TP régression multiple

Master ingénierie mathématiques

Prévisions et intervalles de prévisions.

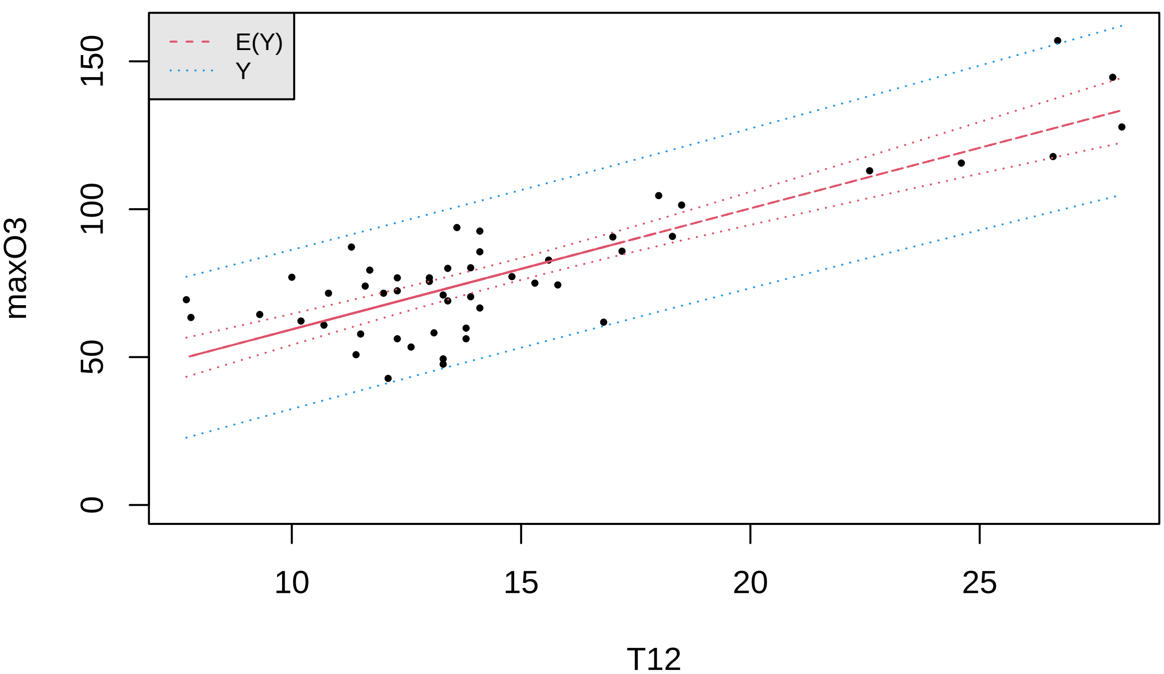
Dans cet exercice on veut modéliser l'ozone (`maxO3`) en fonction de la température (`T12`). On considère le modèle

$$\text{maxO3}_i = \beta_0 + \beta_1 \text{T12}_i + \varepsilon_i$$

où sont i.i.d gaussiens centrés de variance σ^2. L'estimation des paramètres du modèle se fera avec la méthode des MCO.

Note : On ne prendra que les 50 premières données. du fichier `ozone_complet.txt` disponible sur MADOC.

1. Trouver les estimations des paramètres de la régression en utilisant la fonction `lm`.
2. Ecrire une fonction permettant de calculer l'écart type des estimateurs $\widehat{\beta}_0$ et $\widehat{\beta}_1$.
3. Ecrire une fonction permettant de calculer l'estimateur $\widehat{\sigma}^2$.
4. Retrouver les résultats précédents dans les sorties `summary` de l'objet `lm`.
5. Ecrire une fonction `R` permettant d'obtenir un intervalle de confiance de la droite de régression (i.e. de $X\widehat{\beta}$). Tracer les données, la droite de régression et son intervalle de confiance au niveau 95%. Commenter.
6. Ecrire une fonction `R` permettant d'obtenir un intervalle de confiance de la valeur prédite $\widehat{Y}$. Ajouter cet intervalle de prédiction sur le graphe précédent. On obtiendra le graphe suivant :



7. En utilisant la fonction `predict` de `R` (on regardera l'aide) retrouver les résultats des questions 5 et 6.

Région de confiance versus intervalle de confiance

On ajoute au modèle précédent le vent Vx et la nébulosité (Ne12). Le modèle considéré est :

$$\mathrm{maxO3}_i = \beta_0 + \beta_1 \mathrm{T12}_i + \beta_2 \mathrm{Ne12}_i + \beta_3 \mathrm{Vx}_i + \varepsilon_i$$

où ε_i sont i.i.d gaussiens centrés de variance σ^2. L'estimation des paramètres du modèle se fera là-encore par MCO.

Note : On reste sur les 50 premières données.

1. Calculer les intervalles de confiance des paramètres du modèle de régression. On utilisera la fonction `confint`.
2. En utilisant la librairie `ellipse`, reproduire le graphe ci dessous. Interpréter le fait que certaines ellipses ont des axes quasiement parallèles aux axes du repère et d'autres non.

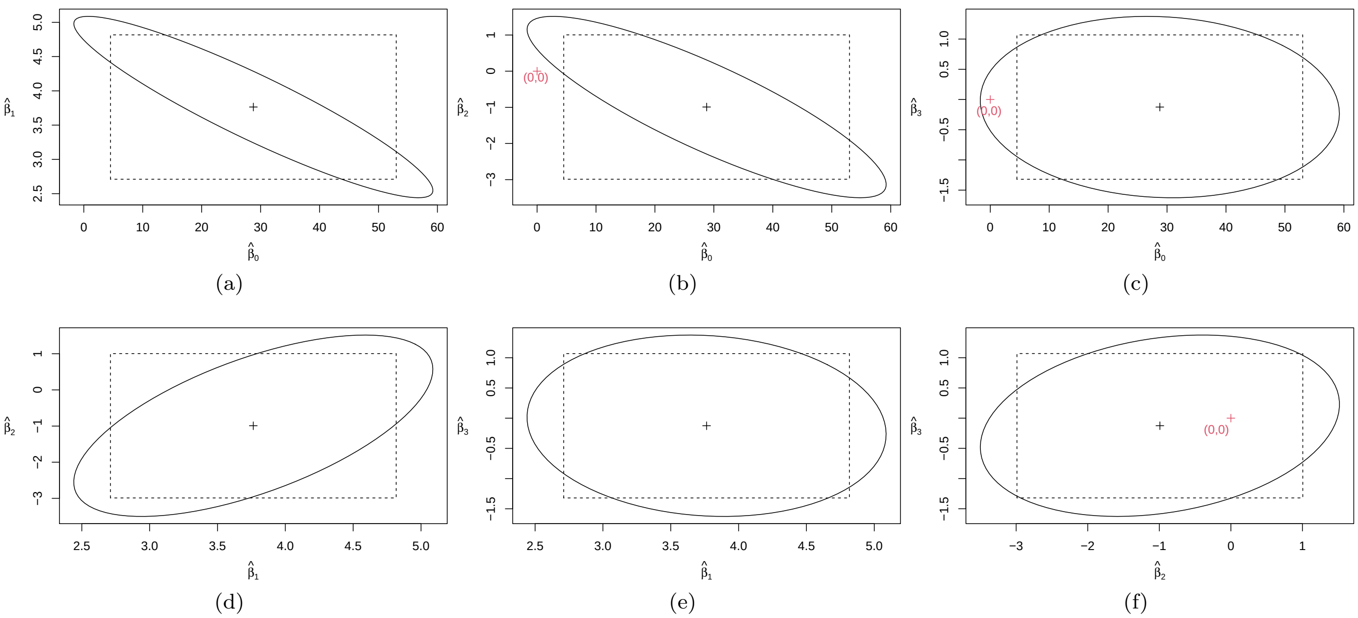


Figure 1: RC et IC des coefficients de régression

3. Peut-on conclure à la nullité de certains coefficients en utilisant le graphe précédent ? Les résultats obtenus avec la région de confiance et avec les intervalles de confiance sont-ils toujours en accord ?
4. Retrouver les résultats des tests pour $H_0 : \beta_i = 0$ en utilisant la fonction `summary`.
5. D'après le modèle, quel impact moyen aura une hausse de la température de 1°C ?

Corrélation partielle :

Soit un modèle de régression multiple

$$Y_i = \beta_0 + \beta_1 X_{1,i} + ... + \beta_p X_{p,i} + \varepsilon_i$$

où les ε_i sont centrés, non corrélés, de variance commune σ^2.

1. Expliquer en quoi la procédure suivante permet de calculer le coefficient β_1 par la méthode MCO :

Etape 1 : Régresser par MCO Y sur $X_2, ..., X_p$, les résidus de ce modèle sont notés $\varepsilon_{Y|X_2,...,X_p}$.

Etape 2 : Régresser par MCO X_1 sur $X_2, ..., X_p$, les résidus de ce modèle sont notés $\varepsilon_{X_1|X_2,...,X_p}$.

Etape 3 : Régresser par MCO $\varepsilon_{Y|X_2,...,X_p}$ sur $\varepsilon_{X_1|X_2,...,X_p}$. La pente de ce modèle de régression simple est égal à β_1.

2. Appliquer cette procédure au modèle de l'exercice précédent.
3. On appelle corrélation partielle entre X_1 et Y en contrôlant $X_2, ..., X_p$ le coefficient de corrélation $r_{Y,X_1|X_2,...,X_p}$ entre $\varepsilon_{Y|X_2,...,X_p}$ et $\varepsilon_{X_1|X_2,...,X_p}$. Calculer ce coefficient avec les données précédentes.
4. Comment tester simplement $H_0 : r_{Y,X_1|X_2,...,X_p} = 0$?