Analyse de variance Exercice 3 TD 3…

Sébastien Lê

Exemple extrait de Wonnacott et Wonnacott (tableau 1)

47	55	54
49	58	51
53	54	50
50	61	51
46	52	49
49	**56**	**51**

Pour une colonne fixée, la dernière ligne du tableau correspond à la moyenne des 5 valeurs de la colonne considérée.

Exemple extrait de Wonnacott et Wonnacott (tableau 2)

49	52	55
55	51	51
51	55	52
52	58	52
48	49	50
51	**53**	**52**

Pour une colonne fixée, la dernière ligne du tableau correspond à la moyenne des 5 valeurs de la colonne considérée.

Exemple extrait de Wonnacott et Wonnacott

47	55	54
49	58	51
53	54	50
50	61	51
46	52	49
49	**56**	**51**

49	52	55
55	51	51
51	55	52
52	58	52
48	49	50
51	**53**	**52**

Exemple extrait de Wonnacott et Wonnacott (tableau 1)

47	55	54
49	58	51
53	54	50
50	61	51
46	52	49
49	**56**	**51**

1. **3 échantillons de la production d'une seule machine**
2. **Echantillons de la production de 3 machines**

Exemple extrait de Wonnacott et Wonnacott (tableau 2)

49	52	55
55	51	51
51	55	52
52	58	52
48	49	50
51	**53**	**52**

1. **3 échantillons de la production d'une seule machine**
2. **Echantillons de la production de 3 machines**

Echantillons de la production de 3 machines (Wonnacott et Wonnacott)

Machine 1	Machine 2	Machine 3
47	55	54
49	58	51
53	54	50
50	61	51
46	52	49
49	**56**	**51**

Tableau 1

3 échantillons de la production d'une seule machine (Wonnacott et Wonnacott)

Echantillon 1	Echantillon 2	Echantillon 3
49	52	55
55	51	51
51	55	52
52	58	52
48	49	50
51	**53**	**52**

Tableau 2

Exemple extrait de Wonnacott et Wonnacott

Machine 1	Machine 2	Machine 3
47	55	54
49	58	51
53	54	50
50	61	51
46	52	49
49	**56**	**51**

Echantillon 1	Echantillon 2	Echantillon 3
49	52	55
55	51	51
51	55	52
52	58	52
48	49	50
51	**53**	**52**

Question

- Quand on regarde les tableaux 1 et 2, se pose la question suivante : « Les différences entres les moyennes du premier tableau sont-elles du même ordre de grandeur que celles du deuxième tableau, ou sont elles assez grandes pour indiquer une différence entre les « vraies » moyennes sous-jacentes? »

Echantillons de la production de 3 machines (Wonnacott et Wonnacott)

Machine 1	Machine 2	Machine 3
50	48	57
42	57	59
53	65	48
45	59	46
55	51	45
49	**56**	**51**

Tableau 3

Question

- Quand on regarde le tableau 3, on se dit que « les machines sont si irrégulières que tous les échantillons pourraient provenir de la même population (…). Par contre, (dans le tableau 1), les même différences entre les moyennes d'échantillon peuvent difficilement être expliqués par le hasard, parce que les machines *n'ont pas*, dans ce cas, une production aussi irrégulière.

Notations

y_{ij} Est la valeur prise par la *j*^{e} mesure pour la *i*^{e} machine

$y_{24} = 61$ Est la valeur prise par la *4*^{e} mesure pour la *2*^{e} machine dans le tableau 1

$y_{i.}$ Est la valeur de la moyenne des mesures pour la *i*^{e} machine

$y_{..}$ Est la valeur de la moyenne des mesures toutes machines confondues

n_i Est le nombre de mesures pour la *i*^{e} machine

Décomposition de la variabilité: Equation d'analyse de la variance

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-y_{..})^2=\sum_{i=1}^{I}n_i(y_{i.}-y_{..})^2+\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-y_{i.})^2$$

$SCE_{Total} = SCE_{Inter} + SCE_{Intra}$

Part de variabilité imputable au facteur Coefficient de détermination R^2

$$R^2 = \frac{SCE_{Inter}}{SCE_{Total}} = 1 - \frac{SCE_{Intra}}{SCE_{Total}}$$

Effet du facteur

H_0 : le facteur n'a pas d'effet sur Y
H_1 : le facteur a un effet sur Y

Effet du facteur

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-y_{..})^2 = \sum_{i=1}^{I} n_i(y_{i.}-y_{..})^2 + \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-y_{i.})^2$$

$SCE_{Total} = SCE_{Inter} + SCE_{Intra}$

Effet du facteur

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-y_{..})^2=\sum_{i=1}^{I}n_i(y_{i.}-y_{..})^2+\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-y_{i.})^2$$

$SCE_{Total} = SCE_{Inter} + SCE_{Intra}$

Sous l'hypothèse H_0

$$\frac{\frac{SCE_{Inter}}{I-1}}{\frac{SCE_{Intra}}{n-I}}=\frac{CM_F}{CM_R}\sim F_{n-I}^{I-1}$$

CM_R : « carré moyen de la résiduelle »

Modèle d’analyse de la variance

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

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

$$Cov(\epsilon_{ij}, \epsilon_{kl}) = 0, \forall i, j, k, l$$

Modèle d'analyse de la variance

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

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

$$Cov(\epsilon_{ij}, \epsilon_{kl}) = 0, \forall i, j, k, l$$

Modèle d'analyse de la variance

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

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

$$Cov(\epsilon_{ij}, \epsilon_{kl}) = 0, \forall i, j, k, l$$

$$\sum_{i=1}^{I} \alpha_i = 0$$

Adopter cette contrainte revient à comparer les modalités du facteur à la moyenne

Estimation des paramètres

$$y_{ij} = \hat{\mu} + \hat{\alpha}_i + e_{ij}$$

Estimation des paramètres

$$y_{ij} = \hat{\mu} + \hat{\alpha}_i + e_{ij}$$

On cherche

$$\hat{\mu} \qquad \hat{\alpha}_i$$

Qui minimisent $\sum e_{ij}^2$

Propriétés des estimateurs (dispositif équi-répété)

$$n_i = r, \forall i$$

$$V(\hat{\alpha}_i) = \sigma^2_{\hat{\alpha}_i} = \frac{I-1}{I}\frac{\sigma^2}{r}$$

Ecarts au modèle : résidus

$$\hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i$$

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

Ecarts au modèle : résidus

$$\hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i$$

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

$$\hat{\sigma}^2 = \frac{\sum_{i,j} e_{ij}^2}{n - I}$$