Modèles Linéaires

Correction Séance 5
Licence 3 MIASHS (2022-2023)

Guillaume Metzler
Institut de Communication (ICOM)
Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC UR 3083, Lyon, France
guillaume.metzler@univ-lyon2.fr

Dans cette séance, nous allons traiter les exercices 1 et 2 de la fiche de TD 3. Les deux exercices sont liés, nous les traiterons donc simultanément.

Ce TD vise à vous faire pratiquer l'Analyse de Variance en exploitant le formalisme vu dans le cadre de ce cours. On va donc se détacher de ce que nous avons pu voir dans le cadre du cours de Statistiques Inférentielles.

Les détails théoriques exploités pour ces deux exercices sont présentés dans l'énoncé de la fiche de TD, directement disponible sur le site de Stéphane Chrétien.

Exposition à un produit toxique dans une usine

On s'intéresse à un problème de santé en milieu industriel et plus précisément à l'exposition à un produit toxique dans une usine fabriquant des produits chimiques. Cette usine dispose de 4 ateliers et le plan proposé pour étudier l'exposition à la toxicité des produits est le suivant : on analyse un certain indicateur sur 20 personnes prises au hasard dans chaque atelier. Le résultat des analyses révèle un résultat entier noté x qui indique le degré d'exposition auquel la personne a été soumise au cours de son travail. Les résultats sont résumés dans le table ci-dessous

	Atelier 1	Atelier 2	Atelier 3	Atelier 4
$y = 7$	5	0	1	7
$y = 8$	4	4	8	5
$y = 9$	3	4	3	8
$y = 10$	4	5	5	0
$y = 11$	4	7	3	0

La première question qui vient à l'esprit est bien sûr de se demander si les ateliers sont équivalents du point de vue du risque d'exposition de ses employés.

Pour répondre à cette question, on sera amené à considérer la statistiques de test suivante

$$r = \frac{\dfrac{\|Y - p_V(Y)\|_2^2}{n-4}}{\dfrac{\|p_V(Y) - p_U(Y)\|_2^2}{4-1}},$$

où

- V est l'image de la matrice X et la projection de Y sur l'image de X est donnée par
$$p_V(Y) = X(X^T X)^{-1} X^T Y.$$
- U est l'image du vecteur X, $\mathbf{e}$ où $\mathbf{e}$ est le vecteur dont les composantes sont des "1" et la projection est donnée par
$$p_V(Y) = \mathbf{e}(\mathbf{e}^T\mathbf{e})^{-1}\mathbf{e}^T Y.$$

Notre objectif est maintenant de calculer cette statistique de test, on va donc commencer par définir les différents objets dont nous avons besoin :

```
# Création des objets nécessaires à l'étude

# La matrice de design
X = matrix(0, nrow=80, ncol=4)
X[1:20,1] = rep(1,20)
X[21:40,2] = rep(1,20)
X[41:60,3] = rep(1,20)
X[61:80,4] = rep(1,20)


# Le vecteur des réponses
a1 = c(rep(7,5),rep(8,4),rep(9,3),rep(10,4),rep(11,4))
a2 = c(rep(7,0),rep(8,4),rep(9,4),rep(10,5),rep(11,7))
a3 = c(rep(7,1),rep(8,8),rep(9,3),rep(10,5),rep(11,3))
a4 = c(rep(7,7),rep(8,5),rep(9,8),rep(10,0),rep(11,0))
y = c(a1,a2,a3,a4)
```

```r
# Notre vecteur unité
e = rep(1,80)
```

On doit ensuite calculer les différentes projections sur les espaces U et V, ce qui nous donne

```r
# Projection sur l'espace V

Pv = X%*%solve(t(X)%*%X)%*%t(X)%*%y

# Projection sur l'espace U

Pu = e%*%solve(t(e)%*%e)%*%t(e)%*%y
```

Il nous reste à calculer le numérateur et le dénominateur qui servent à définir notre statistique de test r pour la déterminer complètement

```r
# Norme du numérateur
num = sum((y-Pv)^2)


# Norme du dénominateur
denom = sum((Pv-Pu)^2)

# Statistique de test
r = (num/76)/(denom/3)
r

## [1] 0.1531694
```

Pour conclure au rejet ou non de l'hypothèse H_0, on va regarder si la valeur de la statistique de test se trouve ou non dans l'intervalle de confiance de niveau $1-\alpha$, $I_{1-\alpha}$ de la loi de Fisher à 76 et 3 degrés de libertés, *i.e.*, regardons si

$$r \in \left[f_{\alpha/2}, f_{\alpha/2}\right],$$

où $\alpha = 0.05$.
Or ces quantiles sont données par

```r
# Borne inférieure de l'intervalle de confiance
borne_inf = qf(0.025,76,3)
# Borne supérieure de l'intervalle de confiance
borne_sup = qf(0.975,76,3)

# Il reste à faire le test
ifelse(((r<borne_inf)|(r>borne_sup)),"On rejette H0",
"On ne rejette pas H0")

## [1] "On rejette H0"
```

Le test nous conduit donc à rejeter l'hypothèse H_0. On peut donc en conclure qu'au moins deux ateliers sont exposés à des degrés différents.