Plans d'Expériences

Josephson Junior R.

February 17, 2024

Table des matières

On cherche à vérifier s'il existe une différence significative entre les moyennes de deux différentes échantillons.
On suppose que les données $Y_{ij} \sim \mathcal{N}(\mu_i, \sigma_i^2)$ de plus **iid** càd plan d'expérience complètement randomisé défini par :

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

Soit les hypothèses à tester :

$$\begin{cases} H0 : & \mu_1 = \mu_2 \\ H1 : & \mu_1 \neq \mu_2 \end{cases}$$

La statistique de test est défini sous 3 cas possibles :

- Cas 1 : $\boldsymbol{\sigma_1^2 \neq \sigma_2^2}$ **connues**

$$\mathbf{Sous\ \ H0} : \mathbf{T} = \frac{\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2}{\sqrt{\frac{\sigma_1^2}{\mathbf{n_1}} + \frac{\sigma_2^2}{\mathbf{n_2}}}} \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$$

- Cas 2 : $\sigma_1^2 = \sigma_2^2 = \sigma^2$ **inconnues**

$$\textbf{Sous} \quad \mathbf{H0} : \mathbf{T} = \frac{\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2}{\hat{\sigma}\sqrt{\frac{1}{\mathbf{n}_1} + \frac{1}{\mathbf{n}_2}}} \sim \mathcal{T}(\mathbf{n}_1 + \mathbf{n}_2 - 2)$$

$$\hat{\sigma}^2 = \frac{(\mathbf{n}_1 - 1)\mathbf{S}_1^2 + (\mathbf{n}_2 - 1)\mathbf{S}_2^2}{\mathbf{n}_1 + \mathbf{n}_2 - 2}$$

- Cas 3 : $\sigma_1^2 \neq \sigma_2^2$ **inconnues**

$$\textbf{Sous} \quad \mathbf{H0} : \mathbf{T} = \frac{\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2}{\sqrt{\frac{\mathbf{S}_1^2}{\mathbf{n}_1} + \frac{\mathbf{S}_2^2}{\mathbf{n}_2}}} \sim \mathcal{T}(\mathbf{ddl})$$

La règle de décision :

$$\begin{cases} Si\ |T| \leq\ t_{\alpha/2} : & On\ accepte\ H0 \\ Si\ |T|\ >\ t_{\alpha/2} : & On\ rejette\ H0 \end{cases}$$

Soit $\theta = \mu_1 - \mu_2$, son estimateur est $\hat{\theta} = \bar{Y}_1 - \bar{Y}_2$. L'intervalle de confiance de θ est définit par :

$$\textbf{Sous H0}\ : \mathbf{IC}_{\theta}^{\mathbf{1}-\alpha} = \hat{\theta}\ \pm\ \mathbf{t}_{\alpha/\mathbf{2}}\ \hat{\sigma}\sqrt{\frac{\mathbf{1}}{\mathbf{n_1}} + \frac{\mathbf{1}}{\mathbf{n_2}}}$$

On se donne deux différentes formulation d'un mortier : **formulation modifiée** et **formulation non-modifiée** . Deux échantillons de 10 observations pour chaque formulation sont disponibles. On veut savoir s'**il existe une différence significative entre les moyennes des deux formulations**

	Mortier modifié $\mathbf{Y_{1j}}$	Mortier non-modifié $\mathbf{Y_{2j}}$
1	16.85	17.5
2	16.4	17.63
3	17.21	18.25
4	16.35	18
5	16.52	17.86
6	17.04	17.75
7	16.96	18.22
8	17.15	17.9
9	16.59	17.96
10	16.57	18.15

Application numérique

Soient les moyennes des deux formulations :

$$\bar{Y}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} Y_{1j} = 16.76 \ ; \bar{Y}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{2j} = 17.92$$

On teste l'égalité des moyennes des deux formulations selon deux cas imposés :

- Cas 1 : $\sigma_{\mathbf{1}}^{\mathbf{2}} = \sigma_{\mathbf{2}}^{\mathbf{2}} = \sigma^2$ **inconnues**

$$Sous \quad H0 : T = \frac{\bar{Y}_1 - \bar{Y}_2}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathcal{T}(18)$$

$$S_1^2 = \frac{1}{n_1 - 1}\left(\sum_{j=1}^{n_1} Y_{1j}^2 - n_1 \bar{Y}_1^2\right) = 0.1 \ ; \ S_2^2 = 0.061$$

$$\Longrightarrow \hat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 0.0805$$

$$\Longrightarrow T = \frac{16.76 - 17.93}{0.284 \times 0.45} = -9.143$$

Alors on rejette H0 car $|T| > t_{2.5\%}(18) = 2.1$.

- Cas 2 : $\boldsymbol{\sigma_1^2 \neq \sigma_2^2}$ **inconnues**

$$Sous \quad H0 : T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \mathcal{T}(ddl)$$

$$ddl = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}} = 18.46 \simeq 19$$

$$\Longrightarrow T = \frac{16.76 - 17.93}{0.1269} = -9.22$$

Alors on rejette H0 car $\mathbf{|T| > t_{2.5\%}(19) = 2.093}$.
Sous H0 (cas 1) l'IC est donné par :

$$IC_{\theta}^{1-\alpha} = [-1.16 \pm 0.27] = [-1.43, -0.89]$$

On remarque $\hat{\theta} \notin \mathbf{IC}_{\theta}^{1-\alpha}$ alors on rejette H0 c'est-à-dire qu'il existe une différence significative des moyennes des deux formulations.

Soit le modèle statistique suivant :

$$Y_{ij} = \mu_i + \beta_j + \varepsilon_{ij} \quad ; \quad \beta_{\mathbf{j}} : \ \textbf{effet additif de j}$$

Le modèle de comparaison appariée dans le cas où l'on a deux plans est :

$$d_j = Y_{1j} - Y_{2j}$$

On remarque que **l'effet additif disparait** pour ce modèle ; dans certains cas ce modèle permet d'améliorer la précison des résultats.
On note :

$$\mu_d = E(d_j) = \mu_1 - \mu_2$$

On teste la différence appariée par :

$$\begin{cases} H0 : & \mu_d = 0 \\ H1 : & \mu_d \neq 0 \end{cases}$$

Sous H0 on définit la statistique de test par :

$$\mathbf{t} = \frac{\bar{\mathbf{d}}}{\frac{\mathbf{S_d}}{\sqrt{\mathbf{n}}}} \sim \mathcal{T}(\mathbf{n}-\mathbf{1})\ ;\ \bar{d} = \frac{1}{n}\sum_{j=1}^{n} d_j\ ;\ S_d^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(d_j - \bar{d}\right)^2$$

Dureté d'u tuyau

On décide de travailler avec 10 specimens métalliques qui répondent aux conditions appliquées en comparaison de plan apparié. Les données sont présentes dans le tableau ci-dessous

T_1	7	3	3	4	8	3	2	9	5	4
T_2	6	3	5	3	8	2	4	9	4	5

T-test apparié

On définit :

$$d_j = T_{1j} - T_{2j}$$

Les hypothèses du T-test apparié est :

$$\begin{cases} H0: & \mu_d = 0 \\ H1: & \mu_d \neq 0 \end{cases}$$

Sous H0 la statistique de test est la suivante :

$$\mathbf{t} = \frac{\bar{\mathbf{d}}}{\frac{\mathbf{S_d}}{\sqrt{\mathbf{n}}}} \sim \mathcal{T}(\mathbf{9})$$

Application numérique :

$$\bar{d} = -\frac{1}{10} = -0.1 \quad ; \quad S_d^2 = \frac{12.9}{9} = 1.43 \Rightarrow S_d = 1.197$$

$$\mathbf{t} = \frac{\bar{\mathbf{d}}}{\frac{\mathbf{S_d}}{\sqrt{\mathbf{n}}}} = -\frac{0.1}{0.3785} = -0.264$$

On accepte H0 car $|\mathbf{t}| < \mathbf{t^c} = \mathbf{2.262} \Longrightarrow$ les moyennes de deux tuyaux sont significativement égales.

T-test simple

Sous H0 :

$$\mathbf{T} = \frac{\bar{\mathbf{T}}_1 - \bar{\mathbf{T}}_2}{\hat{\sigma}\sqrt{\frac{1}{\mathbf{n_1}} + \frac{1}{\mathbf{n_2}}}} \sim \mathcal{T}(\mathbf{18})$$

Applications numériques :

$$\bar{T}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} T_{1j} = 4.8 \quad ; \quad \bar{T}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} = T_{2j} = 4.9$$

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (T_{1j} - T_1)^2 = 5.73 \quad ; \quad S_2^2 = 5 \; ; \; \hat{\sigma}^2 = 5.365$$

$$T = \frac{4.8 - 4.9}{2.316 \times 0.447} = -0.096$$

On a que $|\mathbf{T}| > \mathbf{t^c} = \mathbf{2.101}$ alors on accepte H0 c'est à dire les moyennes des deux tuyaux sont significativement égales.

Intervalle de confiance

Pour le T-test apparié :

$$IC_{\mu_d}^{1-\alpha} = [-0.1 \pm 0.856] = [-0.956, 0.756]$$

Pour le T-test simple :

$$IC_{\theta}^{1-\alpha} = [-0.1 \pm 2.175] = [-2.275, 2.075]$$

Interprétation

Par le principe de blocage (**appariement**) on assiste à une perte de degré de liberté c-à-d une soit disant depréciatiob du test appliqué mais en contre partie **un gain de précision de la situation par l'élimination de l'effet additif entre les blocs**. On peut remarquer que le plan de comparaison appariée a réduit la variabilité d'estimation de **50 %** $\Rightarrow$ une IC plus étroite ce qui montre l'efficacité du test.

Test sur la variance d'une population normale

Sous les hypothèses suivantes :

$$\begin{cases} H0: & \sigma^2 = \sigma_0^2 \\ H1: & \sigma^2 \neq \sigma_0^2 \end{cases}$$

La statistique du test est définie par :

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1) \quad où \;\; S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Règle de décision $\Rightarrow \boldsymbol{\chi_0^2 \geq \chi_{1-\alpha/2}^2}$: **On rejette H0**
L'intervalle de confiance :

$$\mathbf{IC}_{\sigma^2}^{1-\alpha} = \left[\frac{(\mathbf{n}-\mathbf{1})\mathbf{S}^2}{\chi_{\alpha/2}^2}, \frac{(\mathbf{n}-\mathbf{1})\mathbf{S}^2}{\chi_{1-\alpha/2}^2}\right]$$

Egalité des variances d'une population normale iid

Sous les hypothèses suivantes :

$$\begin{cases} H0: & \sigma_1^2 = \sigma_2^2 \\ H1: & \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

Sous H0 on définit la statistique du test :

$$F_0 = \frac{S_1^2}{S_2^2} \sim \mathcal{F}(n_1 - 1, n_2 - 1)$$

Pour $\mathbf{F_0} > \mathbf{F}_{\alpha/\mathbf{2}}$ on rejette H0.
L'intervalle de confiance est donnée par :

$$\mathbf{IC}_\theta^{\mathbf{1}-\alpha} = \left[\frac{\mathbf{S_1^2}}{\mathbf{S_2^2}}\ \mathbf{F}_{\mathbf{1}-\alpha/\mathbf{2}}, \frac{\mathbf{S_1^2}}{\mathbf{S_2^2}}\ \mathbf{F}_{\alpha/\mathbf{2}}\right]$$

Application

On souhaite étudier la variabilité de deux équipements 1 et 2. On soupçonne que l'équipement de type 1 ait une plus grande variabilité que celle du type 2.

$$\begin{cases} H0: & \sigma_1^2 = \sigma_2^2 \\ H1: & \sigma_1^2 > \sigma_2^2 \end{cases}$$

On donne $\mathbf{n_1 = 12}$; $\mathbf{n_2 = 10}$; $\mathbf{S_1 = 14.5}$; $\mathbf{S_2 = 10.8}$

Solution

Sous H0 on définit la statistique de test :

$$F_0 = \frac{S_1^2}{S_2^2} \sim \mathcal{F}(11, 9) \Longrightarrow \mathbf{F_0} = \mathbf{\frac{14.5^2}{10.8^2}} = \mathbf{1.803}$$

On accepte H0 car $\mathbf{F_0} < \mathbf{F^c} = \mathbf{3.10}$ donc les deux équipements sont de mêmes variabilités.