Modèles Linéaires
Chapitre 3: Analyse de la Variance à un Facteur

Pr. Mokhtar KOUKI

Université de Carthage (ESSAI)
mokhtar.kouki@essai.ucar.tn

Novembre 2021

Contenu

On considère une expérience de plantation de quatre variétés d'un arbre fruitier plantées sur la même parcelle. Et on s'intéresse au rendement annuel par arbre. On dispose des informations suivantes :

Variété j	N_j	$Q_{1j} = \sum_{i=1}^{N_j} R_{ij}$	$Q_{2j} = \sum_{i=1}^{N_j} \left(R_{ij} - \overline{R}_j\right)^2$
1	30	53.84	64.56
2	20	66.00	30.04
3	40	183.65	90.40
4	35	34.47	50.44

On considère une expérience de plantation de quatre variétés d'un arbre fruitier plantées sur la même parcelle. Et on s'intéresse au rendement annuel par arbre. On dispose des informations suivantes :

Variété j	N_j	$Q_{1j} = \sum_{i=1}^{N_j} R_{ij}$	$Q_{2j} = \sum_{i=1}^{N_j} (R_{ij} - \overline{R}_j)^2$
1	30	53.84	64.56
2	20	66.00	30.04
3	40	183.65	90.40
4	35	34.47	50.44

Question :Y-a-t-il une différence significative entre les variétés en terme de rendement ? On parle d'effet variété dans le rendement.

On considère R_{ij} le rendement de l'arbre i de la variété j.
Et on suppose que $R_{ij} \rightsquigarrow iid\ N(m_j, \sigma_j^2)$.

On considère R_{ij} le rendement de l'arbre i de la variété j.
Et on suppose que $R_{ij} \rightsquigarrow iid\ N(m_j, \sigma_j^2)$.
On admet que $\sigma_j^2 = \sigma^2\ \forall j$. Dans un premier temps, on s'intéresse à la comparaison entre les deux variétés 1 et 2.

On considère R_{ij} le rendement de l'arbre i de la variété j.
Et on suppose que $R_{ij} \rightsquigarrow iid\ N(m_j, \sigma_j^2)$.
On admet que $\sigma_j^2 = \sigma^2\ \forall j$. Dans un premier temps, on s'intéresse à la comparaison entre les deux variétés 1 et 2.
Sachant que $E(R_{ij}) = m_j$, $\hat{m}_j = \overline{R}_j$.
Tester que les deux variétés donneraient le même rendement, ça revient à tester l'hypothèse :

$$H_0 : m_1 = m_2 \text{ vs } H_1 : m_1 \neq m_2$$

On considère R_{ij} le rendement de l'arbre i de la variété j.
Et on suppose que $R_{ij} \rightsquigarrow iid\ N(m_j, \sigma_j^2)$.
On admet que $\sigma_j^2 = \sigma^2\ \forall j$. Dans un premier temps, on s'intéresse à la comparaison entre les deux variétés 1 et 2.
Sachant que $E(R_{ij}) = m_j$, $\hat{m}_j = \overline{R}_j$.
Tester que les deux variétés donneraient le même rendement, ça revient à tester l'hypothèse :

$$H_0 : m_1 = m_2 \text{ vs } H_1 : m_1 \neq m_2$$

Sous l'hypothèse de la normalité :

$$\overline{R}_1 \rightsquigarrow N\left(m_1, \frac{\sigma^2}{N_1}\right) \text{ et } \overline{R}_2 \rightsquigarrow N\left(m_2, \frac{\sigma^2}{N_2}\right)$$

Par conséquent :

$$\overline{R}_1 - \overline{R}_2 \rightsquigarrow N\left(m_1 - m_2, \sigma^2\left[\frac{1}{N_1} + \frac{1}{N_2}\right]\right)$$

Par conséquent :

$$\overline{R}_1 - \overline{R}_2 \rightsquigarrow N\left(m_1 - m_2, \sigma^2\left[\frac{1}{N_1} + \frac{1}{N_2}\right]\right)$$

Sous H_0

$$T = \frac{\overline{R}_1 - \overline{R}_2}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \rightsquigarrow N\left(0, 1\right)$$

Par conséquent :

$$\overline{R}_1 - \overline{R}_2 \rightsquigarrow N\left(m_1 - m_2, \sigma^2\left[\frac{1}{N_1} + \frac{1}{N_2}\right]\right)$$

Sous H_0

$$T = \frac{\overline{R}_1 - \overline{R}_2}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \rightsquigarrow N(0, 1)$$

L'estimateur sans biais de la variance est défini par :

$$\hat{\sigma}^2 = \frac{1}{N_1 + N_2 - 2}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\left(R_{ij} - \overline{R}_j\right)^2$$

Remarque : Construction de l'estimateur sans biais de l'estimateur de la variance :

Remarque : Construction de l'estimateur sans biais de l'estimateur de la variance : La vraisemblance des deux échantillons définie par :

$$f(R_{ij,i=1,\cdots,N_j,j=1,2}) = \prod_{i=1}^{N_1} f(R_{i1}) \prod_{i=1}^{N_2} f(R_{i2})$$

Remarque : Construction de l'estimateur sans biais de l'estimateur de la variance : La vraisemblance des deux échantillons définie par :

$$f(R_{ij,i=1,\cdots,N_j,j=1,2}) = \prod_{i=1}^{N_1} f(R_{i1}) \prod_{i=1}^{N_2} f(R_{i2})$$

$$= \prod_{i=1}^{N_1} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(R_{i1} - m_1)^2\right)$$

Remarque : Construction de l'estimateur sans biais de l'estimateur de la variance : La vraisemblance des deux échantillons définie par :

$$f(R_{ij,i=1,\cdots,N_j,j=1,2}) = \prod_{i=1}^{N_1} f(R_{i1}) \prod_{i=1}^{N_2} f(R_{i2})$$

$$= \prod_{i=1}^{N_1} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(R_{i1}-m_1)^2\right)$$

$$\times \prod_{i=1}^{N_2} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(R_{i2}-m_2)^2\right)$$

Remarque : Construction de l'estimateur sans biais de l'estimateur de la variance : La vraisemblance des deux échantillons définie par :

$$f(R_{ij,i=1,\cdots,N_j,j=1,2}) = \prod_{i=1}^{N_1} f(R_{i1}) \prod_{i=1}^{N_2} f(R_{i2})$$

$$\begin{aligned}
&= \prod_{i=1}^{N_1} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(R_{i1}-m_1)^2\right) \\
&\quad \times \prod_{i=1}^{N_2} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(R_{i2}-m_2)^2\right) \\
&= (2\pi)^{-(N_1+N_2)/2} \left(\sigma^2\right)^{-(N_1+N_2)/2} \\
&\quad \times \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N_1}(R_{i1}-m_1)^2 + \sum_{i=1}^{N_2}(R_{i2}-m_2)^2\right)\right]
\end{aligned}$$

Et le logarithme de la vraisemblance est :

$$L = -\frac{(N_1 + N_2)}{2}\ln(2\pi) - \frac{(N_1 + N_2)}{2}\ln(\sigma^2)$$

Et le logarithme de la vraisemblance est :

$$\begin{aligned} L &= -\frac{(N_1+N_2)}{2}\ln(2\pi) - \frac{(N_1+N_2)}{2}\ln(\sigma^2) \\ &\quad -\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N_1}(R_{i1}-m_1)^2 + \sum_{i=1}^{N_2}(R_{i2}-m_2)^2\right) \end{aligned}$$

Et le logarithme de la vraisemblance est :

$$\begin{aligned} L &= -\frac{(N_1+N_2)}{2}\ln(2\pi) - \frac{(N_1+N_2)}{2}\ln(\sigma^2) \\ &\quad -\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N_1}(R_{i1}-m_1)^2 + \sum_{i=1}^{N_2}(R_{i2}-m_2)^2\right) \end{aligned}$$

Les conditions nécessaires de premier ordre de la maximisation de L par rapport à m_1, m_2 et σ^2 sont :

Et le logarithme de la vraisemblance est :

$$\begin{aligned} L &= -\frac{(N_1+N_2)}{2}\ln(2\pi) - \frac{(N_1+N_2)}{2}\ln(\sigma^2) \\ &\quad -\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N_1}(R_{i1}-m_1)^2 + \sum_{i=1}^{N_2}(R_{i2}-m_2)^2\right) \end{aligned}$$

Les conditions nécessaires de premier ordre de la maximisation de L par rapport à m_1, m_2 et σ^2 sont :

$$\frac{\partial L}{\partial m_1} = \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{N_1} 2(R_{i1}-\hat{m}_1) = 0$$

Et le logarithme de la vraisemblance est :

$$\begin{aligned} L &= -\frac{(N_1+N_2)}{2}\ln(2\pi) - \frac{(N_1+N_2)}{2}\ln(\sigma^2) \\ &\quad -\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N_1}(R_{i1}-m_1)^2 + \sum_{i=1}^{N_2}(R_{i2}-m_2)^2\right) \end{aligned}$$

Les conditions nécessaires de premier ordre de la maximisation de L par rapport à m_1, m_2 et σ^2 sont :

$$\begin{aligned} \frac{\partial L}{\partial m_1} &= \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{N_1} 2(R_{i1}-\hat{m}_1) = 0 \\ \frac{\partial L}{\partial m_2} &= \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{N_2} 2(R_{i2}-\hat{m}_2) = 0 \end{aligned}$$

Et le logarithme de la vraisemblance est :

$$
\begin{aligned}
L &= -\frac{(N_1+N_2)}{2}\ln(2\pi) - \frac{(N_1+N_2)}{2}\ln(\sigma^2) \\
&\quad -\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N_1}(R_{i1}-m_1)^2 + \sum_{i=1}^{N_2}(R_{i2}-m_2)^2\right)
\end{aligned}
$$

Les conditions nécessaires de premier ordre de la maximisation de L par rapport à m_1, m_2 et σ^2 sont :

$$
\begin{aligned}
\frac{\partial L}{\partial m_1} &= \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{N_1} 2(R_{i1}-\hat{m}_1) = 0 \\
\frac{\partial L}{\partial m_2} &= \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{N_2} 2(R_{i2}-\hat{m}_2) = 0 \\
\frac{\partial L}{\partial \sigma^2} &= -\frac{(N_1+N_2)}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\left(\sum_{i=1}^{N_1}(R_{i1}-\hat{m}_1)^2 + \sum_{i=1}^{N_2}(R_{i2}-\hat{m}_2)^2\right) = 0
\end{aligned}
$$

Ce qui donne comme estimateurs :

$$\begin{aligned}
\hat{m}_1 &= \frac{1}{N_1}\sum_{i=1}^{N_1} R_{i1} = \overline{R}_1 \\
\hat{m}_2 &= \frac{1}{N_2}\sum_{i=1}^{N_2} R_{i2} = \overline{R}_2 \\
\hat{\sigma}^2_{mv} &= \frac{1}{N_2 + N_2}\left(\sum_{i=1}^{N_1}(R_{i1} - \overline{R}_1)^2 + \sum_{i=1}^{N_2}(R_{i2} - \overline{R}_2)^2\right)
\end{aligned}$$

Ce qui donne comme estimateurs :

$$\hat{m}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} R_{i1} = \overline{R}_1$$

$$\hat{m}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} R_{i2} = \overline{R}_2$$

$$\hat{\sigma}^2_{mv} = \frac{1}{N_2+N_2}\left(\sum_{i=1}^{N_1}(R_{i1}-\overline{R}_1)^2 + \sum_{i=1}^{N_2}(R_{i2}-\overline{R}_2)^2\right)$$

Sur un autre plan :

$$\frac{\sum_{i=1}^{N_1}\left(R_{i1}-\overline{R}_1\right)^2}{\sigma^2} \rightsquigarrow \chi^2(N_1-1) \text{ et } \frac{\sum_{i=1}^{N_2}\left(R_{i2}-\overline{R}_2\right)^2}{\sigma^2} \rightsquigarrow \chi^2(N_2-1)$$

Ce qui implique que :

$$\frac{\sum_{j=1}^{2} \sum_{i=1}^{N_j} \left(R_{ij} - \overline{R}_j\right)^2}{\sigma^2} =$$

Ce qui implique que :

$$\frac{\sum_{j=1}^{2}\sum_{i=1}^{N_j}\left(R_{ij}-\overline{R}_j\right)^2}{\sigma^2} = (N_1+N_2-2)\,\frac{\hat{\sigma}^2}{\sigma^2} \rightsquigarrow \chi^2(N_1+N_2-2)$$

Et

$$E\left(\hat{\sigma}^2_{mv}\right) = \frac{1}{N_1+N_2}\sigma^2 E\left(\frac{\sum_{j=1}^{2}\sum_{i=1}^{N_j}\left(R_{ij}-\overline{R}_j\right)^2}{\sigma^2}\right)$$

Ce qui implique que :

$$\frac{\sum_{j=1}^{2}\sum_{i=1}^{N_j}\left(R_{ij}-\overline{R}_j\right)^2}{\sigma^2} = (N_1+N_2-2)\,\frac{\hat{\sigma}^2}{\sigma^2} \rightsquigarrow \chi^2(N_1+N_2-2)$$

Et

$$\begin{aligned} E\left(\hat{\sigma}^2_{mv}\right) &= \frac{1}{N_1+N_2}\sigma^2 E\left(\frac{\sum_{j=1}^{2}\sum_{i=1}^{N_j}\left(R_{ij}-\overline{R}_j\right)^2}{\sigma^2}\right) \\ &= \frac{1}{N_1+N_2}\sigma^2 \times (N_1+N_2-2) \end{aligned}$$

Ce qui implique que :

$$\frac{\sum_{j=1}^{2}\sum_{i=1}^{N_j}\left(R_{ij}-\overline{R}_j\right)^2}{\sigma^2} = (N_1+N_2-2)\,\frac{\hat{\sigma}^2}{\sigma^2} \rightsquigarrow \chi^2(N_1+N_2-2)$$

Et

$$\begin{aligned} E\left(\hat{\sigma}^2_{mv}\right) &= \frac{1}{N_1+N_2}\sigma^2 E\left(\frac{\sum_{j=1}^{2}\sum_{i=1}^{N_j}\left(R_{ij}-\overline{R}_j\right)^2}{\sigma^2}\right) \\ &= \frac{1}{N_1+N_2}\sigma^2\times(N_1+N_2-2) \end{aligned}$$

D'où, l'estimateur sans biais de σ^2 :

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{2}\sum_{i=1}^{N_j}\left(R_{ij}-\overline{R}_j\right)^2}{N_1+N_2-2}$$

En conclusion :

$$\hat{T} = \frac{\overline{R}_1 - \overline{R}_2}{\hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} =$$

En conclusion :

$$\hat{T} = \frac{\overline{R}_1 - \overline{R}_2}{\hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = \frac{\overline{R}_1 - \overline{R}_2}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \Big/ \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} =$$

En conclusion :

$$\hat{T} = \frac{\overline{R}_1 - \overline{R}_2}{\hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = \frac{\overline{R}_1 - \overline{R}_2}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \Big/ \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} = \frac{N(0,1)}{\sqrt{\frac{\chi^2(N_1+N_2-2)}{N_1+N_2-2}}} \rightsquigarrow St(N_1+N_2-2$$

En conclusion :

$$\hat{T} = \frac{\overline{R}_1 - \overline{R}_2}{\hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = \frac{\overline{R}_1 - \overline{R}_2}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \Big/ \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} = \frac{N(0,1)}{\sqrt{\frac{\chi^2(N_1+N_2-2)}{N_1+N_2-2}}} \rightsquigarrow St(N_1+N_2-2$$

Et pour α donné, on rejette l'hypthèse H_0 si $\left|\hat{T}\right| > t^{1-\alpha/2}(N_1 + N_2 - 2)$.
Observation 1 : L'estimateur de la variance est un estimateur pondéré :

$$\hat{\sigma}^2 = \frac{1}{N_1 + N_2 - 2} \sum_{j=1}^{2} \sum_{i=1}^{N_j} \left(R_{ij} - \overline{R}_j\right)^2$$

En conclusion :

$$\hat{T} = \frac{\overline{R}_1 - \overline{R}_2}{\hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = \frac{\overline{R}_1 - \overline{R}_2}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \Big/ \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} = \frac{N(0,1)}{\sqrt{\frac{\chi^2(N_1+N_2-2)}{N_1+N_2-2}}} \rightsquigarrow St(N_1+N_2-2$$

Et pour α donné, on rejette l'hypthèse H_0 si $\left|\hat{T}\right| > t^{1-\alpha/2}(N_1 + N_2 - 2)$.

Observation 1 : L'estimateur de la variance est un estimateur pondéré :

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N_1 + N_2 - 2} \sum_{j=1}^{2} \sum_{i=1}^{N_j} \left(R_{ij} - \overline{R}_j\right)^2 \\ &= \frac{1}{N_1 + N_2 - 2} \left[(N_1 - 1) \frac{\sum_{i=1}^{N_1} \left(R_{i1} - \overline{R}_1\right)^2}{N_1 - 1} \right. \end{aligned}$$

En conclusion :

$$\hat{T} = \frac{\overline{R}_1 - \overline{R}_2}{\hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = \frac{\overline{R}_1 - \overline{R}_2}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \Big/ \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} = \frac{N(0,1)}{\sqrt{\frac{\chi^2(N_1+N_2-2)}{N_1+N_2-2}}} \rightsquigarrow St(N_1+N_2-2$$

Et pour α donné, on rejette l'hypthèse H_0 si $\left|\hat{T}\right| > t^{1-\alpha/2}(N_1 + N_2 - 2)$.
Observation 1 : L'estimateur de la variance est un estimateur pondéré :

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{N_1 + N_2 - 2} \sum_{j=1}^{2} \sum_{i=1}^{N_j} \left(R_{ij} - \overline{R}_j\right)^2 \\
&= \frac{1}{N_1 + N_2 - 2} \left[(N_1 - 1) \frac{\sum_{i=1}^{N_1} \left(R_{i1} - \overline{R}_1\right)^2}{N_1 - 1} \right. \\
&\quad \left. + (N_2 - 1) \frac{\sum_{i=1}^{N_2} \left(R_{i2} - \overline{R}_2\right)^2}{N_2 - 1} \right]
\end{aligned}$$

D'où :

$$\hat{\sigma}^2 = \frac{1}{N_1 + N_2 - 2}\left[(N_1 - 1)\hat{\sigma}_1^2 + (N_2 - 1)\hat{\sigma}_2^2\right]$$

Obersevation 2 : $\hat{T}^2 \rightsquigarrow F(1, N_1 + N_2 - 2)$

$$\hat{T}^2 = \frac{N(0,1)^2}{\frac{\chi^2(N_1+N_2-2)}{N_1+N_2-2}}$$

D'où :

$$\hat{\sigma}^2 = \frac{1}{N_1 + N_2 - 2}\left[(N_1 - 1)\hat{\sigma}_1^2 + (N_2 - 1)\hat{\sigma}_2^2\right]$$

Obersevation 2 : $\hat{T}^2 \rightsquigarrow F(1, N_1 + N_2 - 2)$

$$\hat{T}^2 = \frac{N(0,1)^2}{\frac{\chi^2(N_1+N_2-2)}{N_1+N_2-2}} = \frac{\frac{\chi^2(1)}{1}}{\frac{\chi^2(N_1+N_2-2)}{N_1+N_2-2}} \rightsquigarrow F(1; N_1 + N_2 - 2)$$

Application numérique (Fichier Excel) :

$$\begin{aligned} N_1 &= 30, N_2 = 20, \overline{R}_1 = 1.795, \overline{R}_2 = 3.3 \\ \alpha &= 5\%, \hat{\sigma}^2 = 1.971 \\ \hat{T} &= -3.71350, t^{0.975}(48) = 2.01063, \text{ et } F^{5\%}(1; 48) = 4.04265 \end{aligned}$$

Conclusion : Le rendement dépend de la variété.

L'analyse de la variance à un facteur permet de généralise la comparaison entre un nombre de modalités supérieur à 2

Définition

On considère une variable aléatoire réel d'espérnce mathématique μ et de variance σ^2. le coefficient d'asymétrie (skewnes, Sk) et le coefficient d'aplatissement (Kurtosis) sont définis par :

$$SK = \frac{E\left((X-\mu)^3\right)}{\sigma^3}$$
$$K = \frac{E\left((X-\mu)^4\right)}{\sigma^4}$$

Cas particulier : Si $X \rightsquigarrow N(\mu, \sigma^2)$, $E\left((X-\mu)^4\right) = 3\sigma^4$. C'est pour cette raison qu'on parle souvent d'un excès de kurtosis.

Si on a un échantillon $(X_1, X_2, \cdots, X_N)$ iid issue, le skewness et le kurtosis sont estimés par :

$$\begin{aligned}
\hat{SK} &= \frac{1}{\hat{\sigma}^3} \frac{\sum_{i=1}^{N}(X_i - \overline{X})^3}{N} \\
\hat{K} &= \frac{1}{\hat{\sigma}^4} \frac{\sum_{i=1}^{N}(X_i - \overline{X})^4}{N} \\
\hat{\sigma}^2 &= \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1}
\end{aligned}$$

Exercice : Utiliser R pour calculer le skewness et le kurtosis pour des séries de variables aléatoires issues d'une loi normale (rnorm), une loi exponentielle (rgamma), une loi de student (rt)

Définition

La statistique test de Jarque-Bera (JB) de l'hypothèse de la normalité d'une série est définie par :

$$JB = \frac{N}{6}\left(\hat{SK}^2 + \frac{(\hat{K}-3)^2}{4}\right) \rightsquigarrow \chi^2(2)$$

NB : Si la série correspond à des résidus d'une régression, il faut remplacer N par $N-k$, k étant le nombre de variables explicatives.

Exercice : A l'aide du logiciel R, réaliser les test de normalité d'une série de variables aléatoires issues d'une loi normale (rnorm), une loi exponentielle (rgamma), une loi de student (rt)
(library(normtest))

Définition

On considère des réalisations Y_{ij}, $i = 1, 2, \cdots, N_j$ et $j = 1, 2, \cdots, g$ issues de variables aléatoires de variances respectives σ_j^2 (i étant l'individu et j le groupe). La statistique de Bartlett de l''hypothèse d'égalité de la variance est définie par :

$$B = (N - g)\ln(\hat{\sigma}^2) - \sum_{j=1}^{g} (N_j - 1)\ln(\hat{\sigma}_j^2)$$

Sous l'hypothèse d'égalité des variances, $B \rightsquigarrow \chi^2(g-1)$

$N = \sum_{j=1}^{g} N_j$, $\hat{\sigma}_j^2 = \frac{1}{N_j - 1}\sum_{i=1}^{N_j}\left(Y_{ij} - \bar{Y}_j\right)^2$, $\hat{\sigma}^2 = \frac{1}{N-g}\sum_{j=1}^{g}(N_j - 1)\hat{\sigma}_j^2$

Exercice 1 : Pour $X \rightsquigarrow N(0;\sigma^2)$**, calculer** $E(X^{2k}), k \geq 1$

On considère $X \rightsquigarrow N(0,\sigma^2)$.

$$E(X^{2k}) = \int_{-\infty}^{+\infty} x^{2k} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}x^2)dx$$

Exercice 1 : Pour $X \rightsquigarrow N(0;\sigma^2)$, calculer $E(X^{2k}), k \geq 1$

On considère $X \rightsquigarrow N(0,\sigma^2)$.

$$\begin{aligned} E(X^{2k}) &= \int_{-\infty}^{+\infty} x^{2k} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}x^2)dx \\ &= 2\int_{0}^{+\infty} x^{2k} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}x^2)dx \end{aligned}$$

Exercice 1 : Pour $X \rightsquigarrow N(0; \sigma^2)$, calculer $E(X^{2k}), k \geq 1$
On considère $X \rightsquigarrow N(0, \sigma^2)$.

$$\begin{aligned} E(X^{2k}) &= \int_{-\infty}^{+\infty} x^{2k} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}x^2)dx \\ &= 2\int_{0}^{+\infty} x^{2k} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}x^2)dx \end{aligned}$$

On considère le changement de variable suivant :

$$y = \frac{1}{2\sigma^2}x^2 \Longrightarrow x = \sigma\sqrt{2}y^{1/2}$$

Exercice 1 : Pour $X \rightsquigarrow N(0;\sigma^2)$, calculer $E(X^{2k}), k \geq 1$

On considère $X \rightsquigarrow N(0,\sigma^2)$.

$$\begin{aligned} E(X^{2k}) &= \int_{-\infty}^{+\infty} x^{2k} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}x^2)dx \\ &= 2\int_{0}^{+\infty} x^{2k} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}x^2)dx \end{aligned}$$

On considère le changement de variable suivant :

$$y = \frac{1}{2\sigma^2}x^2 \Longrightarrow x = \sigma\sqrt{2}y^{1/2}$$

Et

$$dx = \sigma\frac{\sqrt{2}}{2}y^{-1/2}dy$$

Ce qui permet d'écrire :

$$E(X^{2k}) \quad = \quad 2\int_0^{+\infty} \left(\sigma\sqrt{2}y^{1/2}\right)^{2k} \frac{1}{\sqrt{2\pi}\sigma} \exp(-y)\sigma\frac{\sqrt{2}}{2}y^{-1/2}dy$$

Ce qui permet d'écrire :

$$\begin{aligned} E(X^{2k}) &= 2\int_0^{+\infty} \left(\sigma\sqrt{2}y^{1/2}\right)^{2k} \frac{1}{\sqrt{2\pi}\sigma} \exp(-y)\sigma\frac{\sqrt{2}}{2}y^{-1/2}dy \\ &= \int_0^{+\infty} \left(\sigma\sqrt{2}y^{1/2}\right)^{2k} \frac{1}{\sqrt{\pi}} \exp(-y)y^{-1/2}dy \end{aligned}$$

Ce qui permet d'écrire :

$$
\begin{aligned}
E(X^{2k}) &= 2\int_0^{+\infty}\left(\sigma\sqrt{2}y^{1/2}\right)^{2k}\frac{1}{\sqrt{2\pi}\sigma}\exp(-y)\sigma\frac{\sqrt{2}}{2}y^{-1/2}dy \\
&= \int_0^{+\infty}\left(\sigma\sqrt{2}y^{1/2}\right)^{2k}\frac{1}{\sqrt{\pi}}\exp(-y)y^{-1/2}dy \\
&= \int_0^{+\infty}\sigma^{2k}2^k y^{k-\frac{1}{2}}\exp(-y)dy
\end{aligned}
$$

Ce qui permet d'écrire :

$$\begin{aligned}
E(X^{2k}) &= 2\int_0^{+\infty} \left(\sigma\sqrt{2}y^{1/2}\right)^{2k} \frac{1}{\sqrt{2\pi}\sigma} \exp(-y)\sigma\frac{\sqrt{2}}{2}y^{-1/2}dy \\
&= \int_0^{+\infty} \left(\sigma\sqrt{2}y^{1/2}\right)^{2k} \frac{1}{\sqrt{\pi}} \exp(-y)y^{-1/2}dy \\
&= \int_0^{+\infty} \sigma^{2k}2^k y^{k-\frac{1}{2}} \exp(-y)dy \\
&= \frac{1}{\sqrt{\pi}}\sigma^{2k}2^k\Gamma(k+\frac{1}{2})
\end{aligned}$$

Ce qui permet d'écrire :

$$
\begin{aligned}
E(X^{2k}) &= 2\int_0^{+\infty}\left(\sigma\sqrt{2}y^{1/2}\right)^{2k}\frac{1}{\sqrt{2\pi}\sigma}\exp(-y)\sigma\frac{\sqrt{2}}{2}y^{-1/2}dy \\
&= \int_0^{+\infty}\left(\sigma\sqrt{2}y^{1/2}\right)^{2k}\frac{1}{\sqrt{\pi}}\exp(-y)y^{-1/2}dy \\
&= \int_0^{+\infty}\sigma^{2k}2^k y^{k-\frac{1}{2}}\exp(-y)dy \\
&= \frac{1}{\sqrt{\pi}}\sigma^{2k}2^k\Gamma(k+\frac{1}{2})
\end{aligned}
$$

Conclusion : Si $X \rightsquigarrow N(0,\sigma^2)$:

$$E(X^{2k}) = \frac{1}{\sqrt{\pi}}\sigma^{2k}2^k\Gamma(k+\frac{1}{2})$$

Sachant que pour une fonction gamma on a :

$$\Gamma(p+1) = p\Gamma(p) \text{ et } \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

Sachant que pour une fonction gamma on a :

$$\Gamma(p+1) = p\Gamma(p) \text{ et } \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

pour $k = 2$,

$$E(X^4) = \frac{1}{\sqrt{\pi}}\sigma^4 4\Gamma(2+\frac{1}{2}) =$$

Sachant que pour une fonction gamma on a :

$$\Gamma(p+1) = p\Gamma(p) \text{ et } \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

pour $k = 2$,

$$E(X^4) = \frac{1}{\sqrt{\pi}}\sigma^4 4\Gamma(2+\frac{1}{2}) = \frac{1}{\sqrt{\pi}}\sigma^4 4\frac{3}{2}\Gamma(1+\frac{1}{2})$$

Sachant que pour une fonction gamma on a :

$$\Gamma(p+1) = p\Gamma(p) \text{ et } \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

pour $k = 2$,

$$\begin{aligned} E(X^4) &= \frac{1}{\sqrt{\pi}}\sigma^4 4\Gamma(2+\frac{1}{2}) = \frac{1}{\sqrt{\pi}}\sigma^4 4\frac{3}{2}\Gamma(1+\frac{1}{2}) \\ &= \frac{1}{\sqrt{\pi}}\sigma^4 4\frac{3}{2}\frac{1}{2}\Gamma(\frac{1}{2}) \end{aligned}$$

Sachant que pour une fonction gamma on a :

$$\Gamma(p+1) = p\Gamma(p) \text{ et } \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

pour $k = 2$,

$$\begin{aligned} E(X^4) &= \frac{1}{\sqrt{\pi}}\sigma^4 4\Gamma(2+\frac{1}{2}) = \frac{1}{\sqrt{\pi}}\sigma^4 4\frac{3}{2}\Gamma(1+\frac{1}{2}) \\ &= \frac{1}{\sqrt{\pi}}\sigma^4 4\frac{3}{2}\frac{1}{2}\Gamma(\frac{1}{2}) = 3\sigma^4 \end{aligned}$$

Exercice 2 : On considère deux échantillons $(X_1, X_2, \cdots, X_p)$ issu d'une loi $N(m, \sigma_1^2)$ et $(Y_1, Y_2, \cdots, Y_q)$ issu d'une loi $N(m, \sigma_2^2)$. Montrer que l'estimateur par maximum de vraisemblance de m peut être écrit sous forme d'une combinaison convexe dont on déterminera les composantes

Exercice 2 : On considère deux échantillons $(X_1, X_2, \cdots, X_p)$ issu d'une loi $N(m, \sigma_1^2)$ et $(Y_1, Y_2, \cdots, Y_q)$ issu d'une loi $N(m, \sigma_2^2)$. Montrer que l'estimateur par maximum de vraisemblance de m peut être écrit sous forme d'une combinaison convexe dont on déterminera les composantes

La vraisemblance des deux échantillons définie par :

$$f(X_1, X_2, \cdots, X_p, Y_1, Y_2, \cdots, Y_q) = \prod_{i=1}^{p} f_X(X_i) \prod_{j=1}^{q} f_Y(Y_j)$$

Exercice 2 : On considère deux échantillons $(X_1, X_2, \cdots, X_p)$ issu d'une loi $N(m, \sigma_1^2)$ et $(Y_1, Y_2, \cdots, Y_q)$ issu d'une loi $N(m, \sigma_2^2)$. Montrer que l'estimateur par maximum de vraisemblance de m peut être écrit sous forme d'une combinaison convexe dont on déterminera les composantes

La vraisemblance des deux échantillons définie par :

$$f(X_1, X_2, \cdots, X_p, Y_1, Y_2, \cdots, Y_q) = \prod_{i=1}^{p} f_X(X_i) \prod_{j=1}^{q} f_Y(Y_j)$$

$$= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(X_i - m)^2\right)$$

Exercice 2 : On considère deux échantillons $(X_1, X_2, \cdots, X_p)$ issu d'une loi $N(m, \sigma_1^2)$ et $(Y_1, Y_2, \cdots, Y_q)$ issu d'une loi $N(m, \sigma_2^2)$. Montrer que l'estimateur par maximum de vraisemblance de m peut être écrit sous forme d'une combinaison convexe dont on déterminera les composantes

La vraisemblance des deux échantillons définie par :

$$f(X_1, X_2, \cdots, X_p, Y_1, Y_2, \cdots, Y_q) = \prod_{i=1}^{p} f_X(X_i) \prod_{j=1}^{q} f_Y(Y_j)$$

$$= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(X_i - m)^2\right)$$

$$\times \prod_{j=1}^{q} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2\sigma_2^2}(Y_j - m)^2\right)$$

$$
\begin{aligned}
= \quad & (2\pi)^{-(p+q)/2} \left(\sigma_1^2\right)^{-p/2} \left(\sigma_1^2\right)^{-q/2} \\
& \times \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_1^2}\sum_{i=1}^{p}(X_i - m)^2 + \frac{1}{\sigma_2^2}\sum_{j=1}^{q}(Y_j - m)^2\right)\right]
\end{aligned}
$$

Et le logarithme de la vraisemblance est :

$$
L \quad = \quad -\frac{(p+q)}{2}\ln(2\pi) - \frac{p}{2}\ln(\sigma_1^2) - \frac{q}{2}\ln(\sigma_2^2)
$$

$$
\begin{aligned}
&= (2\pi)^{-(p+q)/2} \left(\sigma_1^2\right)^{-p/2} \left(\sigma_1^2\right)^{-q/2} \\
&\quad \times \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_1^2}\sum_{i=1}^{p}(X_i - m)^2 + \frac{1}{\sigma_2^2}\sum_{j=1}^{q}(Y_j - m)^2\right)\right]
\end{aligned}
$$

Et le logarithme de la vraisemblance est :

$$
\begin{aligned}
L &= -\frac{(p+q)}{2}\ln(2\pi) - \frac{p}{2}\ln(\sigma_1^2) - \frac{q}{2}\ln(\sigma_2^2) \\
&\quad -\frac{1}{2\sigma_1^2}\sum_{i=1}^{p}(X_i - m)^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^{q}(Y_j - m)^2
\end{aligned}
$$

Les conditions nécessaires de premier ordre de la maximisation de L par rapport à m, σ_1^2 et σ_2^2 sont :

Les conditions nécessaires de premier ordre de la maximisation de L par rapport à m, σ_1^2 et σ_2^2 sont :

$$\frac{\partial L}{\partial m} = \frac{1}{2\hat{\sigma}_1^2}\sum_{i=1}^{p}(X_i - \hat{m}) + \frac{1}{2\hat{\sigma}_2^2}\sum_{j=1}^{q}(Y_j - \hat{m}) = 0 \qquad (1)$$

Les conditions nécessaires de premier ordre de la maximisation de L par rapport à m, σ_1^2 et σ_2^2 sont :

$$\frac{\partial L}{\partial m} = \frac{1}{2\hat{\sigma}_1^2}\sum_{i=1}^{p}(X_i - \hat{m}) + \frac{1}{2\hat{\sigma}_2^2}\sum_{j=1}^{q}(Y_j - \hat{m}) = 0 \tag{1}$$

$$\frac{\partial L}{\partial \sigma_1^2} = -\frac{p}{2\hat{\sigma}_1^2} + \frac{1}{2\hat{\sigma}_1^4}\sum_{i=1}^{p}(X_i - \hat{m})^2 = 0 \tag{2}$$

Les conditions nécessaires de premier ordre de la maximisation de L par rapport à m, σ_1^2 et σ_2^2 sont :

$$\frac{\partial L}{\partial m} = \frac{1}{2\hat{\sigma}_1^2}\sum_{i=1}^{p}(X_i - \hat{m}) + \frac{1}{2\hat{\sigma}_2^2}\sum_{j=1}^{q}(Y_j - \hat{m}) = 0 \qquad (1)$$

$$\frac{\partial L}{\partial \sigma_1^2} = -\frac{p}{2\hat{\sigma}_1^2} + \frac{1}{2\hat{\sigma}_1^4}\sum_{i=1}^{p}(X_i - \hat{m})^2 = 0 \qquad (2)$$

$$\frac{\partial L}{\partial \sigma_2^2} = -\frac{q}{2\hat{\sigma}_2^2} + \frac{1}{2\hat{\sigma}_2^4}\sum_{j=1}^{q}(Y_j - \hat{m})^2 = 0 \qquad (3)$$

Les conditions nécessaires de premier ordre de la maximisation de L par rapport à m, σ_1^2 et σ_2^2 sont :

$$\frac{\partial L}{\partial m} = \frac{1}{2\hat{\sigma}_1^2}\sum_{i=1}^{p}(X_i - \hat{m}) + \frac{1}{2\hat{\sigma}_2^2}\sum_{j=1}^{q}(Y_j - \hat{m}) = 0 \quad (1)$$

$$\frac{\partial L}{\partial \sigma_1^2} = -\frac{p}{2\hat{\sigma}_1^2} + \frac{1}{2\hat{\sigma}_1^4}\sum_{i=1}^{p}(X_i - \hat{m})^2 = 0 \quad (2)$$

$$\frac{\partial L}{\partial \sigma_2^2} = -\frac{q}{2\hat{\sigma}_2^2} + \frac{1}{2\hat{\sigma}_2^4}\sum_{j=1}^{q}(Y_j - \hat{m})^2 = 0 \quad (3)$$

Pour l'équation (1) :

$$\frac{1}{\hat{\sigma}_1^2}\sum_{i=1}^{p} X_i + -\frac{p}{\hat{\sigma}_1^2}\hat{m} + \frac{1}{\hat{\sigma}_2^2}\sum_{j=1}^{q} Y_j - \frac{q}{\hat{\sigma}_2^2}\hat{m} = 0$$

$$\frac{1}{\hat{\sigma}_1^2}\sum_{i=1}^{p} X_i - \frac{p}{\hat{\sigma}_1^2}\hat{m} + \frac{1}{\hat{\sigma}_2^2}\sum_{j=1}^{q} Y_j - \frac{q}{\hat{\sigma}_2^2}\hat{m} = 0$$

$$\frac{1}{\hat{\sigma}_1^2}\sum_{i=1}^{p} X_i - \frac{p}{\hat{\sigma}_1^2}\hat{m} + \frac{1}{\hat{\sigma}_2^2}\sum_{j=1}^{q} Y_j - \frac{q}{\hat{\sigma}_2^2}\hat{m} = 0$$

$$\Longleftrightarrow$$

$$\frac{p}{\hat{\sigma}_1^2}\bar{X} - \frac{p}{\hat{\sigma}_1^2}\hat{m} + \frac{q}{\hat{\sigma}_2^2}\bar{Y} - \frac{q}{\hat{\sigma}_2^2}\hat{m} = 0$$

Par conséquent :

$$\hat{m} = \frac{\frac{p}{\hat{\sigma}_1^2}\bar{X} + \frac{q}{\hat{\sigma}_2^2}\bar{Y}}{\frac{p}{\hat{\sigma}_1^2} + \frac{q}{\hat{\sigma}_2^2}}$$

Pour les estimateurs des variances :

$$\hat{\sigma}_1^2 = \frac{1}{p}\sum_{i=1}^{p}(X_i - \hat{m})^2 \text{ et } \hat{\sigma}_2^2 = \frac{1}{q}\sum_{j=1}^{q}(Y_j - \hat{m})^2$$

NB : $V(\bar{X}) = \frac{\sigma_1^2}{p}$ **et** $V(\bar{Y}) = \frac{\sigma_2^2}{q}$:

Conclusion : L'estimateur de m est une moyenne pondérée des estimateurs de m à partir de chaque échantillon et la pondération se fait par la précision (inverse de la variance). L'estimateur le plus précis aura plus de poids

Définition

On considère Y une variable **continue** et F une variable **qualitative** (catégorielle) appelé facteur. Un modèle d'analyse de la variance de Y par le facteur F est défini par la relation linéaire suivante :

$$Y_{ij} = \theta_j + \epsilon_{ij}, i = 1, 2, \cdots, N_j \text{ et } j = 1, 2, \cdots, g$$

où g le nombre de modalités du facteur F et N_j le nombre d'individus ayant la modalité j (appartenant au groupe j)

Les termes d'erreurs ϵ_{ij} vérifient les hypothèses suivantes :

- H_1 : $E(\epsilon_{ij} = 0 \forall i, j$
- H_2 : $Var(\epsilon_{ij}) = \sigma_\epsilon^2 \forall i, j$
- H_3 $E(\epsilon_{ij}\epsilon_{lk}) = 0 \forall i \neq l \text{ ou } j \neq k$
- H_4 $\epsilon_{ij} \rightsquigarrow N(0, \sigma_\epsilon^2)$

L'estimation par moindres carrés ordinaires :

$$\hat{\theta} = \arg \underset{\theta}{Min} \sum_{i=1}^{N_j} \sum_{j=1}^{g} \epsilon_{ij}^2 = (Y_{ij} - \theta_j)^2 \tag{4}$$

L'estimation par moindres carrés ordinaires :

$$\hat{\theta} = \arg \underset{\theta}{Min} \sum_{i=1}^{N_j} \sum_{j=1}^{g} \epsilon_{ij}^2 = (Y_{ij} - \theta_j)^2 \tag{4}$$

Les conditions nécessaires de premier ordre sont :

$$\frac{\partial \sum_{i=1}^{N_j} \sum_{j=1}^{g} \epsilon_{ij}^2}{\partial \theta_j} = 0$$

L'estimation par moindres carrés ordinaires :

$$\hat{\theta} = \arg \underset{\theta}{Min} \sum_{i=1}^{N_j} \sum_{j=1}^{g} \epsilon_{ij}^2 = (Y_{ij} - \theta_j)^2 \tag{4}$$

Les conditions nécessaires de premier ordre sont :

$$\frac{\partial \sum_{i=1}^{N_j} \sum_{j=1}^{g} \epsilon_{ij}^2}{\partial \theta_j} = 0$$

$\Longrightarrow$

$$\sum_{i=1}^{N_j} \hat{\epsilon}_{ij} = \sum_{i=1}^{N_j} (Y_{ij} - \hat{\theta}_j) = 0$$

L'estimation par moindres carrés ordinaires :

$$\hat{\theta} = \arg \underset{\theta}{Min} \sum_{i=1}^{N_j} \sum_{j=1}^{g} \epsilon_{ij}^2 = (Y_{ij} - \theta_j)^2 \tag{4}$$

Les conditions nécessaires de premier ordre sont :

$$\frac{\partial \sum_{i=1}^{N_j} \sum_{j=1}^{g} \epsilon_{ij}^2}{\partial \theta_j} = 0$$

$\Longrightarrow$

$$\sum_{i=1}^{N_j} \hat{\epsilon}_{ij} = \sum_{i=1}^{N_j} (Y_{ij} - \hat{\theta}_j) = 0 \Longrightarrow \hat{\theta}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} Y_{ij} = \bar{Y}_j \ \forall j$$

avec

$$E(\hat{\theta}_j) = \theta_j \text{ et } V(\hat{\theta}_j) = \frac{\sigma_\epsilon^2}{N_j}$$

La somme des carrés des résidus :

$$SCR = \sum_{j=1}^{g} \sum_{i=1}^{N_j} \hat{\epsilon}_{ij}^2$$

La somme des carrés des résidus :

$$\begin{aligned} SCR &= \sum_{j=1}^{g} \sum_{i=1}^{N_j} \hat{\epsilon}_{ij}^2 \\ &= \sum_{j=1}^{g} \sum_{i=1}^{N_j} (Y_{ij} - \hat{Y}_{ij})^2 \end{aligned}$$

La somme des carrés des résidus :

$$\begin{aligned} SCR &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}\hat{\epsilon}_{ij}^2 \\ &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\hat{Y}_{ij})^2 \\ &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\hat{\theta}_j)^2 \end{aligned}$$

La somme des carrés des résidus :

$$\begin{aligned} SCR &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}\hat{\epsilon}_{ij}^2 \\ &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\hat{Y}_{ij})^2 \\ &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\hat{\theta}_{j})^2 \\ &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y}_{j})^2 \end{aligned}$$

La somme des carrés des résidus :

$$\begin{aligned}
SCR &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}\hat{\epsilon}_{ij}^2 \\
&= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\hat{Y}_{ij})^2 \\
&= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\hat{\theta}_j)^2 \\
&= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y}_j)^2 \\
&= \sum_{j=1}^{g}(N_j-1)\left(\frac{1}{N_j-1}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y}_j)^2\right)
\end{aligned}$$

Sachant que pour chaque groupe (j) :

$$\frac{1}{N_j - 1}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2$$

est un estimateur sans biais de la variance (σ_ϵ^2)

$$E(SCR) = \sum_{j=1}^{g}(N_j - 1)E\left(\frac{1}{N_j - 1}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2\right)$$

Sachant que pour chaque groupe (j) :

$$\frac{1}{N_j - 1}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2$$

est un estimateur sans biais de la variance (σ_ϵ^2)

$$\begin{aligned} E(SCR) &= \sum_{j=1}^{g}(N_j - 1)E\left(\frac{1}{N_j - 1}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2\right) \\ &= \sum_{j=1}^{g}(N_j - 1)\sigma^2 \end{aligned}$$

Sachant que pour chaque groupe (j) :

$$\frac{1}{N_j - 1}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2$$

est un estimateur sans biais de la variance (σ_ϵ^2)

$$\begin{aligned} E(SCR) &= \sum_{j=1}^{g}(N_j - 1)E\left(\frac{1}{N_j - 1}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2\right) \\ &= \sum_{j=1}^{g}(N_j - 1)\sigma^2 = \left(\sum_{j=1}^{g} N_j - g\right)\sigma_\epsilon^2 = (N - g)\sigma_\epsilon^2 \end{aligned}$$

Ainsi, l'estimateur sans biais de la variance :

$$\hat{\sigma}_\epsilon^2 = \frac{1}{N - g}\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2$$

Décomposition de la variance

On a :

$$SCT = \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y})^2$$

Décomposition de la variance

On a :

$$SCT = \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y})^2 = \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y}_j+\bar{Y}_j-\bar{Y})^2$$

Décomposition de la variance

On a :

$$\begin{aligned} SCT &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y})^2 = \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y}_j+\bar{Y}_j-\bar{Y})^2 \\ &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}\left[(Y_{ij}-\bar{Y}_j)^2+(\bar{Y}_j-\bar{Y})^2-2(Y_{ij}-\bar{Y}_j)(\bar{Y}_j-\bar{Y})\right] \end{aligned}$$

Décomposition de la variance

On a :

$$\begin{aligned}
SCT &= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y})^2 = \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y}_j+\bar{Y}_j-\bar{Y})^2 \\
&= \sum_{j=1}^{g}\sum_{i=1}^{N_j}\left[(Y_{ij}-\bar{Y}_j)^2+(\bar{Y}_j-\bar{Y})^2-2(Y_{ij}-\bar{Y}_j)(\bar{Y}_j-\bar{Y})\right] \\
&= \sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y}_j)^2+\sum_{j=1}^{g}N_j(\bar{Y}_j-\bar{Y})^2 \\
&\quad+\sum_{j=1}^{g}(\bar{Y}_j-\bar{Y})\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y}_j)
\end{aligned}$$

Conclusion :

$$SCT = \underbrace{\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2}_{\text{Variance Intra ou Within}}$$

Conclusion :

$$SCT = \underbrace{\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2}_{\text{Variance Intra ou Within}} + \underbrace{\sum_{j=1}^{g} N_j(\bar{Y}_j - \bar{Y})^2}_{\text{Variance Inter ou Between}}$$

Conclusion :

$$
\begin{aligned}
SCT &= \underbrace{\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y}_j)^2}_{\text{Variance Intra ou Within}} \\
&\quad + \underbrace{\sum_{j=1}^{g} N_j(\bar{Y}_j-\bar{Y})^2}_{\text{Variance Inter ou Between}} \\
&= SCR + \underbrace{\sum_{j=1}^{g} N_j(\bar{Y}_j-\bar{Y})^2}_{\text{Variance Inter ou Between}}
\end{aligned}
$$

Tableau ANOVA et test $H_0 : \theta_j = \theta \forall j$

Le tableau d'analyse de la variance est défini comme suit :

Source	$\sum$ des carrés	ddl	Carré moyen
SCE=Between	$\sum_{j=1}^{g} N_j(\bar{Y}_j - \bar{Y})^2$	$g-1$	Between$/g-1$
SCR=Within	$\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2$	$\sum_{j=1}^{g} N_j - g$	Within$/N-g$
Total	$\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y})^2$	$\sum_{j=1}^{g} N_j - 1$	

Sous l'hypothèse de la normalité des résidus :

$$\frac{SCT}{\sigma_\epsilon^2} = \frac{1}{\sigma_\epsilon^2}\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij}-\bar{Y})^2 \rightsquigarrow \chi^2(\sum_{j=1}^{g} N_j - 1)$$

Sous l'hypothèse de la normalité des résidus :

$$\frac{SCT}{\sigma_\epsilon^2} = \frac{1}{\sigma_\epsilon^2}\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y})^2 \rightsquigarrow \chi^2(\sum_{j=1}^{g} N_j - 1)$$

$$\frac{SCR}{\sigma_\epsilon^2} = \frac{1}{\sigma_\epsilon^2}\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2 \rightsquigarrow \chi^2(\sum_{j=1}^{g} N_j - g)$$

Sous l'hypothèse de la normalité des résidus :

$$\frac{SCT}{\sigma_\epsilon^2} = \frac{1}{\sigma_\epsilon^2}\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y})^2 \rightsquigarrow \chi^2(\sum_{j=1}^{g} N_j - 1)$$

$$\frac{SCR}{\sigma_\epsilon^2} = \frac{1}{\sigma_\epsilon^2}\sum_{j=1}^{g}\sum_{i=1}^{N_j}(Y_{ij} - \bar{Y}_j)^2 \rightsquigarrow \chi^2(\sum_{j=1}^{g} N_j - g)$$

$$\frac{SCE}{\sigma_\epsilon^2} = \frac{1}{\sigma_\epsilon^2}\sum_{j=1}^{g}\sum_{i=1}^{N_j}N_j(Y_{ij} - \bar{Y}_j)^2 \rightsquigarrow \chi^2(g - 1)$$

Test H_0 **:** $\theta_j = \theta \forall j$.
Ainsi, sous H_0, le modèle devient (modèle contraint) :

$$Y_{ij} = \theta + \epsilon_{ij}, i = 1, 2, \cdots, N_j \text{ et } j = 1, 2, \cdots, g$$

$$SCR_c = \sum_{j=1}^{g} \sum_{i=1}^{N_j} (Y_{ij} - \bar{Y})^2) = SCT$$

La statistique du test est définie par :

$$F = \frac{(SCR_c - SCR_{nc}) \big/ (ddl_c - ddl_{nc})}{SCR_{nc} / ddl_{nc}}$$

Test H_0 **:** $\theta_j = \theta \forall j$.
Ainsi, sous H_0, le modèle devient (modèle contraint) :

$$Y_{ij} = \theta + \epsilon_{ij}, i = 1, 2, \cdots, N_j \text{ et } j = 1, 2, \cdots, g$$

$$SCR_c = \sum_{j=1}^{g} \sum_{i=1}^{N_j} (Y_{ij} - \bar{Y})^2) = SCT$$

La statistique du test est définie par :

$$\begin{aligned} F &= \frac{(SCR_c - SCR_{nc}) / (ddl_c - ddl_{nc})}{SCR_{nc}/ddl_{nc}} \\ &= \frac{\sum_{j=1}^{g} N_j(\bar{Y}_j - \bar{Y})^2 / (ddl_c - ddl_{nc})}{\sum_{j=1}^{g} \sum_{i=1}^{N_j} (Y_{ij} - \bar{Y}_j)^2 / ddl_{nc}} \end{aligned}$$

Test H_0 **:** $\theta_j = \theta \forall j$.
Ainsi, sous H_0, le modèle devient (modèle contraint) :

$$Y_{ij} = \theta + \epsilon_{ij}, i = 1, 2, \cdots, N_j \text{ et } j = 1, 2, \cdots, g$$

$$SCR_c = \sum_{j=1}^{g} \sum_{i=1}^{N_j} (Y_{ij} - \bar{Y})^2) = SCT$$

La statistique du test est définie par :

$$\begin{aligned} F &= \frac{(SCR_c - SCR_{nc}) / (ddl_c - ddl_{nc})}{SCR_{nc} / ddl_{nc}} \\ &= \frac{\sum_{j=1}^{g} N_j (\bar{Y}_j - \bar{Y})^2 / (ddl_c - ddl_{nc})}{\sum_{j=1}^{g} \sum_{i=1}^{N_j} (Y_{ij} - \bar{Y}_j)^2 / ddl_{nc}} \\ &= \frac{\text{Between} / g - 1}{\text{Within} / N - g} \rightsquigarrow F(g-1; N-g), N = \sum_{j=1}^{g} N_j \end{aligned}$$

MERCI POUR VOTRE ATTENTION