Tableaux d'ANOVA pour quelques plans d'expériences classiques

Jean Vaillant

Le nombre total d'unités expérimentales utilisées dans le plan est noté n . La série statistique des n observations a pour moyenne $\bar{x}$ et pour variance s^2.

I. Plan en randomisation totale - Completely randomized design

I.1. Plan à un facteur avec répétitions :

On étudie un facteur A ayant I niveaux. On a n_i répétitions pour le niveau i.

Données : x_{ij} est la valeur observée pour la $j$ème répétition du niveau i de A; $\bar{x}_{i.}$ est la moyenne observée pour le niveau i de A.

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Totale	SCT	$n-1$	CMT=$\frac{SCT}{n-1}$	
Facteur A	SCF_A	$I-1$	$CMF_A = \frac{SCF_A}{I-1}$	$F_A = \frac{CMF_A}{CMR}$
Résiduelle	SCR	$n-I$	$CMR = \frac{SCR}{n-I}$	

avec

$$SCT = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2 = ns^2;\ SCF_A = \sum_{i=1}^{I} n_i(\bar{x}_{i.}-\bar{x})^2;\ SCR = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2$$

Equation d'ANOVA : $SCT = SCF_A + SCR$

Modèle sous-jacent : $x_{ij} = \mu + \alpha_i + e_{ij}$ où les e_{ij} sont i.i.d. selon $\mathcal{N}(0, \sigma^2)$.

I.2. Plan à deux facteurs sans répétitions :

On étudie deux facteurs A et B ayant respectivement I et J niveaux.

Données : x_{ij} est la valeur observée pour le traitement (i,j); $\bar{x}_{i.}$ est la moyenne observée pour le niveau i de A; $\bar{x}_{.j}$ est la moyenne observée pour le niveau j de B.

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Totale	SCT	$IJ-1$	$CMT = \dfrac{SCT}{n-1}$	
Facteur A	SCF_A	$I-1$	$CMF_A = \dfrac{SCF_A}{I-1}$	$F_A = \dfrac{CMF_A}{CMR}$
Facteur B	SCF_B	$J-1$	$CMF_B = \dfrac{SCF_B}{J-1}$	$F_B = \dfrac{CMF_B}{CMR}$
Résiduelle	SCR	$(I-1)(J-1)$	$CMR = \dfrac{SCR}{(I-1)(J-1)}$	

avec

$$SCT = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij}-\bar{x})^2; \quad SCF_A = \sum_{i=1}^{I} J(\bar{x}_{i.}-\bar{x})^2; \quad SCF_B = \sum_{j=1}^{J} I(\bar{x}_{.j}-\bar{x})^2$$

et

$$SCR = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x})^2$$

Equation d'ANOVA : $SCT = SCF_A + SCF_B + SCR$

Modèle sous-jacent : $x_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$ où les e_{ij} sont i.i.d. selon $\mathcal{N}(0,\sigma^2)$.

I.3. Plan à deux facteurs avec répétitions :

On étudie deux facteurs A et B ayant respectivement I et J niveaux.

Données : x_{ijk} est la valeur observée pour la $k$ème répétition du traitement (i, j); $\bar{x}_{i..}$ est la moyenne observée pour le niveau i de A; $\bar{x}_{.j.}$ celle pour le niveau j de B; $\bar{x}_{ij.}$ celle pour le traitement (i, j); n_{ij} est le nombre de répétitions du traitement (i, j); n_{i+} est le nombre de répétitions du niveau i de A; n_{+j} est le nombre de répétitions du niveau j de B.

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Totale	SCT	$n-1$	$CMT = \dfrac{SCT}{n-1}$	
Facteur A	SCF_A	$I-1$	$CMF_A = \dfrac{SCF_A}{I-1}$	$F_A = \dfrac{CMF_A}{CMR}$
Facteur B	SCF_B	$J-1$	$CMF_B = \dfrac{SCF_B}{J-1}$	$F_B = \dfrac{CMF_B}{CMR}$
Interaction (A, B)	SCF_{AB}	$(I-1)(J-1)$	$CMF_{AB} = \dfrac{SCF_{AB}}{(I-1)(J-1)}$	$F_{AB} = \dfrac{CMF_{AB}}{CMR}$
Résiduelle	SCR	$n-IJ$	$CMR = \dfrac{SCR}{n-IJ}$	

avec $SCT = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(x_{ijk}-\bar{x})^2$; $SCF_A = \sum_{i=1}^{I} n_{i+}(\bar{x}_{i..}-\bar{x})^2$; $SCF_B = \sum_{j=1}^{J} n_{+j}(\bar{x}_{.j.}-\bar{x})^2$

$$SCF_{AB} = \sum_{i=1}^{I}\sum_{i=1}^{J} n_{ij}(\bar{x}_{ij.}-\bar{x}_{i..}-\bar{x}_{.j.}+\bar{x})^2 \text{ et } SCR = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(x_{ijk}-\bar{x}_{ij.})^2$$

Equation d'ANOVA : $SCT = SCF_A + SCF_B + SCF_{AB} + SCR$

Modèle sous-jacent : $x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$ où les e_{ijk} sont i.i.d. selon $\mathcal{N}(0, \sigma^2)$.

II. Plan en blocs randomisés - Randomized block design

II.1. Plan à un facteur en blocs complets sans répétition :

Données : b est le nombre de blocs, I est le nombre de niveaux du facteur A; puisqu'il n'y a pas de répétitions dans les blocs, le nombre total d'unités expérimentales est $n = Ib$; x_{ij} est la valeur observée pour le niveau i de A et le bloc j; $\bar{x}_{i.}$ est la moyenne observée pour le niveau i de A; $\bar{x}_{.j}$ est la moyenne observée pour le bloc j.

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Totale	SCT	$Ib-1$	CMT=$\dfrac{SCT}{Ib-1}$	
Facteur A	SCF_A	$I-1$	$CMF_A = \dfrac{SCF_A}{I-1}$	$F_A = \dfrac{CMF_A}{CMR}$
Facteur bloc	SCF_{bloc}	$b-1$	$CMF_{\text{bloc}} = \dfrac{SCF_{\text{bloc}}}{b-1}$	$F_{\text{bloc}} = \dfrac{CMF_{\text{bloc}}}{CMR}$
Résiduelle	SCR	$(I-1)(b-1)$	$CMR = \dfrac{SCR}{(I-1)(b-1)}$	

$$\text{avec } SCT = \sum_{i=1}^{I}\sum_{j=1}^{b}(x_{ij}-\bar{x})^2; SCF_A = \sum_{i=1}^{I} b(\bar{x}_{i.}-\bar{x})^2\,; SCF_{\text{bloc}} = \sum_{j=1}^{b} I(\bar{x}_{.j}-\bar{x})^2$$

Les sommes de carrés sont calculées comme dans le cas du plan à deux facteurs sans répétitions (voir I.2.) en considérant l'effet bloc comme second facteur.

Equation d'ANOVA : $SCT = SCF_A + SCF_{\text{bloc}} + SCR$

Modèle sous-jacent : $x_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$ où les e_{ij} sont i.i.d. selon $\mathcal{N}(0,\, \sigma^2)$.

II.2. Plan à un facteur en blocs complets équilibrés :

On étudie un facteur A ayant I niveaux. Les b blocs étant complets équilibrés, on note K le nombre de répétitions d'un niveau donné de A à l'intérieur d'un bloc.

Données : x_{ijk} est la valeur observée pour la $k^{\text{ème}}$ répétition du niveau i de A dans le bloc j; $\bar{x}_{i..}$ est la moyenne observée pour le niveau i de A; $\bar{x}_{.j.}$ est celle pour le bloc j; $\bar{x}_{ij.}$ est celle pour le niveau i dans le bloc j.

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Totale	SCT	$IbK-1$	$CMT = \dfrac{SCT}{IbK-1}$	
Facteur A	SCF_A	$I-1$	$CMF_A = \dfrac{SCF_A}{I-1}$	$F_A = \dfrac{CMF_A}{CMR}$
Bloc	SCF_{bl}	$b-1$	$CMF_{\text{bl}} = \dfrac{SCF_{\text{bl}}}{b-1}$	$F_{\text{bl}} = \dfrac{CMF_{\text{bl}}}{CMR}$
Interaction (A,Bloc)	$SCF_{A,\text{bl}}$	$(I-1)(b-1)$	$CMF_{A,\text{bl}} = \dfrac{SCF_{A,\text{bl}}}{(I-1)(b-1)}$	$F_{A,\text{bl}} = \dfrac{CMF_{A,\text{bl}}}{CMR}$
Résiduelle	SCR	$Ib(K-1)$	$CMR = \dfrac{SCR}{Ib(K-1)}$	

avec $SCT = \sum_{i=1}^{I}\sum_{j=1}^{b}\sum_{k=1}^{K}(x_{ijk}-\bar{x})^2$; $SCF_A = \sum_{i=1}^{I} bK(\bar{x}_{i..}-\bar{x})^2$; $SCF_{\text{bl}} = \sum_{j=1}^{b} IK(\bar{x}_{.j.}-\bar{x})^2$;

$SCF_{A,\text{bl}} = \sum_{i=1}^{I}\sum_{i=1}^{b} K(\bar{x}_{ij.}-\bar{x}_{i..}-\bar{x}_{.j.}+\bar{x})^2$ et $SCR = \sum_{i=1}^{I}\sum_{j=1}^{b}\sum_{k=1}^{K}(x_{ijk}-\bar{x}_{ij.})^2$

Equation d'ANOVA : $SCT = SCF_A + SCF_{\text{bl}} + SCF_{A,\text{bl}} + SCR$

Modèle sous-jacent : $x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$ où les e_{ijk} sont i.i.d. selon $\mathcal{N}(0,\, \sigma^2)$.

III. Plan avec plusieurs contrôles d'hétérogénéité - design with several controls of heterogeneity

<u>III.1. Carré latin t x t</u> : on étudie un facteur A qui a t niveaux ($t > 2$). On a deux facteurs de contrôle à t niveaux chacun. Le premier facteur de contrôle est appelé *ligne*, et le second, *colonne*.

Données : x_{ijk} est la valeur observée pour le niveau i de A, la ligne j et la colonne k du carré; $\bar{x}_{i..}$ est la moyenne observée pour le niveau i de A; $\bar{x}_{.j.}$ est la moyenne observée pour la ligne j; $\bar{x}_{..k}$ est celle pour la colonne k.

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Totale	SCT	$t^2 - 1$	$CMT = \dfrac{SCT}{t^2-1}$	
Facteur A	SCF_A	$t-1$	$CMF_A = \dfrac{SCF_A}{t-1}$	$F_A = \dfrac{CMF_A}{CMR}$
Facteur ligne	SCF_L	$t-1$	$CMF_L = \dfrac{SCF_L}{t-1}$	$F_L = \dfrac{CMF_L}{CMR}$
Facteur colonne	SCF_C	$t-1$	$CMF_C = \dfrac{SCF_C}{t-1}$	$F_C = \dfrac{CMF_C}{CMR}$
Résiduelle	SCR	$(t-1)(t-2)$	$CMR = \dfrac{SCR}{(t-1)(t-2)}$	

avec $SCT = t^2 s^2$; $SCF_A = \sum_{i=1}^{t} t(\bar{x}_{i..} - \bar{x})^2$ $SCF_L = \sum_{j=1}^{t} t(\bar{x}_{.j.} - \bar{x})^2$

$SCF_C = \sum_{k=1}^{t} t(\bar{x}_{..k} - \bar{x})^2$ et $SCR = \sum_{i=1}^{t}\sum_{j=1}^{t}\sum_{k=1}^{t}(x_{ijk} - \bar{x}_{i..} - \bar{x}_{.j.} - \bar{x}_{..k} + 2\bar{x})^2$

Equation d'ANOVA : $SCT = SCF_A + SCF_L + SCF_C + SCR$

Modèle sous-jacent : $x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + e_{ijk}$ où les e_{ijk} sont i.i.d. selon $\mathcal{N}(0,\, \sigma^2)$.

III.2. Carré gréco-latin t x t : deux facteurs A et B sont étudiés, chacun ayant t niveaux ($t > 3$). On a en outre deux facteurs de contrôle à t niveaux chacun. Le premier facteur de contrôle est appelé *ligne*, et le second *colonne*.

Données : x_{ijkl} est la valeur observée pour le niveau i de A, le niveau j de B, la ligne k et la colonne l du carré; $\bar{x}_{i...}$ est la moyenne observée pour le niveau i de A; $\bar{x}_{.j..}$ est celle pour le niveau j de B; $\bar{x}_{..k.}$ est celle pour la ligne k; $\bar{x}_{...l}$ est celle pour la colonne l.

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Totale	SCT	t^2-1	$CMT = \dfrac{SCT}{t^2-1}$	
Facteur A	SCF_A	$t-1$	$CMF_A = \dfrac{SCF_A}{t-1}$	$F_A = \dfrac{CMF_A}{CMR}$
Facteur B	SCF_B	$t-1$	$CMF_B = \dfrac{SCF_B}{t-1}$	$F_B = \dfrac{CMF_B}{CMR}$
Facteur ligne	SCF_L	$t-1$	$CMF_L = \dfrac{SCF_L}{t-1}$	$F_L = \dfrac{CMF_L}{CMR}$
Facteur colonne	SCF_C	$t-1$	$CMF_C = \dfrac{SCF_C}{t-1}$	$F_C = \dfrac{CMF_C}{CMR}$
Résiduelle	SCR	$(t-1)(t-3)$	$CMR = \dfrac{SCR}{(t-1)(t-3)}$	

$$\text{avec } SCT = t^2 s^2,\ SCF_A = \sum_{i=1}^{t} t(\bar{x}_{i...} - \bar{x})^2,\ SCF_B = \sum_{j=1}^{t} t(\bar{x}_{.j..} - \bar{x})^2,\ SCF_L = \sum_{j=1}^{t} t(\bar{x}_{..k.} - \bar{x})^2$$

$$SCF_C = \sum_{k=1}^{t} t(\bar{x}_{...l} - \bar{x})^2,\ SCR = \sum_{i=1}^{t}\sum_{j=1}^{t}\sum_{k=1}^{t}\sum_{l=1}^{t}(x_{ijkl} - \bar{x}_{i...} - \bar{x}_{.j..} - \bar{x}_{..k.} - \bar{x}_{...l} + 3\bar{x})^2$$

Equation d'ANOVA : $SCT = SCF_A + SCF_B + SCF_L + SCF_C + SCR$

Modèle sous-jacent : $x_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + e_{ijkl}$ où les e_{ijkl} sont i.i.d. selon $\mathcal{N}(0, \sigma^2)$.

IV. Plan en parcelles divisées, plan en bandes croisées - Split-plot design, criss-cross design

Les ANOVAs associées à ces plans ne sont pas au programme.

IV.1. Split-plot

Ce plan est utilisé quand les plus petites unités auxquelles on peut affecter un niveau de facteur sont de taille bien plus élevée que la taille requise pour l'application des niveaux des autres facteurs. Ainsi, en expérimentation agronomique, le facteur pratique culturale ne peut être appliqué qu'à une échelle suffisamment grande par rapport au facteur variété ou au facteur espacement des rangées de culture.

Un autre exemple concerne l'étude de la production laitière par rapport aux facteurs pâturage et méthode de traite. L'unité expérimentale idéale pour le facteur méthode de traite est la vache mais pour le facteur pâturage, il ne peut s'agir que d'un groupe d'animaux.

Le split-plot est un plan en blocs équilibrés qui consiste à subdiviser les blocs en sous-blocs et à affecter à toutes les unités expérimentales d'un même sous-bloc le même niveau de l'un des facteurs étudiés. Dans le cas d'un split-plot à deux facteurs étudiés A et B, on procède de la façon suivante : dans chaque bloc, les niveaux du facteur A sont attribués par tirages aléatoires aux sous-blocs (randomisation intra-bloc), puis dans chaque sous-bloc, l'attribution des niveaux du facteur B se fait par tirages aléatoires (randomisation intra-sous-bloc). B est appelé facteur split-plot.

Exemple : On dispose de 48 parcelles expérimentales en 2 blocs ayant chacun 4 sous-blocs. Chaque sous-bloc contient 6 parcelles. On attribue aléatoirement un des 4 niveaux c_1, c_2, c_3 et c_4 du facteur *Procédé cultural* à chaque sous-bloc d'un bloc. A l'intérieur d'un sous-bloc, on attribue aléatoirement un des 6 niveaux du facteur *Variété* à chaque parcelle.

Bloc 1

c_3v_2	c_3v_1	c_1v_5	c_1v_6	c_4v_1	c_4v_4	c_2v_3	c_2v_4
c_3v_5	c_3v_3	c_1v_2	c_1v_1	c_4v_5	c_4v_6	c_2v_1	c_2v_5
c_3v_4	c_3v_6	c_1v_3	c_1v_4	c_4v_2	c_4v_3	c_2v_6	c_2v_2

Bloc 2

c_2v_3	c_2v_5	c_3v_6	c_3v_3	c_1v_6	c_1v_4	c_4v_4	c_4v_5
c_2v_2	c_2v_4	c_3v_4	c_3v_1	c_1v_3	c_1v_2	c_4v_3	c_4v_2
c_2v_6	c_2v_1	c_3v_2	c_3v_5	c_1v_1	c_1v_5	c_4v_1	c_4v_6

IV.2. Criss-cross

Le plan criss-cross est un plan utilisé essentiellement en expérimentation agronomique et appliqué à cause de certaines contraintes techniques. C'est un plan à deux facteurs étudiés qui sont appliqués en bandes perpendiculaires.

Exemple : On désire étudier l'influence sur le rendement du facteur *Technique de travail du sol* (à 3 niveaux a_1, a_2 et a_3) et du facteur *Irrigation* (à deux niveaux b_1=irrigué; b_2=non irrigué). Le seul plan que l'on puisse pratiquement mettre en place est :

a_1	a_2	a_3
$\downarrow$	$\downarrow$	$\downarrow$

$b_1 \rightarrow$	a_1b_1	a_2b_1	a_3b_1
$b_2 \rightarrow$	a_1b_2	a_2b_2	a_3b_2

Le criss-cross est un plan relativement peu utilisé : il faut vraiment que, pratiquement, l'expérimentateur ne puisse disposer ses parcelles d'une autre façon.

V. Test d'hypothèse concernant l'influence de facteur

A partir du tableau d'ANOVA, on peut tester l'influence des différents facteurs étudiés ou de leurs interactions éventuelles. En effet, chaque statistique F de Fisher, obtenue dans la dernière colonne de ce tableau, correspond à une source de variation précisée dans la ligne concernée en première colonne. L'hypothèse nulle H_0 que l'on veut tester est alors

H_0 : "la source de variation est sans effet" contre l'hypothèse alternative

H_1 : "Il y a un effet dû à la source de variation"

L'hypothèse H_0 peut être formulée également de la façon suivante :

"les valeurs espérées (moyennes théoriques) sont identiques pour la source de variation".

Règle de décison au niveau α :

Rejet de H_0 si $F > f_{\nu_1;\nu_2;1-\alpha}$

Non rejet de H_0 si $F \leq f_{\nu_1;\nu_2;1-\alpha}$

où $f_{\nu_1;\nu_2;1-\alpha}$ est un fractile de la loi de Fisher-Snedecor à ν_1 et ν_2 degrés de liberté,

ν_1 est le nombre de degrés de liberté associé à la source de variation testée,

ν_2 est le nombre de degrés de liberté associé à la variation résiduelle.

Conclusion du test :

Le rejet de l'hypothèse nulle H_0 signifie que l'expérience met en évidence un effet significatif de la source de variation étudiée sur la variable principale : au moins une moyenne théorique est différente des autres.

Le non rejet de l'hypothèse nulle H_0 signifie que l'expérience n'a pu mettre en évidence une influence significative de la source de variation étudiée sur la variable principale : il n'y a pas d'écart significatif entre les différentes moyennes théoriques.

Utilisation des probabilités critiques (p-values)

Les logiciels modernes fournissent les résultats des règles de décision sous forme de probabilités critiques. Par définition, la probabilité critique p_c (en anglais p-value) du test de Fisher est la probabilité qu'une réalisation de la loi de Fisher-Snedecor soit plus élevée que la valeur de la statistique de Fisher F sous l'hypothèse H_0. Ainsi, une très faible valeur pour p_c indique que l'hypothèse privilégiée H_0 est vraisemblablement fausse. Plus p_c est faible, plus les données témoignent de l'effet significatif de la source de variation étudiée. Elles nous conduisent alors à rejeter H_0.

Remarque :

Avec MS EXCEL ou OpenOffice.org Calc, la fonction SOMME.CARRES.ECARTS permet de calculer la somme des carrés d'écart pour une série de données (utiliser le bouton fx puis chercher dans la catégorie "Statistique").