

TD 3 : Sélection de modèle

Exercice 1. Fisher global

1. Rappeler la définition des lois du chi-deux, de Student, et de Fisher.
2. On se place dans le cadre d'une régression linéaire gaussienne multiple $Y = X\beta + \varepsilon$. Rappeler les expressions et les lois de $\hat{\beta}$ et $\hat{\sigma}^2$.
3. En déduire que la statistique de test de Fisher global

$$F = \frac{1}{p\hat{\sigma}^2} \hat{\beta}^T (X^T X) \hat{\beta}$$

suit sous H_0 une loi de Fisher de paramètre $p, n - p$.

4. Réécrire F en fonction de Y et \hat{Y} .
5. Que teste cette statistique ? Que peut-on en dire en pratique ?

Exercice 2. Fisher emboîté

1. Rappeler le test entre modèles emboîtés et donner la statistique de test F en fonction de Y, \hat{Y} , et \hat{Y}_0 . Dans quel contexte retrouve-t-on l'expression de l'exercice 1 ?
2. Réécrire cette quantité en fonction de SCR et SCR_0 .
3. Montrer que

$$F = \frac{n-p}{q} \frac{R^2 - R_0^2}{1-R^2},$$

où R^2 et R_0^2 sont les coefficients de détermination associés respectivement au modèle complet et au modèle emboîté.

Exercice 3. Tests de Student et de Fisher

On considère le modèle de régression linéaire classique à n variables et p prédicteurs:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

On souhaite montrer l'équivalence entre les tests de Student et de Fisher pour tester la nullité du dernier coefficient:

$$\mathcal{H}_0 : \beta_p = 0 \quad \text{contre} \quad \mathcal{H}_1 : \beta_p \neq 0.$$

1. Soient $U \sim \mathcal{N}(0, 1)$ et $V \sim \chi_k^2$ deux variables aléatoires indépendantes (avec k un entier strictement positif). Quelle est la loi de $T = \frac{U}{\sqrt{V/k}}$? Quelle est la loi de $F = T^2$?
2. Rappelez les hypothèses classiques du modèle linéaire gaussien. Quelles sont les dimensions de \mathbf{y} , \mathbf{X} , $\boldsymbol{\beta}$ et $\boldsymbol{\epsilon}$? On se place sous ces hypothèses dans toute la suite.
3. Rappelez la statistique T_p du test de Student pour la nullité du coefficient β_p , et sa loi sous l'hypothèse \mathcal{H}_0 .
4. On décompose \mathbf{X} en blocs:

$$\mathbf{X} = [\mathbf{X}_0 \ \mathbf{X}_p] \quad \text{avec} \quad \mathbf{X}_0 = [\mathbf{X}_1 \ \cdots \ \mathbf{X}_{p-1}],$$

où \mathbf{X}_0 est la matrice de taille $n \times (p-1)$ des $(p-1)$ premières colonnes de \mathbf{X} .

Écrivez les deux modèles emboîtés qui correspondent au test de la nullité du coefficient β_p . Donnez la statistique F_p du test de Fisher correspondant, et sa loi sous \mathcal{H}_0 .

5. En utilisant la décomposition $\mathbf{X} = [\mathbf{X}_0 \ \mathbf{X}_p]$, donnez la matrice $\mathbf{X}^T \mathbf{X}$ sous forme de 4 blocs.

6. On admet le lemme d'inversion matricielle par blocs suivant:

Soit \mathbf{M} une matrice par blocs, $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$, avec $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ de dimensions respectives $q \times q$, $q \times r$, $r \times q$, et $r \times r$. On suppose \mathbf{M} et \mathbf{A} inversibles. Alors, on peut écrire \mathbf{M}^{-1} sous la forme:

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}, \quad \text{avec } \mathbf{H}^{-1} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}.$$

Montrez la relation suivante :

$$\frac{1}{[(\mathbf{X}^T \mathbf{X})^{-1}]_{pp}} = \mathbf{X}_p^T \mathbf{X}_p - \mathbf{X}_p^T \mathbf{X}_0 (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{X}_p$$

7. On note \mathbf{P}_0 la matrice de projection orthogonale sur l'espace \mathcal{M}_0 engendré par les $p-1$ colonnes de \mathbf{X}_0 , et \mathbf{P} la matrice de projection orthogonale sur \mathcal{M} engendré par les p colonnes de \mathbf{X} .

Donnez les expressions de \mathbf{P}_0 et \mathbf{P} en fonction de, respectivement, \mathbf{X}_0 et \mathbf{X} , puis montrez la relation suivante :

$$\frac{1}{[(\mathbf{X}^T \mathbf{X})^{-1}]_{pp}} = \mathbf{X}_p^T (\mathbf{I}_n - \mathbf{P}_0) \mathbf{X}_p.$$

8. On décompose $\hat{\beta}$ en deux blocs: $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_p \end{pmatrix}$. Montrez : $\mathbf{X}\hat{\beta} = \mathbf{X}_0\hat{\beta}_0 + \mathbf{X}_p\hat{\beta}_p$.

9. On note $\hat{\mathbf{y}}$ et $\hat{\mathbf{y}}_0$ les projetés orthogonaux de \mathbf{y} sur \mathcal{M} et \mathcal{M}_0 . Justifiez l'égalité:

$$\hat{\mathbf{y}}_0 = \mathbf{P}_0 \hat{\mathbf{y}}.$$

En déduire:

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_0 = (\mathbf{I}_n - \mathbf{P}_0) \mathbf{X}_p \hat{\beta}_p$$

10. Montrez que $T_p^2 = F_p$. En déduire l'équivalence des deux tests.

Exercice 4. Estimation sous contrainte Dans le modèle de régression linéaire, il arrive parfois que l'on souhaite imposer des contraintes linéaires à β , par exemple que sa première coordonnée soit égale à 1. Nous supposerons en général que nous imposons q contraintes linéairement indépendantes à β , ce qui s'écrit sous la forme : $R\beta = r$, où R est une matrice $q \times p$ de rang $q < p$ et r un vecteur de taille q . Montrer que l'estimateur des moindres carrés sous contraintes s'écrit:

$$\hat{\beta}_c = \hat{\beta} + (X^T X)^{-1} R^T [R(X^T X)^{-1} R^T]^{-1} (r - R\hat{\beta}).$$

Montrez qu'il est sans biais, et que sa variance est égale à :

$$\mathbf{V}[\hat{\beta}_c] = \mathbf{V}[\hat{\beta}] - \sigma^2 (X^T X)^{-1} R^T (R(X^T X)^{-1} R^T)^{-1} R (X^T X)^{-1}.$$