

TP n° 3 : Régression linéaire simple 2

Exercice 1. On souhaite expliquer une variable quantitative Y en fonction d'une variable quantitative X_1 . Pour ce faire, on considère le modèle de *rls* :

$$Y = \beta_0 + \beta_1 X_1 + \epsilon,$$

où $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Les paramètres β_0 , β_1 et σ sont des réels inconnus. On les estime alors à l'aide de n observations de (Y, X_1) par la méthode des *mco*. Le tableau de ce modèle de *rls* renvoyé par la commande `summary` est donné ci-dessous :

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	•	19.4194	16.15	0.0000
X1	30.5364	2.0103	•	0.0000

Residual standard error: 35.12 on 17 degrees of freedom

Multiple R-squared: 0.9314, Adjusted R-squared: •

F-statistic: • on 1 and 17 DF, p-value: 2.535e-11

Les points • représente des informations volontairement effacées.

1. Quelle est la valeur de n ?
2. Donner la valeur prédite de Y lorsque $X_1 = 2$.
3. Est-ce que la régression est significative ? Si oui, est-elle "seulement" significative ?
4. Peut-on affirmer que $\beta_1 \neq 30.40$ au risque 5% ?
5. Donner un intervalle de confiance pour β_0 au niveau 95%.
6. Donner la valeur du R^2 ajusté.
7. Donner une estimation ponctuelle de σ^2 .
8. Donner la valeur du f_{obs} du test global de Fisher.
9. Retrouver les résultats numériques précédents avec des commandes R adéquates, le jeu de données étant disponible ici : <https://chesneau.users.lmno.cnrs.fr/scores.txt>

Exercice 2. On souhaite expliquer le chemin de freinage en mètres d'un véhicule (distance parcourue entre le début du freinage et l'arrêt total) (variable Y) en fonction de sa vitesse en kilomètres heure (variable X). Pour $n = 12$ expériences indépendantes, on observe les valeurs $(x_1, y_1), \dots, (x_n, y_n)$ de (X, Y) suivantes :

x_i	40	50	60	70	80	90	100	110	120	130	140	150
y_i	9	11	20	27	39	45	58	78	79	93	108	124

1. Représenter le nuage de points $\{(x_1, y_1), \dots, (x_n, y_n)\}$. À partir de celui-ci, expliquer pourquoi on peut envisager l'existence d'une liaison linéaire entre Y et X .

On adopte le modèle de *rls* : $Y = \beta_0 + \beta_1 X + \epsilon$, avec $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Les paramètres β_0 , β_1 et σ sont des réels inconnus.

2. Reproduire et comprendre l'enjeu des commandes suivantes :

```
X = c(40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150)
Y = c(9, 11, 20, 27, 39, 45, 58, 78, 79, 93, 108, 124)
plot(X, Y)
reg = lm(Y ~ X)
summary(reg)
```

3. Donner le R^2 et le R^2 ajusté. Que peut-on en dire ?
4. Donner les *emco* ponctuels de β_0 et β_1 .
5. Tracer la droite de régression sur le nuage de points.
6. Donner la valeur prédite de Y lorsque $X = 127$.
7. Donner une estimation ponctuelle de σ .
8. Quelle matrice renvoie les commandes `vcov(reg)` ? À quoi correspond la sortie des commandes `sqrt(diag(vcov(reg)))` ?
9. Représenter le graphique des résidus. Est-ce que les hypothèses standards semblent être satisfaites ?
10. Est-ce que la régression est significative ?
11. Donner un intervalle de confiance pour la valeur moyenne de Y lorsque $X = 127$ au niveau 95%.
12. Reproduire et comprendre l'enjeu des commandes suivantes :

```
plot(X, Y)
s = seq(40, 150, length = 100)
ICpred = predict(reg, new = data.frame(X = s), interval = "confidence",
level = 0.95)
matlines(s, cbind(ICpred), lty = c(1, 4, 4), col = c(1, 2, 2),
lwd = c(1, 2, 2))
```

13. Reproduire et comprendre l'enjeu des commandes suivantes :

```
res = summary(reg)
tnu = qt(0.975, df = reg$df)
coef(res)[1, 1] + c(-1, 1) * tnu * coef(res)[1, 2]
coef(res)[2, 1] + c(-1, 1) * tnu * coef(res)[2, 2]
```

Proposer des commandes plus simples qui renvoient les mêmes résultats.

14. Tracer un ellipsoïde de confiance pour $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ au niveau 95%.