Décomposition de la variance

ABOUZAID Mehdi
TOOMEY Damien

Table des matières

1 Notations

c : nombre de groupes
r_i : effectif d'un groupe, $i = 1...c$
n : effectif total

Moyenne globale (tous groupes confondus) :

$$\bar{y} = \frac{1}{c} \cdot \sum_{i=1}^{c} \bar{y}_i$$

Moyenne d'un groupe :

$$\bar{y}_i = \frac{1}{r_i} \cdot \sum_{j=1}^{r_i} y_{ij}$$

2 Modèle

$$y_{ij} = \mu_i + \epsilon_i$$

Dans cette démonstration, pour rester général, on considère que les groupes n'ont pas forcément les mêmes effectifs.

3 Décomposition de la somme des carrés des écarts

$$
\begin{aligned}
\text{SCE} &= \text{somme des carrés des écarts} \\
&= \sum_{i=1}^{c}\sum_{j=1}^{r_i}(y_{ij}-\bar{y})^2 \\
&= \sum_{i=1}^{c}\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i+\bar{y}_i-\bar{y})^2 \\
&= \sum_{i=1}^{c}\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i)^2 + 2\cdot\sum_{i=1}^{c}\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i)\cdot(\bar{y}_i-\bar{y}) + \sum_{i=1}^{c}\sum_{j=1}^{r_i}(\bar{y}_i-\bar{y})^2 \quad \mathbf{(1)} \\
&= \sum_{i=1}^{c}\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i)^2 + \sum_{i=1}^{c}\sum_{j=1}^{r_i}(\bar{y}_i-\bar{y})^2 \quad \mathbf{(2)} \\
&= \underbrace{\sum_{i=1}^{c}\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i)^2}_{\text{Intra-Classe}} + \underbrace{\sum_{i=1}^{c} r_i\cdot(\bar{y}_i-\bar{y})^2}_{\text{Inter-Classes}}
\end{aligned}
$$

Passage des lignes **(1)** à **(2)** :

$$
\begin{aligned}
&\sum_{i=1}^{c}\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i)\cdot(\bar{y}_i-\bar{y}) \\
&= \sum_{i=1}^{c}\sum_{j=1}^{r_i}(y_{ij}\cdot\bar{y}_i - y_{ij}\cdot\bar{y} - \bar{y}_i^2 + \bar{y}_i\cdot\bar{y}) \\
&= \sum_{i=1}^{c}\sum_{j=1}^{r_i}(\bar{y}_i\cdot(y_{ij}-\bar{y}_i) - \bar{y}\cdot(y_{ij}-\bar{y}_i)) \\
&= \sum_{i=1}^{c}(\bar{y}_i\cdot\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i) - \bar{y}\cdot\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i))
\end{aligned}
$$

or

$$
\begin{aligned}
&\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i)\\
&=(\sum_{j=1}^{r_i} y_{ij})-\bar{y}_i\cdot(\sum_{j=1}^{r_i}1)\\
&=(\sum_{j=1}^{r_i} y_{ij})-\bar{y}_i\cdot r_i\\
&=(\sum_{j=1}^{r_i} y_{ij})-\frac{1}{r_i}\cdot\sum_{f=1}^{r_i} y_{if}\cdot r_i\\
&=(\sum_{j=1}^{r_i} y_{ij})-\sum_{f=1}^{r_i} y_{if}\\
&=0
\end{aligned}
$$

En effet, la moyenne des écarts à la moyenne est toujours nulle.

Ainsi,

$$2\cdot\sum_{i=1}^{c}\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i)\cdot(\bar{y}_i-\bar{y})=0$$

4 Décomposition de la variance

$$\Rightarrow \text{Variance}=\frac{\text{SCE}}{n}=\underbrace{\frac{1}{n}\cdot\sum_{i=1}^{c}\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_i)^2}_{\substack{\text{Variance Intra-Classe}\\ \text{Variance résiduelle}}}+\underbrace{\frac{1}{n}\cdot\sum_{i=1}^{c}r_i\cdot(\bar{y}_i-\bar{y})^2}_{\substack{\text{Variance Inter-Classes}\\ \text{Variance expliquée}}}$$

Par la suite, on considérera que l'effectif de chaque groupe est identique ($r_i=r,\ \forall i=1...c$).

5 Bibliographie

- Wikipedia
  https://fr.wikipedia.org/wiki/Analyse_de_la_variance