Examen 2020-2021

Josephson Junior R.

April 28, 2024

Table des matières

1.i Hypothèse que les cinq matériaux aient le même effet ...

Soit le modèle ANOVA à un facteur de contrôle :

$$\mathbf{Y_{ij}} = \mu + \alpha_{\mathbf{i}} + \varepsilon_{\mathbf{ij}}$$

On teste les hypothèses suivantes :

$$\begin{cases} \mathbf{H0} : & \alpha_{\mathbf{i}} = \mathbf{0} \\ \mathbf{Ha} : & \exists \mathbf{i} \ / \ \alpha_{\mathbf{i}} \neq \mathbf{0} \end{cases}$$

Sous H0 on définit la statistique de test :

$$\mathbf{F} = \frac{\frac{\mathbf{SS_{Traitement}}}{\mathbf{a-1}}}{\frac{\mathbf{SS_E}}{\mathbf{N-a}}} \sim \mathcal{F}(\mathbf{a-1, N-a}) \ \ (\mathbf{4, 15})$$

AN :

$$\bar{Y}_{1.} = 159.8 \ ; \ \bar{Y}_{2.} = 6.25 \ ; \ \bar{Y}_{3.} = 2942 \ ; \ \bar{Y}_{4.} = 5723 \ ; \ \bar{Y}_{5.} = 10.75$$

$$\mathbf{SS_{Traitement}} = \sum_{i=1}^{5}\sum_{j=1}^{4}\left(\bar{Y}_i - \bar{Y}\right)^2 = 4 \times \sum_{i=1}^{5}\bar{Y}_i^2 - 20 \times \bar{Y}^2 = 103.2 \times 10^6$$

$$\mathbf{SS_T} = \sum_{i=1}^{5}\sum_{j=1}^{4}\left(Y_{ij} - \bar{Y}\right)^2 = \sum_{i=1}^{5}\sum_{j=1}^{4} Y_{ij}^2 - 20 \times \bar{Y}^2 = 165.7 \times 10^6$$

$$\mathbf{SS_E} = \mathbf{SS_T} - \mathbf{SS_{Traitement}} = 62.5 \times 10^6$$

$$\mathbf{F} = \frac{103.2/4}{62.5/15} = 6.19 \; > \; \mathbf{F^c}$$

Donc on rejette H0 , les matériaux n'ont pas le même effet sur le temps moyen de défaillance.

1.ii Moyenne globale et les effets traitements

La moyenne globale est donnée par :

$$\bar{Y} = \frac{1}{5}\sum_{i=1}^{5} \bar{Y}_{i.} = 1768.4$$

Les effets traitements :

$$\begin{cases} \hat{\alpha}_{\mathbf{1}} = & \bar{Y}_{1.} - \bar{Y} = & -1608.6 \\ \hat{\alpha}_{\mathbf{2}} = & \bar{Y}_{2.} - \bar{Y} = & -1762.15 \\ \hat{\alpha}_{\mathbf{3}} = & \bar{Y}_{3.} - \bar{Y} = & 1173.6 \\ \hat{\alpha}_{\mathbf{4}} = & \bar{Y}_{4.} - \bar{Y} = & 3954.6 \\ \hat{\alpha}_{\mathbf{1}} = & \bar{Y}_{5.} - \bar{Y} = & -1757.65 \end{cases}$$

1.iii Tableau 2 - Mean difference - tstat

Traitement	Mean différence	T-test
1 vs 2	153.5	0.11
1 vs 3	-2782	-1.93
1 vs 4	-5563.25	-3.85
1 vs 5	149	0.10
2 vs 3	-2935.5	-2.03
2 vs 4	-5716.75	-3.96
2 vs 5	-4.5	-0.003
3 vs 4	-2781.25	-1.93
3 vs 5	2931	2.03
4 vs 5	5712.25	3.96

$$\mathbf{T - test = \frac{\bar{Y}_{i.} - \bar{Y}_{j.}}{Strandard\ Error}}$$

2. Analyse graphique

Le premier graphe donne une idée sur l'homoskédasticité de la variance car les valeurs des residus sont bien distincte si l'on regarde la ligne horizontale en 0. Le second graphique fournit la distribution des residus qui semblent suivre une loi normale.

B1. Transformation standard

$$Y_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

Pour standardiser les données il suffit d'appliquer la formule suivante :

$$\frac{Y_i - \bar{Y}_i}{\hat{\sigma}_i} \sim \mathcal{N}(0,1) \quad ; \quad \hat{\sigma}^2 = \frac{1}{4}\sum_{i=1}^{4}\left(Y_i - \bar{Y}_i\right)^2$$

B2. Analyse des données après transformation

AN :

$$\bar{Y}_{1.} = -0.79 \;;\; \bar{Y}_{2.} = -0.86 \;;\; \bar{Y}_{3.} = 0.77 \;;\; \bar{Y}_{4.} = 1.94 \;;\; \bar{Y}_{5.} = -0.86$$

$$SS_{Traitement} = \sum_{i=1}^{5}\sum_{j=1}^{4}\left(\bar{Y}_i - \bar{Y}\right)^2 = 4 \times \sum_{i=1}^{5}\bar{Y}_i^2 - 20 \times \bar{Y}^2 = 25.8$$

$$SS_T = \sum_{i=1}^{5}\sum_{j=1}^{4}\left(Y_{ij} - \bar{Y}\right)^2 = \sum_{i=1}^{5}\sum_{j=1}^{4} Y_{ij}^2 - 20 \times \bar{Y}^2 = 39.7$$

$$SS_E = SS_T - SS_{Traitement} = 6792.576$$

$$F = \frac{25.8/4}{13.9/15} = 6.96 \ > \ F^c$$

Après standardisation des données les matériaux n'ont toujours pas le même effet sur le temps moyen de défaillance.

B3. Tableau 3 - Mean difference - T-test

Traitement	Mean différence	T-test
1 vs 2	0.07	0.1
1 vs 3	-1.56	-2.11
1 vs 4	-2.73	-3.7
1 vs 5	0.07	0.1
2 vs 3	-1.63	-2.202
2 vs 4	-2.8	-3.78
2 vs 5	0	0
3 vs 4	-1.17	-1.58
3 vs 5	1.63	2.202
4 vs 5	2.8	3.78

B4. Analyse des résultats

Après standardisation on peut conclure que les residus présents sur le Normal plot suivent bien la loi Normal centré réduite. Ainsi que la variance semblent être homoskédastique.

B5i. Comparaison des paires moyennes par le test de Tukey

Soit le corps d'hypothèse :

$$\begin{cases} \textbf{H0} : & \mu_{\mathbf{i}} = \mu_{\mathbf{j}} \\ \textbf{Ha} : & \mu_{\mathbf{i}} \neq \mu_{\mathbf{j}} \end{cases}$$

La statistique du test est définie par :

$$q = \frac{\bar{Y}_{max} - \bar{Y}_{min}}{\sqrt{\frac{CM_{SSE}}{n}}}$$

Le seuil critique de Tukey est :

$$T_{\alpha} = q(a, f) \times \sqrt{\frac{CM_{SSE}}{n}} = 12.07$$

On dit que les paires moyennes sont significativement différentes si :

$$\left|\bar{Y}_i - \bar{Y}_j\right| > T_{alpha} \Rightarrow \textit{à vous de conclure}$$

B5ii. LSD-Fisher

$$LSD = t_{\alpha/2,N-a} \times \sqrt{\frac{2CM_{SSE}}{n}} = 1.58$$

On rejette H0 si :

$$\left|\bar{Y}_i - \bar{Y}_j\right| > LSD \Rightarrow \textit{à vous de conclure}$$

B5iii. Comparatif des deux tests

On constate que les deux tests donnent les mêmes résultats quant au rejet de H0. La seule différence entre les deux méthodes est que le Tukey-Test se base sur **le rang studentisé** tandis que le LSD de Fisher sur **la distribution de Student**.

1. $CM_{SSTraitement}$ est un estimateur biaisé de σ^2

$$E\left(\frac{\sum\sum\left(\bar{Y}_i-\bar{Y}\right)^2}{a-1}\right)=\frac{1}{a-1}E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}\left(\bar{Y}_i-\bar{Y}\right)^2\right] \quad (1)$$

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

$$\bar{Y}_i^2=\frac{1}{n^2}\left(\sum_{j=1}^{n}Y_{ij}\right)^2 \quad ; \quad \bar{Y}^2=\frac{1}{N^2}\left(\sum_{i=1}^{a}\sum_{j=1}^{n}Y_{ij}\right)^2$$

$$Y_{ij}=\mu+\alpha_i+\varepsilon_{ij} \quad ; \quad E\left[\sum_{j=1}^{n}\varepsilon_{ij}^2\right]=n\sigma^2 \quad ; \quad E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}\varepsilon_{ij}^2\right]=N\sigma^2$$

$$E\left[n\sum_{i=1}^{a}\bar{Y}_i^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{a}\left(\sum_{j=1}^{n}Y_{ij}\right)^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{a}\left(\sum_{j=1}^{n}\mu+\alpha_i+\varepsilon_{ij}\right)^2\right]$$

$$= \mathbf{N}\mu^{\mathbf{2}} + \mathbf{n}\sum_{\mathbf{i=1}}^{\mathbf{a}}\alpha_{\mathbf{i}}^{\mathbf{2}} + \mathbf{a}\sigma^{\mathbf{2}}$$

$$E\left[N\bar{Y}^2\right] = E\left[\frac{1}{N}\left(\sum_{i=1}^{a}\sum_{j=1}^{n}Y_{ij}\right)^2\right] = E\left[\frac{1}{N}\left(\sum_{i=1}^{a}\sum_{j=1}^{n}\mu+\alpha_i+\varepsilon_{ij}\right)^2\right]$$

$$= \mathbf{N}\mu^{\mathbf{2}} + \frac{\mathbf{n^2}}{\mathbf{N}}\sum_{\mathbf{i=1}}^{\mathbf{a}}\alpha_{\mathbf{i}}^{\mathbf{2}} + \sigma^{\mathbf{2}}$$

$$E\left[n\sum_{i=1}^{a}\bar{Y}_i^2 - N\bar{Y}^2\right] = \left(\frac{N-n}{N}\right)\times n\sum_{i=1}^{a}\alpha_i^2 + (a-1)\sigma^2$$

$$Or \quad \frac{N-n}{N} \approx 1 \Rightarrow \mathbf{(1)} = \sigma^{\mathbf{2}} + \frac{\mathbf{n}\sum_{\mathbf{i=1}}^{\mathbf{a}} \alpha_{\mathbf{i}}^{\mathbf{2}}}{\mathbf{a-1}} \neq \sigma^{\mathbf{2}}$$

Donc le Carré Moyen SStraitement est un estimateur biaisé de σ^2. Il est sans biais sous l'hypothèse **H0** : $\alpha_{\mathbf{i}} = \mathbf{0}$.

2. Sous Ha vraie la statistique de test

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

Le paramètre λ est commun au loi non centré (ici il s'agit d'un Fisher non centré).

$$\lambda = \frac{n\sum_{i=1}^{a} \alpha_i^2}{\sigma^2}$$

3. Procédure 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 :

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

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 ($\Gamma_u = 0$).

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