

Serie de travaux dirigés n°3 (RLM)

Exercice 1 : On dispose de n observations sur une variable dépendante y et on régresse cette variable sur une constante, de sorte que le modèle qu'on cherche à estimer s'écrit :

$$y_i = \beta_0 + \varepsilon_i \quad i = 1, \dots, n$$

- Écrire ce modèle sous forme matricielle .
- Montrer que le paramètre estimé $\hat{\beta}_0$ est égal à la moyenne empirique de la variable Y .
- Montrer que le coefficient de détermination associé à cette régression est nul.

Exercice 2 : Dans un service de maternité souhaite étudier l'influence d'âge de la mère (en années) et durée de la grossesse (en semaine) sur le poids du nouveau-né (en kilogrammes). On dispose d'un échantillon de 6 accouchements .

poids (Kg)	âge (année)	durée(semaine)
3,84	22	9
3,88	28	9
4,30	35	10
3,41	19	7
3,82	31	8
4,13	26	11

1. Présenter le modèle à régresser sous forme matricielle.
2. Estimer les coefficients β_0, β_1 et β_2 du modèle par calcul matriciel.
3. En déduire l'expression de $\hat{Y}_{(6,1)}$, vecteur des valeurs ajustées de y , et celle de $\hat{\varepsilon}_{(6,1)}$, vecteur des résidus estimés.
4. Vérifier que le point moyen appartient à l'hyperplan de régression.
5. Calculer le coefficient de détermination associé à cette estimation.

Exercice 3 : Montrer que le théorème de Gauss-Markov est vérifié dans le cas de modèle de régression linéaire multiple où il y a de p variables aléatoires $(X_1, \dots, X_p)$

. **Exercice 4 :** on suppose que on a n observations $y_1, \dots, y_n$ et que y_i suit une loi normale $y_i \sim \mathcal{N}(\mu, \sigma^2)$

- Écrire la log-vraisemblance $(\log L(\mu, \sigma^2))$.
- Calculer l'estimateur du maximum de vraisemblance de (μ) .
- Montrer que : $\hat{\sigma}_{MLE}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$.