Sorbonne Université Année 2023/2024

M2 Probabilités et finance, Probabilités et Modèles aléatoires Premier semestre

Correction de la feuille de TD 3 : Théorème de Cochran et modèle linéaire

Exercice : Le principe de la régression linéaire est de modéliser une variable y à partir de variables explicatives $\mathbf{x} = \left(x_1, \ldots, x_p\right)^T$, i.e de considérer

$$y = \beta_1 x_1 + \ldots + \beta_p x_p,$$

où $\beta = \left(x_1, \ldots, x_p\right)$ est inconnu. En pratique, on dispose d'un échantillon $\left(\mathbf{x}_1, y_1\right), \ldots, \left(\mathbf{x}_n, y_n\right)$, mais on obtient jamais réellement une droite (erreurs de mesures...). On va donc considérer le modèle linéaire

$$y = \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon$$

avec $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$. On parle alors de modèle linéaire gaussien. On suppose maintenant que les données suivent le modèle suivant :

$$Y_i = \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p} + \epsilon_i,$$

avec

- Y_i est une variable aléatoire et on observe les réalisations y_i.
- Les $\mathbf{x}_i = \left(x_{i,1}, \ldots, x_{i,p}\right)^T$ sont déterministes.
- Le paramètre $\beta = \left(\beta_1, \ldots, \beta_p\right)^T$ est inconnu et déterministe.
- Les ϵ_i sont i.i.d et $\epsilon_1 \sim \mathcal{N}\left(0, \sigma^2\right)$.

1. Ecrire le modèle de manière matricielle.

$$Y = X\beta + \epsilon,$$

avec $Y = \left(Y_1, \ldots, Y_n\right)^T$, $\epsilon = \left(\epsilon_1, \ldots, \epsilon_n\right)^T$ et

$$X = \begin{pmatrix} x_{1,1} & \ldots & x_{1,p} \\ \vdots & \vdots & \vdots \\ x_{n,1} & \vdots & x_{n,p} \end{pmatrix}$$

Pour tout i, on a bien

$$Y[i] = \left(x_{i,1}, \ldots, x_{i,p}\right)^T \beta + \epsilon_i = x_{i,1}\beta_1 + \ldots + x_{i,p}\beta_p + \epsilon_i = Y_i.$$

2. Donner la loi du vecteur Y. Quelle est la loi de Y_i ?
   On a $\epsilon \sim \mathcal{N}\left(0, \sigma^2 I_p\right)$ et donc
   $$Y \sim \mathcal{N}\left(X\beta, \sigma^2 I_p\right)$$
   et en particulier $Y_i \sim \mathcal{N}\left((X\beta)[i], \sigma^2\right)$.

   On considère à partir de maintenant que $\text{rang}(X) = p$, et on note $D = \text{Im}(X)$. On s'intéresse à l'estimateur des moindres carrés défini par
   $$\hat{\beta} = \arg\min_{h \in \mathbb{R}^p} \|Y - Xh\|^2$$
   Montrer que la matrice $X^T X$ est inversible et en déduire $\hat{\beta}$.
   La matrice $X^T X$ est clairement symétrique. De plus pour tout h,
   $$h^T X^T X h = \|Xh\|^2 \geq 0.$$
   De plus, comme X est de taille $n \times d$ et de rang p, elle est injective, donc
   $$Xh = 0 \Longrightarrow h = 0_{\mathbb{R}^p}.$$
   On note $G(h) = \|Y - Xh\|^2$.
   $$\nabla G(h) = -2X^T(Y - Xh) \qquad \text{et} \qquad \nabla^2 G(h) = X^T X.$$
   Comme $X^T X$ est positive, la fonction est fortement convexe et si le gradient admet un 0, c'est donc l'unique minimiseur. On résout donc
   $$\nabla G(h) = 0 \Leftrightarrow X^T Y - -X^T X h = 0 \Leftrightarrow h = \left(X^T X\right)^{-1} X^T Y$$
   On obtient donc $\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$ qui est l'unique minimiseur des moindres carrés.
3. Montrer que $P_D = X\left(X^T X\right)^{-1} X^T$ est le projecteur orthogonale sur D parallèlement à $D^\perp$.
   P_D est clairement symétrique et
   $$P_D^2 = X\left(X^T X\right)^{-1} X^T X \left(X^T X\right)^{-1} X^T = X\left(X^T X\right)^{-1}\left(X^T X\right)\left(X^T X\right)^{-1} X^T = P_D.$$
   De plus pour tout $h' \in D$, il existe $h \in \mathbb{R}^p$ tel que $h' = Xh$. On a donc
   $$P_D(h') = P_D(Xh) = X\left(X^T X\right)^{-1} X^T X h = X\left(X^T X\right)^{-1} (X^T X) h = Xh = h'$$

Pour tout $h' \in D^\perp \Leftrightarrow \forall h \in \mathbb{R}^p, (Xh)^T h' = 0$. De plus, pour tout $h \in \mathbb{R}^p$ et $h' \in D^\perp$, on a

$$P_D(h')^T h = \underbrace{h'}_{\in D^\perp} \underbrace{P_D^T h}_{=P_D(h)\in D} = 0$$

et donc $P_D(h') = 0$.

4. Que pouvez vous en déduire sur $X\hat{\beta}$?

   On a

$$X\hat{\beta} = X\left(X^T X\right)^{-1} X^T Y = P_D Y$$

   et $X\hat{\beta}$ est donc la projection orthogonale de Y sur D.

5. Donner la loi de $X\hat{\beta}$ et en déduire celle de $\hat{\beta}$.

   Comme $Y \sim \mathcal{N}\left(\beta, \sigma^2 I_p\right)$, on obtient

$$X\hat{\beta} \sim \mathcal{N}\left(P_D X\beta, \sigma^2 P_D I_p P_D^T\right)$$

   Comme $X\beta \in D$ et P_D est une projection orthogonale, on obtient

$$X\hat{\beta} \sim \mathcal{N}\left(X\beta, \sigma^2 P_D\right)$$

   En multipliant par X^T, on obtient

$$X^T X\hat{\beta} \sim \mathcal{N}\left(X^T X\beta, \sigma^2 X^T P_D X\right)$$

   et comme $X^T X$ est inversible, et en remarquant que $\left(X^T X\right)^{-1} X^T P_D X = \left(X^T X\right)^{-1}$,

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2\left(X^T X\right)^{-1}\right).$$

6. On suppose σ^2 connu. Soit $x_0 \in \mathbb{R}^p \backslash \{0\}$, donner un intervalle de confiance de niveau au moins $1-\alpha$ de $x_0^T\beta$.

   On a

$$x_0^T\hat{\beta} \sim \mathcal{N}\left(x_0^T\beta, \sigma^2 x_0^T\left(X^T X\right)^{-1} x_0\right).$$

   En remarquant que $x_0^T\left(X^T X\right)^{-1} x_0 \in \mathbb{R}_+^*$, on obtient (en centrant et réduisant)

$$\frac{x_0^T\hat{\beta} - x_0^T\beta}{\sigma\sqrt{x_0^T\left(X^T X\right)^{-1} x_0}} \sim \mathcal{N}(0,1).$$

On a donc

$$1-\alpha = \mathbb{P}\left[-q_{1-\alpha} \leq \frac{x_0^T\hat{\beta} - x_0^T\beta}{\sigma\sqrt{x_0^T\left(X^TX\right)^{-1}x_0}} \leq q_{1-\alpha}\right] = \mathbb{P}\left[x_0^T\beta \in x_0^T\hat{\beta} \pm q_{1-\alpha}\sigma\sqrt{x_0^T\left(X^TX\right)^{-1}x_0}\right].$$

7. On suppose maintenant que σ^2 est inconnu et on considère l'estimateur

$$\hat{\sigma}^2 = \frac{1}{n-p}\left\|Y - X\hat{\beta}\right\|^2$$

   (a) Expliquer ce choix d'estimateur.

   Comme σ^2 est la variance des ϵ_i (et en particulier le moment d'ordre 2), un estimateur naturel aurait été

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - X\beta)^2 = \frac{1}{n}\|Y - X\beta\|^2.$$

   Cependant, comme β est inconnu, on le remplace par son estimateur. Le remplaçement de n par $n-p$ permet (a priori) d'obtenir un estimateur sans biais de σ^2.

   (b) Exprimer $\hat{\sigma}^2$ à l'aide de projections.

   On a

$$\frac{1}{n-p}\hat{\sigma}^2 = \frac{1}{n-p}\left\|Y - X\hat{\beta}\right\|^2 = \frac{1}{n-p}\|Y - P_DY\|^2 = \frac{1}{n-p}\|P_{D^\perp}Y\|^2$$

   (c) Enoncer le théorème de Cochran dans ce cas. On a $\mathbb{R}^p = D \oplus D^\perp$. Ainsi, $X\hat{\beta} = P_D(Y)$ et $P_{D^\perp}(Y)$ sont indépendants et comme $X\beta \in D$,

$$\frac{1}{\sigma^2}\|P_{D^\perp}Y - P_{D^\perp}X\beta\|^2 = \frac{n-p}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{n-p}$$

   et en particulier, $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendants. De plus, on a

$$\frac{1}{\sigma^2}\left\|X\beta - X\hat{\beta}\right\|^2 \sim \chi^2_p$$

   et est également indépendants de $\hat{\sigma}^2$.

   (d) En déduire un intervalle de confiance pour $x_0^T\beta$.

   Grâce à la question précédente, on a

$$\frac{x_0^T\hat{\beta} - x_0^T\beta}{\hat{\sigma}\sqrt{x_0^T\left(X^TX\right)^{-1}x_0}} \sim T_{n-p}.$$

On a donc

$$1-\alpha = \mathbb{P}\left[-t_{n-p,1-\alpha} \leq \frac{x_0^T\hat{\beta} - x_0^T\beta}{\hat{\sigma}\sqrt{x_0^T\left(X^TX\right)^{-1}x_0}} \leq t_{n-p,1-\alpha}\right] = \mathbb{P}\left[x_0^T\beta \in x_0^T\hat{\beta} \pm t_{n-p,1-\alpha}\hat{\sigma}\sqrt{x_0^T\left(X^TX\right)^{-1}x_0}\right]$$

(e) En déduire un test de niveau α pour tout tester $\beta_j = \beta_{j,0}$.

En notant $e_j = (0,\ldots,1,0,\ldots,0)^T$, on a $\beta_j = e_j^T\beta$. On a donc l'intervalle de confiance

$$IC_{1-\alpha}\left(\beta_j\right) = \left[x_0^T\hat{\beta} \pm t_{n-p,1-\alpha}\hat{\sigma}\left(X^TX\right)^{-1}[j,j]\right]$$

où $\left(X^TX\right)^{-1}[j,j]$ est la j,j-ème coordonnée de $\left(X^TX\right)^{-1}$. On rejette donc le test si $\beta_{j,0}$ n'est pas dans l'intervalle de confiance.

(f) Construire un intervalle de confiance pour σ.

Comme

$$\frac{n-p}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{n-p}$$

en notant $k_{\alpha/2}$ et $k_{1-\alpha/2}$ les quantiles d'ordre $\alpha/2$ et $1-\alpha/2$, on obtient l'intervalle

$$IC_{1-\alpha}\left(\sigma^2\right) = \left[\frac{(n-p)\hat{\sigma}^2}{k_{1-\alpha/2}}; \frac{(n-p)\hat{\sigma}^2}{k_{\alpha/2}}\right]$$

(g) Construire un test di niveau α pour tester $\beta = \beta_0$.

D'après la question c), on a

$$\frac{1}{\sigma^2}\left\|X\beta - X\hat{\beta}\right\|^2 \sim \chi^2_p \qquad \text{et} \qquad \frac{n-p}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{n-p}$$

et ces deux variables sont indépendantes. On a donc

$$\frac{\left\|X\beta - X\hat{\beta}\right\|^2}{\hat{\sigma}^2} \sim \mathcal{F}\left(p, n-p\right).$$

On cherche donc c_α tel que

$$\mathbb{P}\left[\frac{\left\|X\beta - X\hat{\beta}\right\|^2}{\hat{\sigma}^2} \geq c_\alpha\right] = \alpha \Leftrightarrow \mathbb{P}\left[\mathcal{F}\left(p,n-p\right) \geq c_\alpha\right] \Leftrightarrow c_\alpha = f_{p,n-p,1-\alpha}$$

où $f_{p,n-p,1-\alpha}$ est le quantile d'ordre $1-\alpha$ d'une loi de Fisher de paramètre $p, n-p$.

8. On considère maintenant une $n+1$-ème donnée x_{n+1} et on souhaite prédire Y_{n+1}.

   (a) Proposer un prédicteur $\hat{Y}_{n+1}$. On considère $\hat{Y}_{n+1} = x_{n+1}^T\hat{\beta}$.

   (b) Donner sa loi.

On a

$$x_{n+1}^T \hat{\beta} \sim \mathcal{N}\left(x_{n+1}^T \beta, \sigma^2 x_{n+1}^T \left(X^T X\right)^{-1} x_{n+1}\right).$$

(c) Donner la loi de l'erreur de prédiction $\hat{\epsilon}_{n+1} = Y_{n+1} - \hat{Y}_{n+1}$.

On a

$$Y_{n+1} - \hat{Y}_{n+1} = x_{n+1}^T \beta - x_{n+1} \hat{\beta} + \epsilon_{n+1} \sim \mathcal{N}\left(0, \sigma^2 x_{n+1}^T \left(X^T X\right)^{-1} x_{n+1} + \sigma^2\right)$$

(d) En déduire un intervalle de prédiction.

D'après la question précédente, et comme $\hat{\sigma}^2$ et $\hat{\beta}$ sont indépendants, on a

$$\frac{Y_{n+1} - \hat{Y}_{n+1}}{\hat{\sigma}\sqrt{x_{n+1}^T \left(X^T X\right)^{-1} x_{n+1} + 1}} \sim T_{n-p}.$$

On obtient donc l'intervalle de prédiction

$$IP_{1-\alpha}\left(Y_{n+1}\right) = \left[\hat{Y}_{n+1} \pm t_{n-p,1-\alpha/2} \hat{\sigma} \sqrt{x_{n+1}^T \left(X^T X\right)^{-1} x_{n+1} + 1}\right]$$