1^{ere} Année Ingénieur

Modèle linéaire

Statistiques TD 9

1 Modèle linéaire

Soit le modèle

$$Y = X\theta + \epsilon, \theta \in R^k$$

X est une matrice $n \times k$ de rang plein et $\epsilon = (\epsilon_i)_{i=1\dots n}$ est une suite i.i.d de loi $\mathcal{N}(0, \sigma^2)$.

- Quel est l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ?
- Quel est l'estimateur du maximum de vraisemblance $\hat{\sigma}^2$ de σ^2?
- Montrer que $\hat{\theta}$ est sans biais mais que cela n'est pas le cas de $\hat{\sigma}^2$. Donner un estimateur sans biais s^2 de σ^2 multiple de $\hat{\sigma}^2$.
- quelles sont les distributions de :
  - $s^2 \frac{n-k}{\sigma^2}$
  - $\frac{\|X(\hat{\theta}-\theta)\|}{\sigma^2}$
  - $\frac{\frac{\|X(\hat{\theta}-\theta)\|}{k}}{s^2}$
- En Déduire une région de confiance pour θ.

2 Régression multiple

3 Régression polynômiale

Soient p et n deux entiers tels que $0 \leq p < n$. Pour tout $1 \leq j \leq n$ on pose $x_j = j - \frac{n+1}{2}$. On considère $\psi_0, \psi_1, \dots, \psi_p$ polynômes fixés tels que :

- degré$(\psi_i) = i, \ \forall 0 \leq i \leq p$
- $\sum_{j=1}^{n} \psi_i(x_j)\psi_k(x_j) = 0, \ \forall 0 \leq i \neq k \leq p$

On a en vue d'étudier le modèle de régression

$$Y_j = \sum_{i=0}^{p} \lambda_i \psi_i(x_j) + \varepsilon_j, \quad 1 \leq j \leq n$$

où $\varepsilon_1, \dots, \varepsilon_n$ sont indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$ et $\lambda_0, \dots, \lambda_p, \sigma^2$ sont des paramètres inconnus.

a) Estimer les paramètres de ce modèle.

b) Ecrire un test de niveau α de "$\lambda_p = 0$" contre "$\lambda_p \neq 0$".

Le revenu net par action de la compagnie Gillette pour les années 57 à 64 est le suivant :

Année (z_j) :	1957	1958	1959	1960	1961	1962	1963	1964
Revenu en \$ $(Y_j(\omega))$:	0,93	0,99	1,11	1,33	1,52	1,60	1,47	1,33

On pose $x_j = z_j - 1960,5$ puis $\psi_0(x) = 1, \psi_1(x) = 2x$ et $\psi_2(x) = x^2 - 21/4$.

c) Vérifier brièvement que le choix ci-dessus de $x_j, \psi_0, \psi_1, \psi_2$ rentre dans le cadre décrit plus haut.

d) En supposant que $Y_j = \sum_{i=0}^{2} \lambda_i \psi_i(x_j) + \varepsilon_j, \quad 1 \leq j \leq 8$ avec $\varepsilon_1, \ldots, \varepsilon_8$ indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$, estimer $\lambda_0, \lambda_1, \lambda_2$ et σ^2. Tester si $\lambda_2 = 0$.

e) Faire une prévision pour le revenu net par action en 1965.

N.B. Pour faciliter les calculs, on indique les valeurs numériques suivantes :
$\sum_j \psi_1^2(x_j) = \sum_j \psi_2^2(x_j) = 168$; $\sum_j Y_j^2(\omega) = 13,65$; $\sum_j \psi_0(x_j) Y_j(\omega) = 10,28$;
$\sum_j \psi_1(x_j) Y_j(\omega) = 6,86$; $\sum_j \psi_2(x_j) Y_j(\omega) = -4,1$.

4 Degré 3

Pour les données ci-dessous, ajuster le modèle :

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon.$$

Tester l'hypothèse $\beta_2 = \beta_3$.

X	-5	-3	-1	1	3	5
Y	13	4	3	4	10	22