Plans d'Expériences

Josephson Junior R.

April 25, 2024

Table des matières

1. ANOVA à un facteur de contrôle
   - Formulation du modèle
   - Analyse du modèle à effets fixes
   - Estimation du modèle
   - Méthode des Contrastes
   - Comparaison des contrastes
   - Comparaison de paires traitements-moyens

Considerons **a** traitements se présentant comme la modalité d'un **seul facteur**. Pour comparer entre ses a traitements on dispose de **n** unités expérimentales. La variable réponse de chacun de ses **a** traitements est une variable aléatoire $\mathbf{Y_{ij}}$ représentant le $\mathbf{j^{ème}}$ obs effectuée sur le traitement i.

Modèle 1

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

- μ_i : moyenne du $\mathbf{i^{ème}}$ niveau
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

Modèle 2

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

- α_i : facteur d'influence de la modalité i appelé **effet principal du** $\mathbf{i^{ème}}$ niveau.

- Total des observations du **traitement i** :

$$Y_i = \sum_{j=1}^{n} Y_{ij}$$

- Moyenne des observation du **traitement i** :

$$\bar{Y}_i = \frac{1}{n} \sum_{j=1}^{n} Y_{ij}$$

- Total de toutes les observations :

$$Y = \sum_{i=1}^{a} \sum_{j=1}^{n} Y_{ij}$$

- Moyenne de toutes les observations :

$$\bar{Y} = \frac{Y}{N} \quad \textbf{où} \quad \mathbf{N} = \mathbf{n} \times \mathbf{a}$$

L'objectif du travail est d'effectuer un test d'égalité des moyennes des **a traitements** :

$$\begin{cases} H0: & \mu_i = \ldots = \mu_a \\ H1: & \mu_i \neq \mu_j \end{cases}$$

Au niveau du modèle à effets fixes on définit la moyenne du traitement i :

$$\mu_i = \mu + \alpha_i \Longrightarrow \mu = \frac{\sum_{i=1}^{a} \mu_i}{a} \Rightarrow \sum_{i=1}^{a} \alpha_i = 0$$

Alternativement on pourrait tester **l'hypothèse de nullité des effets traitements** :

$$\begin{cases} H0: & \alpha_i = \ldots = \alpha_a = 0 \\ H1: & \alpha_i \neq 0 \end{cases}$$

Décomposition de la Variance

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}\right)^2 = n\sum_{i=1}^{a}\left(\bar{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_i\right)^2$$

$$\mathbf{SS_T = SS_{Traitement} + SS_E = Between + Within}$$

Au sens de la distribution χ^2 de Pearson **les degrés de liberté sont additifs** :

$$\mathbf{(N-1) = (a-1) + (N-a)}$$

Carrés Moyens

$$\mathbf{CM_{(SST)}} = \frac{\mathbf{SS_T}}{\mathbf{N-1}}$$

$$\mathbf{CM_{SSTraitement}} = \frac{\mathbf{SS_{Traitement}}}{\mathbf{a-1}}$$

$$\mathbf{CM_{SSE}} = \frac{\mathbf{SS_E}}{\mathbf{N-a}}$$

Approche par Espérance

$$\mathbf{E(CM_{SSE})} = \sigma^2$$

$$\mathbf{E(CM_{SSTraitement})} = \sigma^2 + \frac{\mathbf{n}\sum \alpha_i^2}{\mathbf{a-1}}$$

Test d'hypothèse : Equivalence des traitements

Sous H0 vraie on définie la statistique suivante :

$$\mathbf{F_0} = \frac{\mathbf{CM_{SSTraitement}}}{\mathbf{CM_{SSE}}} \quad \sim \quad \mathcal{F}(\mathbf{a-1}, \mathbf{N-a})$$

Decision : Si $\mathbf{F_0} > \mathbf{F^c}$ alors on rejette H0 (il y a bien des différences dans les traitements moyens).

Loi **Standard** et Loi **non-centré**

- Loi non-centré :

$$Y \sim \mathcal{N}(\mu, I) \;\Rightarrow\; Y^2 \sim \chi^2(n, \lambda) \;\;;\;\; \lambda = \sum \mu_i^2$$

.

- Loi standard :

$$Y^2 \sim \chi^2(n) \;\;;\;\; \lambda = 0$$

.

Rejet de H0 : Conséquence

Sous H1 vraie, qu'est ce qu'on aurait ? (Se souvenir de **E(CMSStraitement)**)

$$\frac{\mathbf{SS_{Traitement}}}{\sigma^2} \sim \chi^2(\mathbf{a}-\mathbf{1},\lambda) \ ; \ \lambda = \frac{\mathbf{n}\sum \alpha_{\mathbf{i}}^2}{\sigma^2}$$

En faite sous H1 vraie la loi de la statistique va changer :

$$\mathbf{F_0} = \frac{\mathbf{CM_{SSTraitement}}}{\mathbf{CM_{SSE}}} \sim \mathcal{F}(\mathbf{a}-\mathbf{1},\mathbf{N}-\mathbf{a},\lambda)$$

L'utilité de ce changement est dans le calcul de la puissance du test :

$$\eta = \mathbf{P}(\textbf{rejet H0}/\textbf{H1 vraie}) = \mathbf{1} - \beta = \mathbf{P}(\mathbf{F} > \mathbf{F_{a-1,N-a}})$$

Sachant que $\mathbf{F} \sim \mathcal{F}(\mathbf{a}-\mathbf{1},\mathbf{N}-\mathbf{a},\lambda)$.

Par MCO nous avons les résultats suivants pour le modèle 2 :

$$\hat{\mu} = \bar{\mathbf{Y}} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\mathbf{N}}\right) \quad ; \quad \hat{\alpha}_{\mathbf{i}} = \bar{\mathbf{Y}}_{\mathbf{i}} - \bar{\mathbf{Y}} \sim \mathcal{N}\left(\mu, \sigma^2\left(\frac{\mathbf{1}}{\mathbf{n}} - \frac{\mathbf{1}}{\mathbf{N}}\right)\right)$$

Pour estimer σ^2 on a :

$$\hat{\sigma}^2 = \frac{\mathbf{1}}{\mathbf{N}-\mathbf{a}} \sum_{\mathbf{i=1}}^{\mathbf{a}} \sum_{\mathbf{j=1}}^{\mathbf{n}} \left(\mathbf{Y}_{\mathbf{ij}} - \bar{\mathbf{Y}}_{\mathbf{i}}\right)^2 = \frac{\mathbf{SS_E}}{\mathbf{N}-\mathbf{a}}$$

Pour le ième traitement moyen on a l'IC suivant :

$$IC_{\mu_i} = \left[\bar{Y}_i \quad \pm \quad t_{\alpha/2,N-a} \times \frac{\hat{\sigma}}{\sqrt{n}}\right]$$

Pour la différence de deux traitements moyens :

$$IC_{\mu_i-\mu_{i'}} = \left[\bar{Y}_i - \bar{Y}_{i'} \quad \pm \quad t_{\alpha/2,N-a} \times \sqrt{\frac{2\hat{\sigma}^2}{n}}\right]$$

Nécessité

Quand le test d'ANOVA arrive à réjeter H0 il est préferable de savoir où se situe les différences. On a recours à la méthode des contrastes.
En général, une contraste est formulée comme suite :

$$\Gamma = \sum_{i=1}^{a} c_i \mu_i = \sum_{i=1}^{a} c_i \alpha_i \quad ; \quad \sum_{i=0}^{a} c_i = 0$$

En termes de contraste on teste les hypothèses suivantes :

$$\begin{cases} \textbf{H0} : & \Gamma = 0 \\ \textbf{Ha} : & \Gamma \neq 0 \end{cases}$$

Approche du **t-test**

On définit le contraste d'intérêt en terme de total traitements :

$$C = \sum_{i=1}^{a} c_i Y_i \quad ; \quad V(C) = n\sigma^2 \sum_{i=1}^{a} c_i^2 \quad ; \quad E(C) = n \sum_{i=1}^{a} c_i \mu_i$$

Sous H0 vraie on définit la statistique comme suite :

$$\mathbf{T} = \frac{\sum \mathbf{c_i Y_i}}{\sqrt{\mathbf{n} \times \sigma^2 \times \sum \mathbf{c_i^2}}} \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$$

Si σ^2 inconnue alors :

$$\mathbf{T} = \frac{\sum \mathbf{c_i Y_i}}{\sqrt{\mathbf{n} \times \hat{\sigma}^2 \times \sum \mathbf{c_i^2}}} \sim \mathcal{T}(\mathbf{N} - \mathbf{a})$$

Approche du **F-test**

On retrouve la loi de Fisher pour un Student au carré :

$$\mathbf{F} = \mathbf{T}^2 = \left(\frac{\sum \mathbf{c_i Y_i}}{\sqrt{\mathbf{n} \times \hat{\sigma}^2 \times \sum \mathbf{c_i^2}}} \right)^2 = \frac{\left(\frac{\sum \mathbf{c_i Y_i}}{\sqrt{\mathbf{n} \sum \mathbf{c_i^2}}} \right)^2}{\mathbf{CM_{SSE}}} = \frac{\mathbf{CM_{SSC}}}{\mathbf{CM_{SSE}}} \sim \mathbf{F}(1, \mathbf{N} - \mathbf{a})$$

Il suffit de remarquer :

$$\mathbf{SS_C} = \left(\frac{\sum \mathbf{c_i Y_i}}{\sqrt{\mathbf{n} \times \sum \mathbf{c_i^2}}} \right)^2 \sim \chi^2(1)$$

Intervalle de Confiance d'un contraste

Comme tester tester par l'hypothèse on va chercher l'IC de Γ :

$$\mathbf{IC_\Gamma = \left[\sum_{i=1}^{a} c_i \bar{Y}_i \ \pm \ t_{\alpha/2,N-a} \times \sqrt{\frac{\sigma^2}{n} \times \sum_{i=1}^{a} c_i^2} \right]}$$

Contrastes Standardisés

Lorsqu'il y a plus qu'un constraste d'intérêt il est utile de les évaluer tous sur la même échelle. On définit alors le contraste standardisé par :

$$c_i^* = \frac{c_i}{\sqrt{n \times \sum c_i^2}} \quad \Rightarrow \quad C = \sum_{i=1}^{a} c_i^* Y_i$$

Remarques

- Sous echantillon de **tailles inégales** :

$$\sum_{i=1}^{a} n_i c_i = 0$$

- Contrastes orthogonaux c_i et d_i :

$$\sum_{i=1}^{a} c_i d_i = 0$$

Méthode de **Scheffé**

Elle définit une comparaison de tous types possibles de contrastes entre traitements moyens. On suppose un ensemble de m contrastes dans les traitements moyens d'intérêts déterminés par :

$$\Gamma = \sum_{i=1}^{a} c_{iu}\mu_i \quad \text{avec} \quad u \in [1, m]$$

Le contraste correspondant dans les moyennes traitements est :

$$C_u = \sum_{i=1}^{a} c_{iu}\bar{Y}_i$$

L'erreur standard est définie par :

$$S_{C_u} = \sqrt{CM_{SSE}\sum_{i=1}^{n}\frac{c_i^2}{n_i}}$$

La valeur critique est :

$$\mathbf{S_{a,u}} = \mathbf{S_{C_u}}\sqrt{(\mathbf{a}-\mathbf{1})\mathbf{F}^{\alpha}_{\mathbf{a-1,N-a}}}$$

Si $|\mathbf{C_u}| > \mathbf{S_{a,u}}$: on rejette H0 ($\Gamma_u = 0$).

$$\mathbf{IC_{\Gamma_u}} = [\mathbf{C_u} \ \pm \ \mathbf{S_{a,u}}]$$

Test de Tukey (1953)

$$\begin{cases} \mathbf{H0} : & \mu_i = \mu_j \\ \mathbf{Ha} : & \mu_i \neq \mu_j \end{cases}$$

La procédure utilise une distribution statistique de rang **studentisé** :

$$\mathbf{q} = \frac{\mathbf{\bar{Y}_{max} - \bar{Y}_{min}}}{\sqrt{\frac{\mathbf{CM_{SSE}}}{\mathbf{n}}}} \Rightarrow \begin{cases} \bar{Y}_{max} : & max\ \bar{Y}_i\ de\ l'échantillon \\ \bar{Y}_{min} : & min\ \bar{Y}_i\ de\ l'échantillon \end{cases}$$

Pour rejeter H0 il faut que :

$$\left|\mathbf{\bar{Y}_i - \bar{Y}_j}\right| > \mathbf{q}_\alpha(\mathbf{a}, \mathbf{f}) \times \sqrt{\frac{\mathbf{CM_{SSE}}}{\mathbf{n}}} \ ; \ \mathbf{f} = \mathbf{ddl}(\mathbf{CM_{SSE}})$$

Pour des tailles d'échantillon deséquilibré :

$$\mathbf{IC}_{\mu_\mathbf{i}-\mu_\mathbf{j}} = \left[\mathbf{\bar{Y}_i - \bar{Y}_j} \pm \frac{\mathbf{q}_\alpha(\mathbf{a}, \mathbf{f})}{\sqrt{\mathbf{2}}} \times \sqrt{\mathbf{CM_{SSE}} \times \left(\frac{\mathbf{1}}{\mathbf{n_i}} + \frac{\mathbf{1}}{\mathbf{n_j}}\right)}\right]$$

LSD - Fisher

C'est l'approche des différences les moins significatives de Fisher. La paire moyenne est declarée significativement différente si :

$$\left|\bar{\mathbf{Y}}_{\mathbf{i}} - \bar{\mathbf{Y}}_{\mathbf{j}}\right| \; > \mathbf{t}_{\alpha/\mathbf{2},\mathbf{N}-\mathbf{a}} \times \sqrt{\mathbf{CM}_{\mathbf{SSE}}\left(\frac{\mathbf{1}}{\mathbf{n_i}} + \frac{\mathbf{1}}{\mathbf{n_j}}\right)}$$

La quantité différence moins significative est définie par :

$$\mathbf{LSD} = \mathbf{t}_{\alpha/\mathbf{2},\mathbf{N}-\mathbf{a}} \times \sqrt{\mathbf{CM}_{\mathbf{SSE}}\left(\frac{\mathbf{1}}{\mathbf{n_i}} + \frac{\mathbf{1}}{\mathbf{n_j}}\right)}$$

Si l'échantillon est de taille égale alors :

$$\mathbf{LSD} = \mathbf{t}_{\alpha/\mathbf{2},\mathbf{N}-\mathbf{a}} \times \sqrt{\frac{\mathbf{2CM}_{\mathbf{SSE}}}{\mathbf{n}}}$$