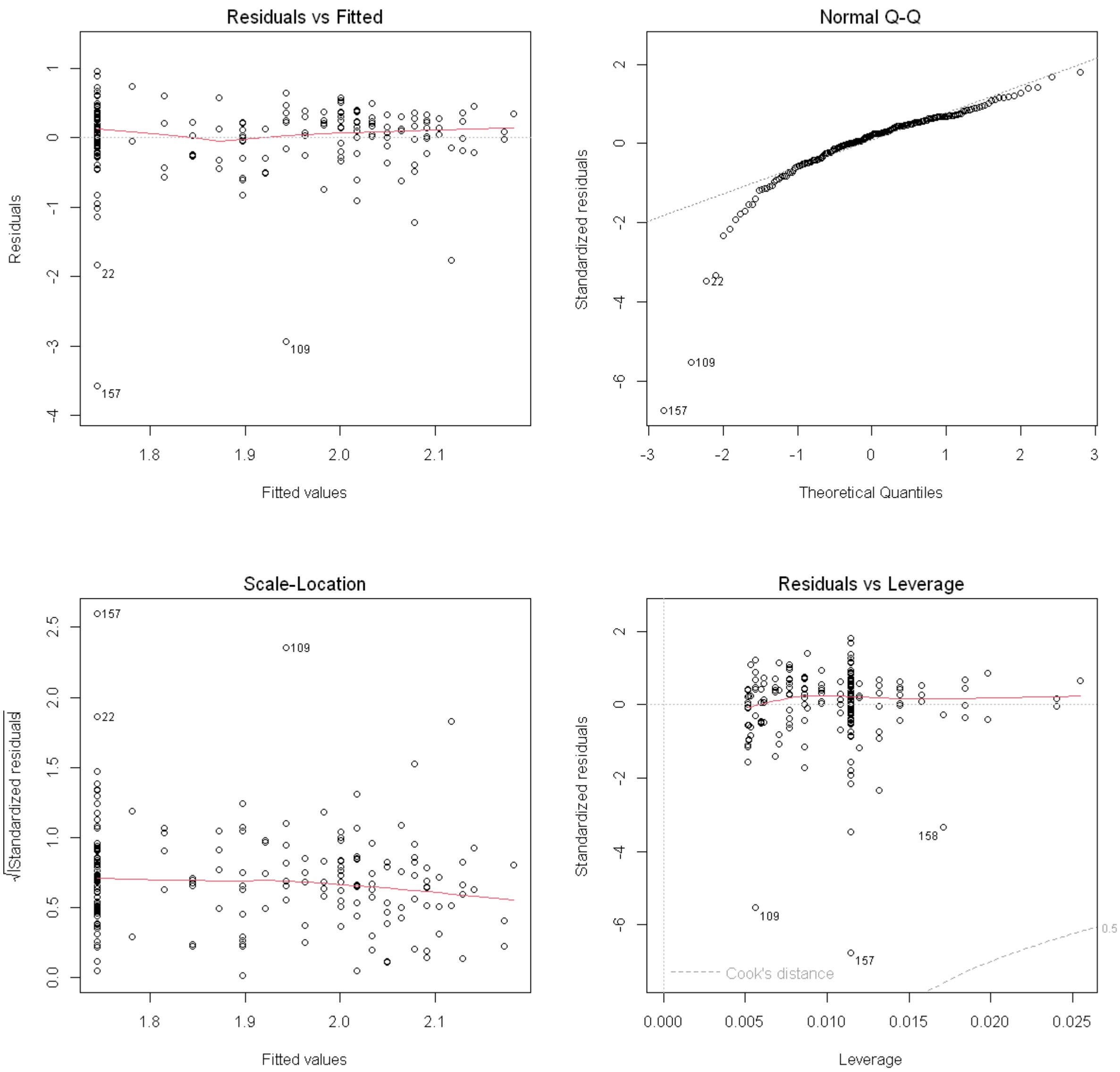
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
log(8 + mondata$Advertising)	1	3.886605	3.8866048	13.7246	0.0002763175
Residuals	193	54.654743	0.2831852	NA	NA

Tester la signification du modèle

```
Call:
lm(formula = log(mondata$Sales) ~ log(8 + mondata$Advertising))

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5771 -0.1940  0.1105  0.2922  0.9569 

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                   1.08800    0.22274   4.885 2.17e-06 ***
log(8 + mondata$Advertising)  0.31570    0.08522   3.705 0.000276 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5322 on 193 degrees of freedom
Multiple R-squared:  0.06639,	Adjusted R-squared:  0.06155 
F-statistic: 13.72 on 1 and 193 DF,  p-value: 0.0002763
```

Nuage de point du modèle 5





Test significativité du modèle 5 :

Hypothèse $H_0 : \beta_1 = 0$ et $H_1 : \beta_1 \neq 0$

On a p-value = 0.0002763 < $\alpha = 0.05$ On rejete H_0

On à donc que le modèle est significatif au seuil $\alpha = 5\%$

Évaluation validité du modèle 5 :

Le modèle 5 est peu représentatif comparativement aux autres modèles, car R^2 = 0.06639.

On sait aussi que plus R^2 est proche de 0, moins la variabilité des valeurs est expliqué par le modèle.

Analyse des résidus du modèle 5 :

- Les résidus suivent assez bien la droite de normalité, l'hypothèse de normalité est donc respectée.
- Les résidus ne sont pas répartis de façon homogène autour de la droite,

en effet on observe une forte concentration de points au début et un peu sur la fin.

- L'homoscédasticité des valeurs est plutôt bonne mais si on peut noté un très léger rétrécissement sur la fin.
- Quelques points atipiques peuvent fausser le modèle. On pourrait les supprimer pour améliorer les résultats.

Intervale de confiance β_0 et β_1 :

A matrix: 2 × 2 of type dbl

	2.5 %	97.5 %
(Intercept)	0.6486804	1.5273240
log(8 + mondata$Advertising)	0.1476221	0.4837681

À 95% on a $\beta_0 \in [0.7805855, 1.480652]$ et $\beta_1 \in [0.1638894, 0.434608]$.

Modèle 6 --- $Y = \beta_0 e^{\beta_1 X_2 + \epsilon}$

Équation transformée : $ln(Y) = ln(\beta_0) + \beta_1 X_2 + \epsilon$

Tableau coefficient de regression

A matrix: 2 × 4 of type dbl

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.75790007	0.054010327	32.547480	2.634337e-80
mondata$Advertising	0.02197562	0.005880101	3.737287	2.449595e-04

Tableau analyse de variance

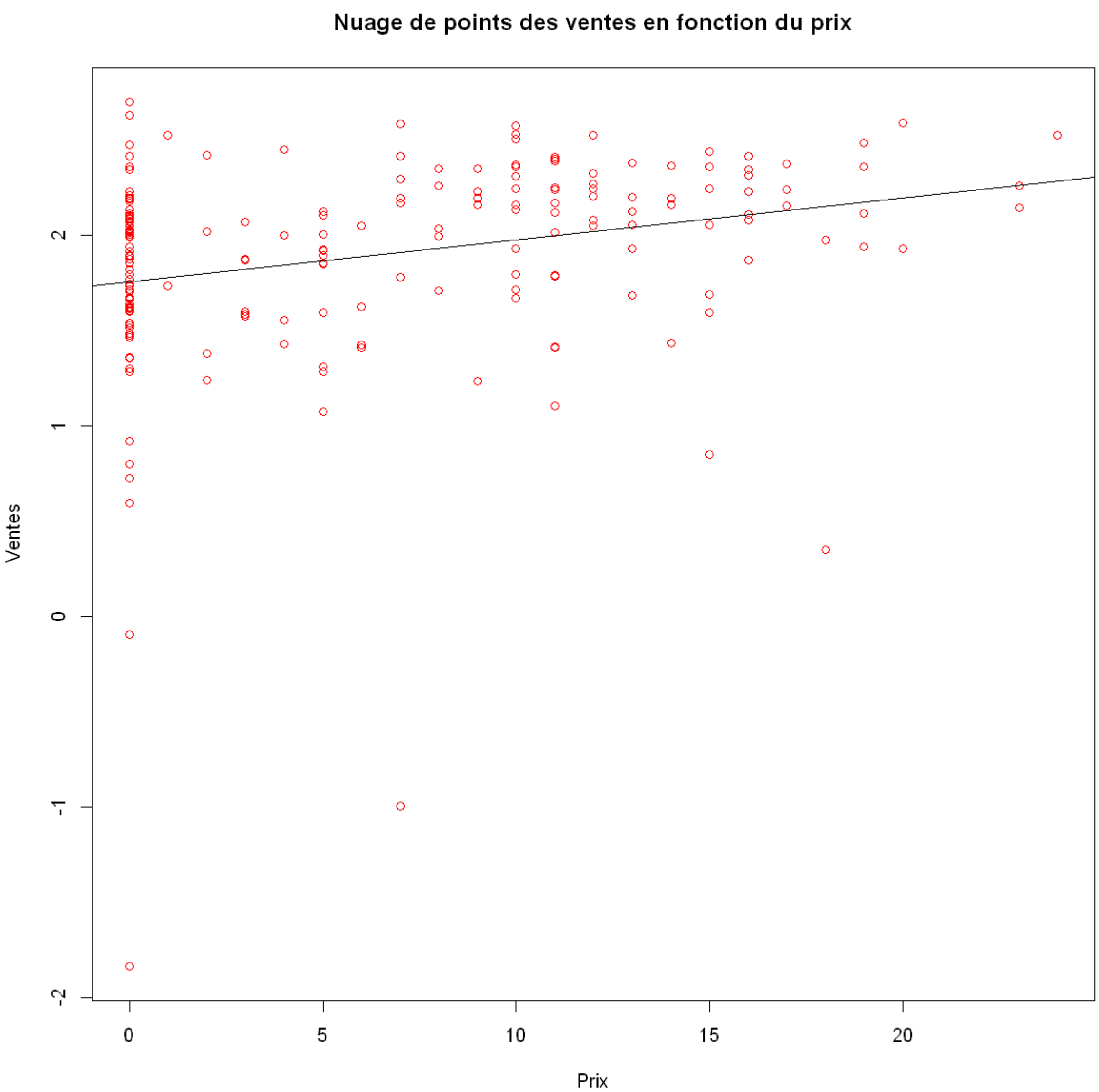
A anova: 2 × 5

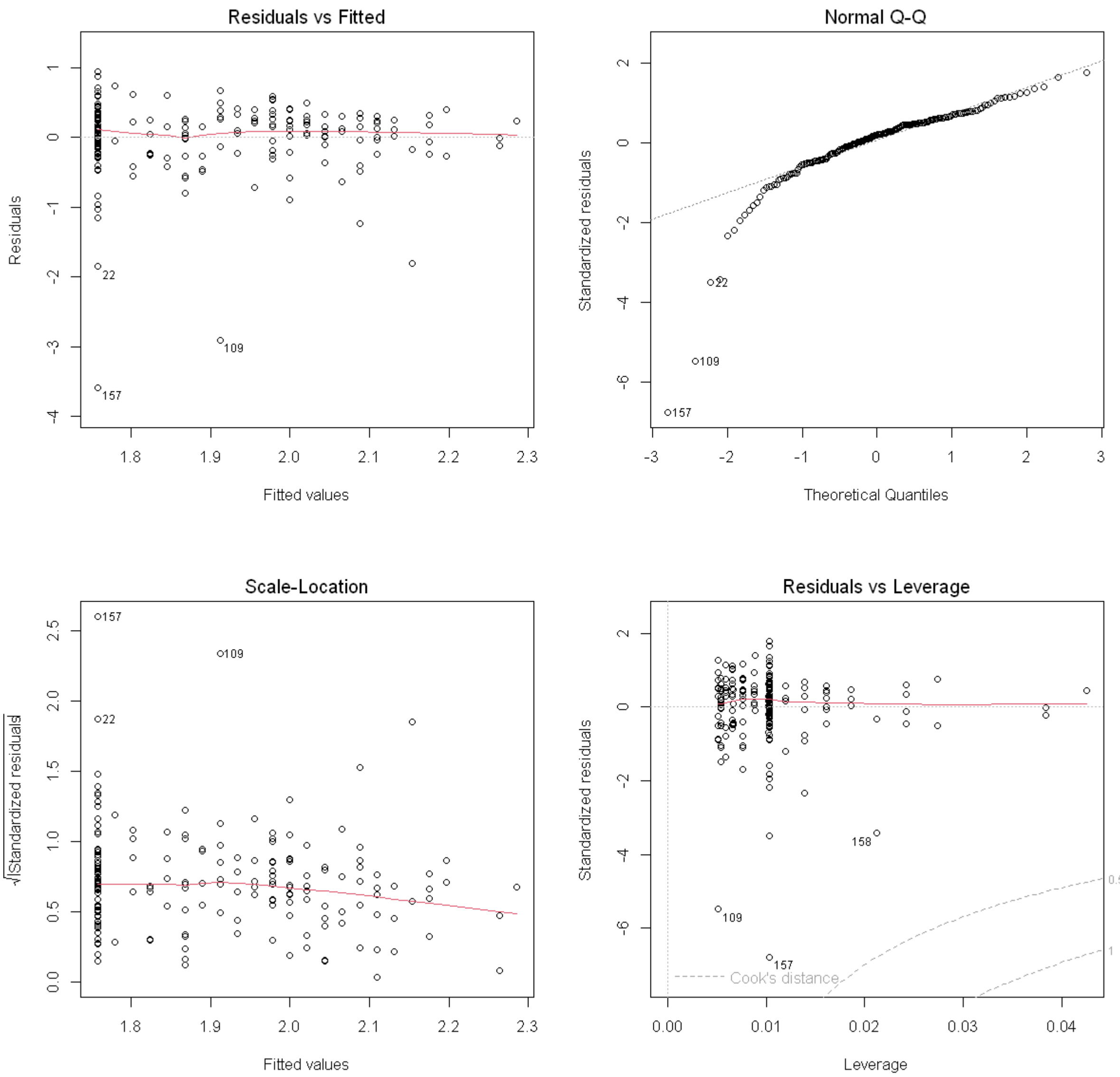
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
mondata$Advertising	1	3.950698	3.9506985	13.96732	0.0002449595
Residuals	193	54.590649	0.2828531	NA	NA

Tester la signification du modèle

```
Call:
lm(formula = log(mondata$Sales) ~ mondata$Advertising)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5905 -0.1936  0.1092  0.2805  0.9435 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)          1.75790    0.05401  32.547  < 2e-16 ***
mondata$Advertising  0.02198    0.00588   3.737 0.000245 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5318 on 193 degrees of freedom
Multiple R-squared:  0.06749,   Adjusted R-squared:  0.06265 
F-statistic: 13.97 on 1 and 193 DF,  p-value: 0.000245
```

Nuage de point du modèle 6





Test significativité du modèle 6 :

Hypothèse $H_0 : \beta_1 = 0$ et $H_1 : \beta_1 \neq 0$

On a p-value = 0.000245 < $\alpha = 0.05$ On rejete H_0

On à donc que le modèle est significatif au seuil $\alpha = 5\%$

Évaluation validité du modèle 6 :

Le modèle 6 est peu valide comparativement aux autres modèles, car R^2 = 0.06749.

On sait aussi que plus R^2 est proche de 0, moins la variabilité des valeurs est expliqué par le modèle.

Analyse des résidus du modèle 6 :

- Les résidus suivent assez bien la droite de normalité, l'hypothèse de normalité est donc respectée.
- Les résidus ne sont pas répartis de façon homogène autour de la droite,

en effet on observe une forte concentration de points au début.

- L'homoscédasticité des valeurs est plutôt bonne mais si on peut noté un très léger rétrécissement sur la fin.
- Quelques points atipiques peuvent fausser le modèle. On pourrait les supprimer pour améliorer les résultats.

Intervale de confiance β_0 et β_1 :

A matrix: 2 × 2 of type dbl

	2.5 %	**97.5 %**
(Intercept)	1.65137379	1.86442635
mondata$Advertising	0.01037812	0.03357313

À 95% on a $\beta_0 \in [1.67805637, 1.84386620]$ et $\beta_1 \in [0.01221427, 0.03111555]$.

d) Choix du meilleur modèle

Après analyse des différents modèles, on a les modèles 1 et 3 qui se destinguent des autres,
car ils sont tout deux significatifs et ont une valeur de R^2 assez élevé par rapport au autre.
Cependant, le modèle 1 est plus représentatif que le modèle 3 car il a une valeur de R^2 de 0.1989, la plus élevé.
C'est donc le modèle 1 qui represente le mieux la relation linéaire entre le prix et le nombre de ventes.

Il est tout de même important de noté qu'aucun des modèles ne peu être totalement validé,
car ils ont tous une valeur de R^2 plutôt éloignée de 1.
Ainsi bien que le modèle 1 soit le plus représentatif et valide, il ne permet pas de bien expliquer
la variabilité des valeurs du modèle.

Équation du Modèle 1 : $Y = \beta_0 + \beta_1 X_1 + \epsilon$

On utilise la valeur de X_1 pour prédire la valeur des ventes Y.

A matrix: 1 × 3 of type dbl

	fit	**lwr**	**upr**
1	7.246117	6.886953	7.605282

Ainsi avec un niveau de confiance de 95%, on peut prédire que le nombre de ventes
sera compris entre **6.886953 et 7.605282** milliers pour un prix de 118$.