TD DE MODÈLES LINÉAIRES - FEUILLE 9
Analyse de variance - partie 2

EXERCICE 1 Dans le cadre de l'analyse de variance à 2 facteurs, le modèle peut être réécrit sous la forme suivante

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

pour $1 \leq i \leq I$, $1 \leq j \leq J$, et $1 \leq k \leq n_{ij}$.

1. Quel est le nombre de paramètres à identifier ?
2. Le modèle est-il identifiable ?
3. Commenter les conditions d'identifiabilité, dites d'orthogonalité.

EXERCICE 2 Considérons le modèle

$$Y_{ij} = \mu + \alpha_i + \beta_j + cP_{ij} + \varepsilon_{ij},$$

pour $1 \leq i \leq I$, $1 \leq j \leq J$, où les P_{ij} sont connus, les ε_{ij} sont indépendants de loi $\mathcal{N}(0, \sigma^2)$ et $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0$.

1. Montrer que $Y = \tilde{X}\theta + \varepsilon$, où $A = [\mathbb{1}|a_1|\cdots|a_I|b_1|\cdots|b_J|P]$ avec $\mathbb{1}, a_1, ..., a_I, b_1, ..., b_J, P$ éléments de $\mathbb{R}^{IJ}$ que l'on précisera.
2. Soit $F_0 = \{\mu 1, \mu \in \mathbb{R}\}, F_1 = \Big\{ \sum_{i=1}^{I} \alpha_i a_i, \alpha_i \in \mathbb{R}, \sum_{i=1}^{I} \alpha_i = 0 \Big\}$ et $F_2 = \Big\{ \sum_{j=1}^{J} \beta_j b_j,\ \beta_j \in \mathbb{R}, \sum_{j=1}^{J} \beta_j = 0 \Big\}$.
   Montrer que F_0, F_1 et F_2 sont des sous-espaces vectoriels deux-à-deux orthogonaux. En déduire que
   $$\mathbb{R}^{IJ} = F_0 \oplus F_1 \oplus F_2 \oplus G.$$
3. Calculer $P_0, P_0 + P_1$. (P_0 étant la projection orthogonale sur F_0). En déduire P_1. Calculer P_G.
4. En déduire que l'on a la décomposition
$$\begin{aligned}\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \mu - \alpha_i - \beta_j - cP_{ij})^2 &= IJ(\bar{Y}_{..} - \mu - c\bar{P}_{..})^2 + J\sum_{i=1}^{I}(\bar{Y}_{i.} - \bar{Y}_{..} - \alpha_i - c(\bar{P}_{i.} - \bar{P}_{..}))^2 \\ &+ I\sum_{j=1}^{J}(\bar{Y}_{.j} - \bar{Y}_{..} - \beta_j - c(\bar{P}_{.j} - \bar{P}_{..}))^2 \\ &+ \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..} - c(P_{ij} - \bar{P}_{i.} - \bar{P}_{.j} + \bar{P}_{..}))^2.\end{aligned}$$
5. En déduire que les ESBVM de $\mu, \alpha_1, ..., \alpha_I, \beta_1, ..., \beta_J$ vérifient $\hat{\mu} = \bar{Y}_{..} - \hat{c}\bar{P}_{..}$, $\hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..} - \hat{c}(\bar{P}_{i.} - \bar{P}_{..})$ et $\hat{\beta}_j = \bar{Y}_{.j} - \bar{Y}_{..} - \hat{c}(\bar{P}_{.j} - \bar{P}_{..})$.
6. Déterminer $\hat{c}$ et $\hat{\sigma}^2$.
7. On suppose que $c = 0$, construire la table d'analyse de la variance dans ce cas.
8. On suppose que $c = 0$, déterminer un intervalle de confiance de β_j au seuil $1 - \alpha$ de type Student pour j tel que $1 \leq j \leq J$.
9. On suppose que $c = 0$, tester au seuil $1 - \alpha$ l'hypothèse que $\beta_j = 0$ pour $1 \leq j \leq J$.
10. Proposer un test pour l'hypothèse $c = 0$.
11. On suppose avoir rejeté l'hypothèse $c = 0$, tester alors l'hypothèse H_0 : "tous les β_j sont nuls" contre H_1 : "tous les β_j ne sont pas nuls".
12. Donner la table d'analyse de la variance dans le cas $c \neq 0$.