

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER BISKRA

Faculté des Sciences Exactes et Sciences de la Nature et de la Vie  
Département de Mathématiques



Première Année Master

Notes de Cours

---

# Analyse de Données

Chapitre 3 : Analyse de la Variance  
(Séance 7)

---

Auteur des notes :

Pr. Djabrane YAHIA  
Dr. Sana BENAMEUR

Année universitaire : 2020-2021

### 2.5.3 Inférence Sur le Modèle

On test l'hypothèse globale

$$\begin{cases} H_0 : \beta = \beta_0, \\ H_1 : \beta \neq \beta_0. \end{cases} \quad \beta_0 \in \mathbb{R}^{p+1}.$$

Le modèle dans ce cas sera tester globalement, on sait que

$$\frac{(\hat{\beta} - \beta)^t (X^t X) (\hat{\beta} - \beta)}{(p+1) \sigma_\varepsilon^2} \sim \frac{\chi_{p+1}^2}{p+1},$$

et

$$\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \frac{\chi_{n-p-1}^2}{(n-p-1)},$$

de plus  $\hat{\beta}$  et  $\hat{\sigma}_\varepsilon^2$  sont indépendante, ce qui implique que

$$Q(\hat{\beta}, \hat{\sigma}_\varepsilon^2) = \frac{(\hat{\beta} - \beta)^t (X^t X) (\hat{\beta} - \beta)}{(p+1) \hat{\sigma}_\varepsilon^2} \sim \mathcal{F}(p+1, n-p-1)$$

$\mathcal{F}(p+1, n-p-1)$  est la loi de Fisher avec  $(p+1)$  et  $(n-p-1)$  ddl. Au niveau  $\alpha$ , on acceptera  $H_0$  si

$$Q(\hat{\beta}, \hat{\sigma}_\varepsilon^2) \leq f$$

où  $f$  est le fractile d'ordre  $(1-\alpha)$  de la loi de Fisher  $\mathcal{F}(p+1, n-p-1)$ .

Les résultats sont habituellement présentés dans un tableau appelé le tableau d'analyse de la variance sous la forme suivante :

Source de Variation	ddl	Somme des carrés	Moyenne des carrés	$F$
Expliquée	$p$	$SCE$	$MCE = \frac{SCE}{p}$	$\frac{MCE}{MCR}$
Résiduelle	$n-p-1$	$SCR$	$MCR = \frac{SCR}{n-p-1}$	
Totale	$n-1$	$SCT$		

On accepte  $H_0$  si la statistique :

$$F = \frac{\frac{SCE}{p}}{\frac{SCR}{n-p-1}} = \frac{MCE}{MCR} \leq f.$$

où  $f$  est le fractile d'ordre  $(1-\alpha)$  de la loi de Fisher  $\mathcal{F}(p, n-p-1)$ .

### 2.5.4 Inférence Sur le Modèle Réduit

On pratique, il est possible de procéder au test de la nullité de certains termes seulement, le problème sera donc à tester :

$$\begin{cases} H_0 : b_1 = b_2 = \dots = b_q = 0, \\ H_1 : b_j \neq 0, j = \overline{1, q}, q < p. \end{cases}$$

$p$  est le nombre exact des paramètres du modèle. Soit  $R_q^2$  le coefficient de détermination du modèle réduit à  $(p - q)$  variables.

Sous  $H_0$ , la statistique

$$Q_q = \frac{(R^2 - R_q^2)(n - p - 1)}{(1 - R^2)q} = \frac{(SCE - SCE_q)/q}{SCR/(n - p - 1)},$$

où  $SCE_q$  est la somme des carrées expliquée du modèle réduit.

Cette statistique suit une loi de Fisher à  $q$  et  $(n - p - 1)$  ddl, et on accepte  $H_0$  si

$$Q_q \leq f',$$

où  $f'$  est le fractile d'ordre  $(1 - \alpha)$  de  $\mathcal{F}(q, n - p - 1)$

## 2.6 Prévision

La prévision dans le cas d'un modèle linéaire multiple consiste à calculer une estimation

$$\hat{Y}_0 = \hat{b}_0 + \hat{b}_1 X_0^1 + \hat{b}_2 X_0^2 + \dots + \hat{b}_p X_0^p$$

avec  $X_0 = (1, X_0^1, X_0^2, \dots, X_0^p)$  sont les observations qui arrivent après avoir écrit le modèle.

Les intervalles de confiance des prévisions de  $Y$  et  $\mathbb{E}(Y)$  au niveau de confiance  $(1 - \alpha)\%$ , sont données respectivement par :

$$\begin{aligned} \text{prév}(Y) &= \hat{Y}_0 \mp t\hat{\sigma}_\varepsilon \left(1 + X_0 (X^t X)^{-1} X_0^t\right)^{1/2} \\ \text{prév}(\mathbb{E}(Y)) &= \hat{Y}_0 \mp t\hat{\sigma}_\varepsilon \left(X_0 (X^t X)^{-1} X_0^t\right)^{1/2} \end{aligned}$$

où  $t$  est le fractile d'ordre  $(1 - \alpha/2)$  de la loi de Student  $\mathcal{T}(n - p - 1)$ .