

Plan d'expérience

Pierre Druilhet

Université Blaise Pascal, Clermont Ferrand

INRA Theix 2008

1 POURQUOI FAIRE UN PLAN D'EXPÉRIENCE ?

- Exemple introductif : le plan de pesées
- Démarche expérimentale

2 PLANS EN BLOCS

- Exemple 1 : choix du nombre de blocs
- Blocage spatial
- Plans en blocs complets équilibrés (PBCE)
- Plans orthogonaux
- Exemple de PBCE
- PBIE
- Conclusion

3 NOTIONS D'OPTIMALITÉ

- Matrice d'information associée à un paramètre d'intérêt
- Optimalité uniforme
- Critères d'optimalité
- Optimalité universelle

4 RANDOMISATION

- Introduction
- Randomisation totale
- Randomisation restreinte
- Notion générale de randomisation

5 PLANS FACTORIELS

- Plan factoriel complet
- Exemple des canards
- Exemple des mesures répétées : carrés latins complets
- Carré gréco-latin
- Carrés de Youden

6 PLANS EN CROSS-OVER

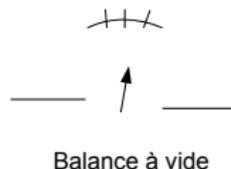
- Définition
- Plans circulaires
- Exemples
- Plan en cross-over AB/BA
- Cross-over avec pré-périodes

Première partie I

POURQUOI FAIRE UN PLAN D'EXPÉRIENCE ?

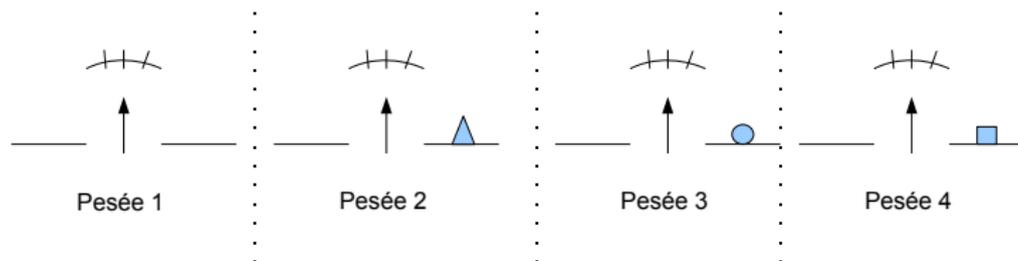
EXEMPLE INTRODUCTIF : LE PLAN DE PESÉES

- On a trois objets à peser à l'aide d'une balance de Roberval.
- On n'a droit qu'à 4 pesées. Au départ, la balance est déséquilibrée (défaut de calibrage)



- La précision de la balance est donnée par la variance des **erreurs de mesures** σ^2 .
- Objectif : estimer avec le plus de précision possible le poids des trois objets.

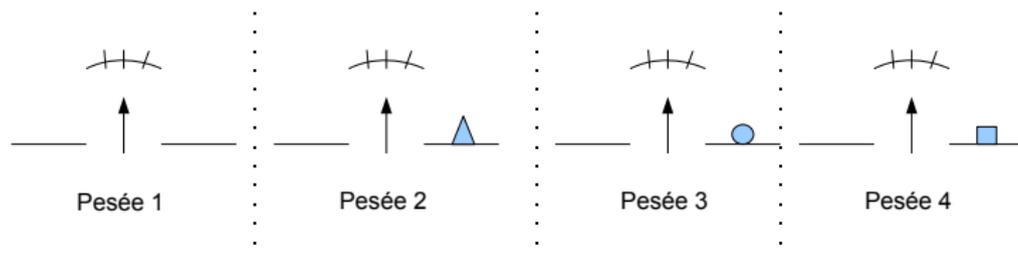
PREMIER PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{cases} y_1 = p_o + \varepsilon_1 \\ y_2 = p_o + p_1 + \varepsilon_2 \\ y_3 = p_o + p_2 + \varepsilon_3 \\ y_4 = p_o + p_3 + \varepsilon_4 \end{cases}$$

PREMIER PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{cases} y_1 = p_o + \varepsilon_1 \\ y_2 = p_o + p_1 + \varepsilon_2 \\ y_3 = p_o + p_2 + \varepsilon_3 \\ y_4 = p_o + p_3 + \varepsilon_4 \end{cases}$$

MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

ESTIMATION PAR MCO

$$\begin{aligned} \hat{p}_1 &= y_2 - y_1 && \rightarrow \text{var}(\hat{p}_1) = 2\sigma^2 \\ \hat{p}_2 &= y_3 - y_1 && \rightarrow \text{var}(\hat{p}_2) = 2\sigma^2 \\ \hat{p}_3 &= y_4 - y_1 && \rightarrow \text{var}(\hat{p}_3) = 2\sigma^2 \end{aligned}$$

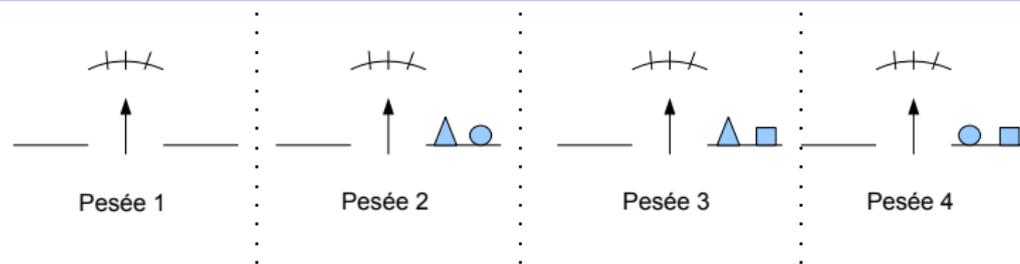
MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

ESTIMATION PAR MCO

$$\begin{aligned} \hat{p}_1 &= y_2 - y_1 && \rightarrow \text{var}(\hat{p}_1) = 2\sigma^2 \\ \hat{p}_2 &= y_3 - y_1 && \rightarrow \text{var}(\hat{p}_2) = 2\sigma^2 \\ \hat{p}_3 &= y_4 - y_1 && \rightarrow \text{var}(\hat{p}_3) = 2\sigma^2 \end{aligned}$$

DEUXIÈME PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

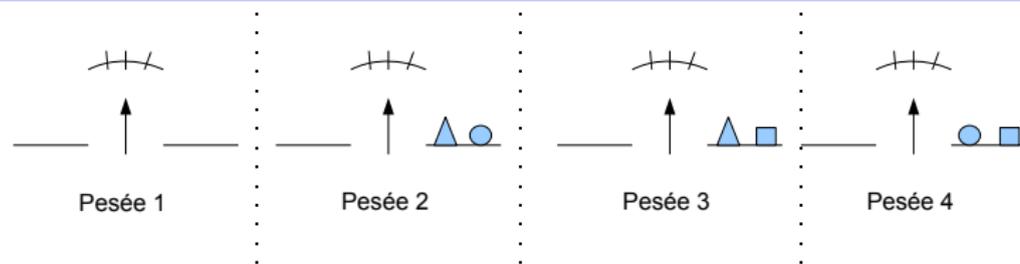
ESTIMATION PAR MCO

$$\hat{p}_1 = \frac{1}{2}(y_2 + y_3 - y_1 - y_4) \rightarrow \text{var}(\hat{p}_1) = \sigma^2$$

$$\hat{p}_2 = \frac{1}{2}(y_2 + y_4 - y_1 - y_3) \rightarrow \text{var}(\hat{p}_2) = \sigma^2$$

$$\hat{p}_3 = \frac{1}{2}(y_3 + y_4 - y_1 - y_2) \rightarrow \text{var}(\hat{p}_3) = \sigma^2$$

DEUXIÈME PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

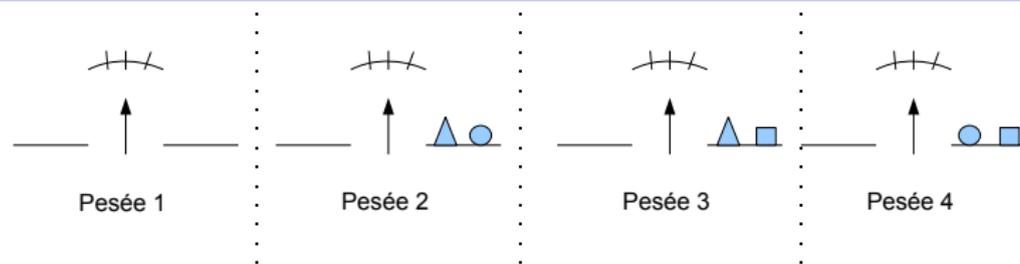
ESTIMATION PAR MCO

$$\hat{p}_1 = \frac{1}{2}(y_2 + y_3 - y_1 - y_4) \rightarrow \text{var}(\hat{p}_1) = \sigma^2$$

$$\hat{p}_2 = \frac{1}{2}(y_2 + y_4 - y_1 - y_3) \rightarrow \text{var}(\hat{p}_2) = \sigma^2$$

$$\hat{p}_3 = \frac{1}{2}(y_3 + y_4 - y_1 - y_2) \rightarrow \text{var}(\hat{p}_3) = \sigma^2$$

DEUXIÈME PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

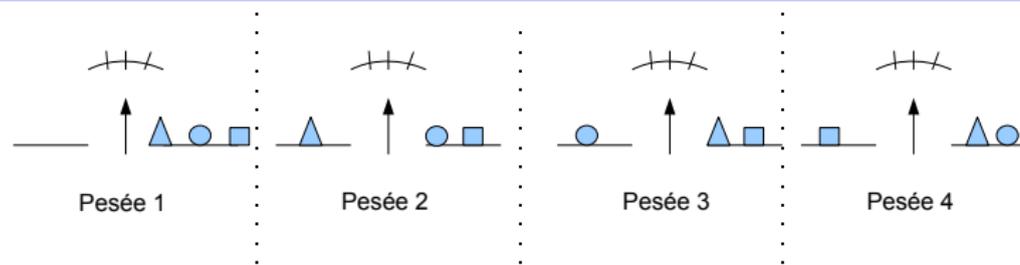
ESTIMATION PAR MCO

$$\hat{p}_1 = \frac{1}{2}(y_2 + y_3 - y_1 - y_4) \rightarrow \text{var}(\hat{p}_1) = \sigma^2$$

$$\hat{p}_2 = \frac{1}{2}(y_2 + y_4 - y_1 - y_3) \rightarrow \text{var}(\hat{p}_2) = \sigma^2$$

$$\hat{p}_3 = \frac{1}{2}(y_3 + y_4 - y_1 - y_2) \rightarrow \text{var}(\hat{p}_3) = \sigma^2$$

TROISIÈME PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

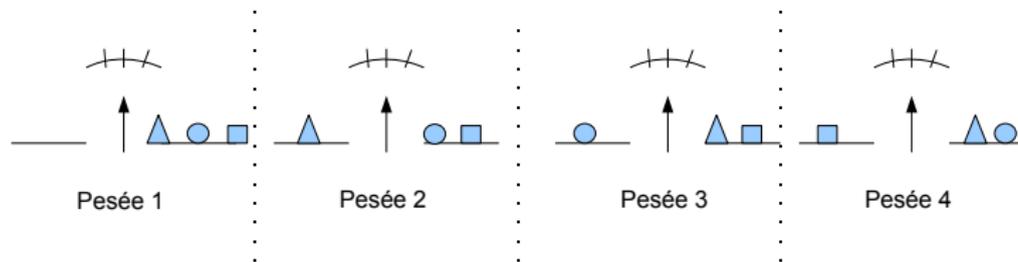
ESTIMATION PAR MCO

$$\hat{p}_1 = \frac{1}{2}(y_1 - y_2) \rightarrow \text{var}(\hat{p}_1) = \frac{1}{2}\sigma^2$$

$$\hat{p}_2 = \frac{1}{2}(y_1 - y_3) \rightarrow \text{var}(\hat{p}_2) = \frac{1}{2}\sigma^2$$

$$\hat{p}_3 = \frac{1}{2}(y_1 - y_4) \rightarrow \text{var}(\hat{p}_3) = \frac{1}{2}\sigma^2$$

TROISIÈME PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

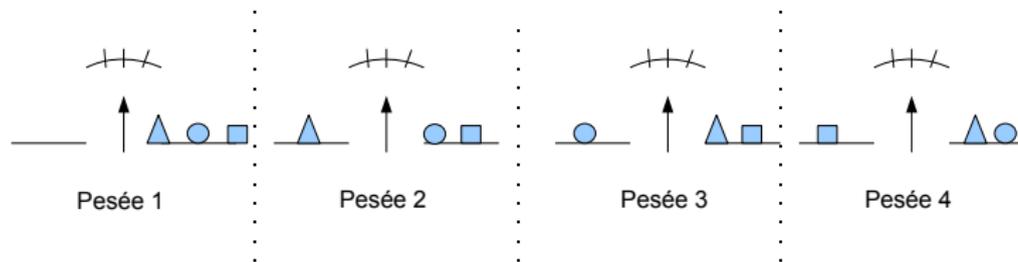
ESTIMATION PAR MCO

$$\hat{p}_1 = \frac{1}{2}(y_1 - y_2) \rightarrow \text{var}(\hat{p}_1) = \frac{1}{2}\sigma^2$$

$$\hat{p}_2 = \frac{1}{2}(y_1 - y_3) \rightarrow \text{var}(\hat{p}_2) = \frac{1}{2}\sigma^2$$

$$\hat{p}_3 = \frac{1}{2}(y_1 - y_4) \rightarrow \text{var}(\hat{p}_3) = \frac{1}{2}\sigma^2$$

TROISIÈME PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

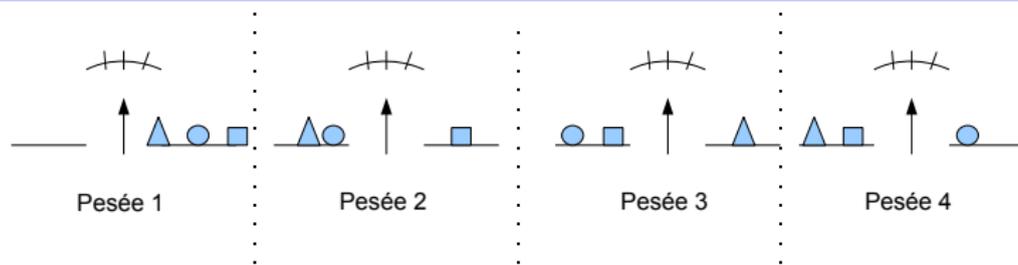
ESTIMATION PAR MCO

$$\hat{p}_1 = \frac{1}{2}(y_1 - y_2) \rightarrow \text{var}(\hat{p}_1) = \frac{1}{2}\sigma^2$$

$$\hat{p}_2 = \frac{1}{2}(y_1 - y_3) \rightarrow \text{var}(\hat{p}_2) = \frac{1}{2}\sigma^2$$

$$\hat{p}_3 = \frac{1}{2}(y_1 - y_4) \rightarrow \text{var}(\hat{p}_3) = \frac{1}{2}\sigma^2$$

QUATRIÈME PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

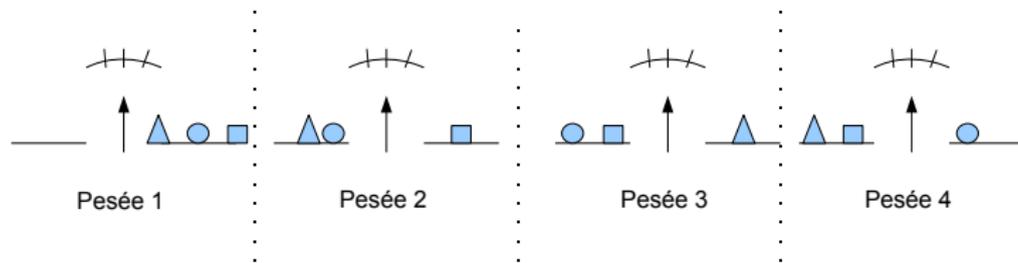
ESTIMATION PAR MCO

$$\hat{p}_1 = \frac{1}{4}(y_1 - y_2 + y_3 - y_4) \rightarrow \text{var}(\hat{p}_1) = \frac{1}{4}\sigma^2$$

$$\hat{p}_2 = \frac{1}{4}(y_1 - y_2 - y_3 + y_4) \rightarrow \text{var}(\hat{p}_2) = \frac{1}{4}\sigma^2$$

$$\hat{p}_3 = \frac{1}{4}(y_1 + y_2 - y_3 - y_4) \rightarrow \text{var}(\hat{p}_3) = \frac{1}{4}\sigma^2$$

QUATRIÈME PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

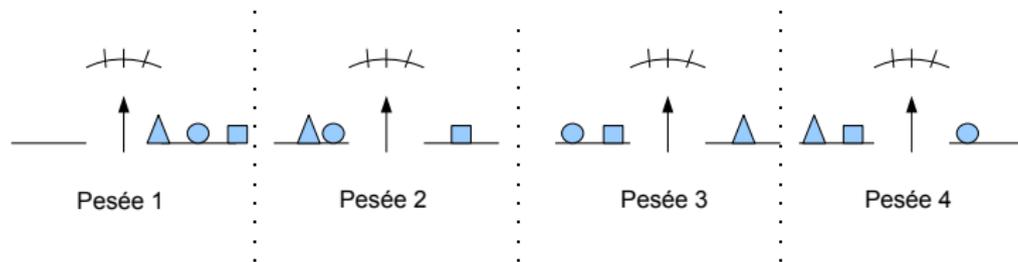
ESTIMATION PAR MCO

$$\hat{p}_1 = \frac{1}{4}(y_1 - y_2 + y_3 - y_4) \rightarrow \text{var}(\hat{p}_1) = \frac{1}{4}\sigma^2$$

$$\hat{p}_2 = \frac{1}{4}(y_1 - y_2 - y_3 + y_4) \rightarrow \text{var}(\hat{p}_2) = \frac{1}{4}\sigma^2$$

$$\hat{p}_3 = \frac{1}{4}(y_1 + y_2 - y_3 - y_4) \rightarrow \text{var}(\hat{p}_3) = \frac{1}{4}\sigma^2$$

QUATRIÈME PLAN DE PESÉES



MODÈLE STATISTIQUE ASSOCIÉ

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = Xp + \varepsilon$$

ESTIMATION PAR MCO

$$\hat{p}_1 = \frac{1}{4}(y_1 - y_2 + y_3 - y_4) \rightarrow \text{var}(\hat{p}_1) = \frac{1}{4}\sigma^2$$

$$\hat{p}_2 = \frac{1}{4}(y_1 - y_2 - y_3 + y_4) \rightarrow \text{var}(\hat{p}_2) = \frac{1}{4}\sigma^2$$

$$\hat{p}_3 = \frac{1}{4}(y_1 + y_2 - y_3 - y_4) \rightarrow \text{var}(\hat{p}_3) = \frac{1}{4}\sigma^2$$

CONCLUSION DE L'EXEMPLE

Que retenir de l'exemple :

- La stratégie de planification va influencer directement la précision statistique des résultats.
- On peut montrer que le dernière plan de pesées est optimal : la matrice du plan est une matrice d'Hadamard.
- Dans le dernier plan de pesées, le plan est orthogonal. Tout se passe comme si l'on avait pesé 4 fois chaque objet avec une balance équilibré.

PRINCIPE D'UNE DÉMARCHE EXPÉRIMENTALE

1 Définition des objectifs de l'expérience par l'expérimentateur :

- estimation : quelle va être l'espérance de vie moyen d'un individu malade sous traitement ?
- test d'hypothèse : Un nouveau médicament est-il plus efficace que le médicament de référence ?
- prévision : dans un élevage laitier, quelle quantité de lait moyen par vache prévu pour un type d'alimentation donnée : évaluation du coût et de la rentabilité ?
- ...

2 Quelle sont les contraintes expérimentales :

- contraintes de temps, j'ai à ma disposition le matériel pour 1 semaine.
- contrainte d'espace : je veux tester 5 variétés de blé et dispose de 7 parcelles de terrains.
- contraintes de coût.
- ...

PRINCIPE D'UNE DÉMARCHE EXPÉRIMENTALE

- 1 Définition des objectifs de l'expérience par l'expérimentateur :
 - estimation : quelle va être l'espérance de vie moyen d'un individu malade sous traitement ?
 - test d'hypothèse : Un nouveau médicament est-il plus efficace que le médicament de référence ?
 - prévision : dans un élevage laitier, quelle quantité de lait moyen par vache prévu pour un type d'alimentation donnée : évaluation du coût et de la rentabilité ?
 - ...
- 2 Quelle sont les contraintes expérimentales :
 - contraintes de temps, j'ai à ma disposition le matériel pour 1 semaine.
 - contrainte d'espace : je veux tester 5 variétés de blé et dispose de 7 parcelles de terrains.
 - contraintes de coût.
 - ...

PRINCIPE D'UNE DÉMARCHE EXPÉRIMENTALE

- ① Définition des objectifs de l'expérience par l'expérimentateur :
 - estimation : quelle va être l'espérance de vie moyen d'un individu malade sous traitement ?
 - test d'hypothèse : Un nouveau médicament est-il plus efficace que le médicament de référence ?
 - prévision : dans un élevage laitier, quelle quantité de lait moyen par vache prévu pour un type d'alimentation donnée : évaluation du coût et de la rentabilité ?
 - ...
- ② Quelle sont les contraintes expérimentales :
 - contraintes de temps, j'ai à ma disposition le matériel pour 1 semaine.
 - contrainte d'espace : je veux tester 5 variétés de blé et dispose de 7 parcelles de terrains.
 - contraintes de coût.
 - ...

PRINCIPE D'UNE DÉMARCHE EXPÉRIMENTALE

- ① Définition des objectifs de l'expérience par l'expérimentateur :
 - estimation : quelle va être l'espérance de vie moyen d'un individu malade sous traitement ?
 - test d'hypothèse : Un nouveau médicament est-il plus efficace que le médicament de référence ?
 - prévision : dans un élevage laitier, quelle quantité de lait moyen par vache prévu pour un type d'alimentation donnée : évaluation du coût et de la rentabilité ?
 - ...
- ② Quelle sont les contraintes expérimentales :
 - contraintes de temps, j'ai à ma disposition le matériel pour 1 semaine.
 - contrainte d'espace : je veux tester 5 variétés de blé et dispose de 7 parcelles de terrains.
 - contraintes de coût.
 - ...

PRINCIPE D'UNE DÉMARCHE EXPÉRIMENTALE

- ① Définition des objectifs de l'expérience par l'expérimentateur :
 - estimation : quelle va être l'espérance de vie moyen d'un individu malade sous traitement ?
 - test d'hypothèse : Un nouveau médicament est-il plus efficace que le médicament de référence ?
 - prévision : dans un élevage laitier, quelle quantité de lait moyen par vache prévu pour un type d'alimentation donnée : évaluation du coût et de la rentabilité ?
 - ...
- ② Quelle sont les contraintes expérimentales :
 - contraintes de temps, j'ai à ma disposition le matériel pour 1 semaine.
 - contrainte d'espace : je veux tester 5 variétés de blé et dispose de 7 parcelles de terrains.
 - contraintes de coût.
 - ...

PRINCIPE D'UNE DÉMARCHE EXPÉRIMENTALE

- ① Définition des objectifs de l'expérience par l'expérimentateur :
 - estimation : quelle va être l'espérance de vie moyen d'un individu malade sous traitement ?
 - test d'hypothèse : Un nouveau médicament est-il plus efficace que le médicament de référence ?
 - prévision : dans un élevage laitier, quelle quantité de lait moyen par vache prévu pour un type d'alimentation donnée : évaluation du coût et de la rentabilité ?
 - ...
- ② Quelle sont les contraintes expérimentales :
 - contraintes de temps, j'ai à ma disposition le matériel pour 1 semaine.
 - contrainte d'espace : je veux tester 5 variétés de blé et dispose de 7 parcelles de terrains.
 - contraintes de coût.
 - ...

PRINCIPE D'UNE DÉMARCHE EXPÉRIMENTALE

- ① Définition des objectifs de l'expérience par l'expérimentateur :
 - estimation : quelle va être l'espérance de vie moyen d'un individu malade sous traitement ?
 - test d'hypothèse : Un nouveau médicament est-il plus efficace que le médicament de référence ?
 - prévision : dans un élevage laitier, quelle quantité de lait moyen par vache prévu pour un type d'alimentation donnée : évaluation du coût et de la rentabilité ?
 - ...
- ② Quelle sont les contraintes expérimentales :
 - contraintes de temps, j'ai à ma disposition le matériel pour 1 semaine.
 - contrainte d'espace : je veux tester 5 variétés de blé et dispose de 7 parcelles de terrains.
 - contraintes de coût.
 - ...

PRINCIPE D'UNE DÉMARCHE EXPÉRIMENTALE

- ① Définition des objectifs de l'expérience par l'expérimentateur :
 - estimation : quelle va être l'espérance de vie moyen d'un individu malade sous traitement ?
 - test d'hypothèse : Un nouveau médicament est-il plus efficace que le médicament de référence ?
 - prévision : dans un élevage laitier, quelle quantité de lait moyen par vache prévu pour un type d'alimentation donnée : évaluation du coût et de la rentabilité ?
 - ...
- ② Quelle sont les contraintes expérimentales :
 - contraintes de temps, j'ai à ma disposition le matériel pour 1 semaine.
 - contrainte d'espace : je veux tester 5 variétés de blé et dispose de 7 parcelles de terrains.
 - contraintes de coût.
 - ...

PRINCIPE D'UNE DÉMARCHE EXPÉRIMENTALE

- ① Définition des objectifs de l'expérience par l'expérimentateur :
 - estimation : quelle va être l'espérance de vie moyen d'un individu malade sous traitement ?
 - test d'hypothèse : Un nouveau médicament est-il plus efficace que le médicament de référence ?
 - prévision : dans un élevage laitier, quelle quantité de lait moyen par vache prévu pour un type d'alimentation donnée : évaluation du coût et de la rentabilité ?
 - ...
- ② Quelle sont les contraintes expérimentales :
 - contraintes de temps, j'ai à ma disposition le matériel pour 1 semaine.
 - contrainte d'espace : je veux tester 5 variétés de blé et dispose de 7 parcelles de terrains.
 - contraintes de coût.
 - ...

PRINCIPE D'UNE DÉMARCHE EXPÉRIMENTALE

- ① Définition des objectifs de l'expérience par l'expérimentateur :
 - estimation : quelle va être l'espérance de vie moyen d'un individu malade sous traitement ?
 - test d'hypothèse : Un nouveau médicament est-il plus efficace que le médicament de référence ?
 - prévision : dans un élevage laitier, quelle quantité de lait moyen par vache prévu pour un type d'alimentation donnée : évaluation du coût et de la rentabilité ?
 - ...
- ② Quelle sont les contraintes expérimentales :
 - contraintes de temps, j'ai à ma disposition le matériel pour 1 semaine.
 - contrainte d'espace : je veux tester 5 variétés de blé et dispose de 7 parcelles de terrains.
 - contraintes de coût.
 - ...

3 Définition des facteurs pouvant influencer l'expérience :

- température pour une expérience biologique.
- âge, état de santé, sexe d'un patient
- effets périodes pour des mesures répétées dans le temps.
- ...

4 Détermination des sources de variabilité possibles :

- hétérogénéité du terrain,
- différents fournisseurs d'animaux,
- expérience sur plusieurs terrains exposés différemment,
- effet du technicien,
- effet du centre d'expérimentation (essais multicentrique).
- effet de l'étude (méta-analyse)
- ...

5 Définition des unités expérimentales :

- individu,
- individu x période,
- parcelle de terrain,
- ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 3 Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- 4 Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- 5 Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- ③ Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- ④ Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- ⑤ Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- ③ Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- ④ Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- ⑤ Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- ③ Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- ④ Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- ⑤ Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- ③ Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- ④ Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- ⑤ Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- ③ Définition des facteurs pouvant influencer l'expérience :
 - température pour une expérience biologique.
 - âge, état de santé, sexe d'un patient
 - effets périodes pour des mesures répétées dans le temps.
 - ...
- ④ Détermination des sources de variabilité possibles :
 - hétérogénéité du terrain,
 - différents fournisseurs d'animaux,
 - expérience sur plusieurs terrains exposés différemment,
 - effet du technicien,
 - effet du centre d'expérimentation (essais multicentrique).
 - effet de l'étude (méta-analyse)
 - ...
- ⑤ Définition des unités expérimentales :
 - individu,
 - individu x période,
 - parcelle de terrain,
 - ...

- 6 Élaboration de l'expérience = **planification** : à partir des données précédentes, élaborer un plan qui
 - permette la meilleure précision statistique des résultats,
 - garantisse l'impartialité de l'étude (randomisation + simple ou double aveugle si possible).
- 7 Réalisation des expériences,
- 8 Analyse des résultats.

- 6 Élaboration de l'expérience = **planification** : à partir des données précédentes, élaborer un plan qui
 - permette la meilleure précision statistique des résultats,
 - garantisse l'impartialité de l'étude (randomisation + simple ou double aveugle si possible).
- 7 Réalisation des expériences,
- 8 Analyse des résultats.

- ⑥ Élaboration de l'expérience = **planification** : à partir des données précédentes, élaborer un plan qui
 - permette la meilleure précision statistique des résultats,
 - garantisse l'impartialité de l'étude (randomisation + simple ou double aveugle si possible).
- ⑦ Réalisation des expériences,
- ⑧ Analyse des résultats.

- 6 Élaboration de l'expérience = **planification** : à partir des données précédentes, élaborer un plan qui
 - permette la meilleure précision statistique des résultats,
 - garantisse l'impartialité de l'étude (randomisation + simple ou double aveugle si possible).
- 7 Réalisation des expériences,
- 8 Analyse des résultats.

- 6 Élaboration de l'expérience = **planification** : à partir des données précédentes, élaborer un plan qui
 - permette la meilleure précision statistique des résultats,
 - garantisse l'impartialité de l'étude (randomisation + simple ou double aveugle si possible).
- 7 Réalisation des expériences,
- 8 Analyse des résultats.

COMMENT GAGNER DE LA PRÉCISION STATISTIQUE ?

- Augmenter le nombre de répétitions :
 - la variance des estimateurs diminue proportionnellement au nombre d'observations si les répétitions sont réalisées dans les mêmes conditions,
 - mais la variance résiduelle peut augmenter si on est obligé de prendre une population plus hétérogène.
Exemple : tests sur des rats élevés dans un même laboratoire vs rats obtenus par différents fournisseurs. Ou bien nécessité de mener l'expérience sur plusieurs champs, plusieurs centres.
- Contrôle local (ou blocking). On regroupe les unités expérimentales en fonctions de critères ou facteurs : on essaie d'avoir la population la plus homogène possible à l'intérieur de chaque groupe.
 - diminue la variance et augmente la puissance des tests en enlevant des sources de variabilité.
 - peut augmenter la variance si le plan n'est pas orthogonal et l'homogénéisation insuffisante.
 - accroît la complexité de l'analyse.

COMMENT GAGNER DE LA PRÉCISION STATISTIQUE ?

- Augmenter le nombre de répétitions :
 - la variance des estimateurs diminue proportionnellement au nombre d'observations si les répétitions sont réalisées dans les mêmes conditions,
 - mais la variance résiduelle peut augmenter si on est obligé de prendre une population plus hétérogène.
Exemple : tests sur des rats élevés dans un même laboratoire vs rats obtenus par différents fournisseurs. Ou bien nécessité de mener l'expérience sur plusieurs champs, plusieurs centres.
- Contrôle local (ou blocking). On regroupe les unités expérimentales en fonctions de critères ou facteurs : on essaie d'avoir la population la plus homogène possible à l'intérieur de chaque groupe.
 - diminue la variance et augmente la puissance des tests en enlevant des sources de variabilité.
 - peut augmenter la variance si le plan n'est pas orthogonal et l'homogénéisation insuffisante.
 - accroît la complexité de l'analyse.

COMMENT GAGNER DE LA PRÉCISION STATISTIQUE ?

- Augmenter le nombre de répétitions :
 - la variance des estimateurs diminue proportionnellement au nombre d'observations si les répétitions sont réalisées dans les mêmes conditions,
 - mais la variance résiduelle peut augmenter si on est obligé de prendre une population plus hétérogène.
Exemple : tests sur des rats élevés dans un même laboratoire vs rats obtenus par différents fournisseurs. Ou bien nécessité de mener l'expérience sur plusieurs champs, plusieurs centres.
- Contrôle local (ou blocking). On regroupe les unités expérimentales en fonctions de critères ou facteurs : on essaie d'avoir la population la plus homogène possible à l'intérieur de chaque groupe.
 - diminue la variance et augmente la puissance des tests en enlevant des sources de variabilité.
 - peut augmenter la variance si le plan n'est pas orthogonal et l'homogénéisation insuffisante.
 - accroît la complexité de l'analyse.

COMMENT GAGNER DE LA PRÉCISION STATISTIQUE ?

- Augmenter le nombre de répétitions :
 - la variance des estimateurs diminue proportionnellement au nombre d'observations si les répétitions sont réalisées dans les mêmes conditions,
 - mais la variance résiduelle peut augmenter si on est obligé de prendre une population plus hétérogène.
Exemple : tests sur des rats élevés dans un même laboratoire vs rats obtenus par différents fournisseurs. Ou bien nécessité de mener l'expérience sur plusieurs champs, plusieurs centres.
- Contrôle local (ou blocking). On regroupe les unités expérimentales en fonctions de critères ou facteurs : on essaie d'avoir la population la plus homogène possible à l'intérieur de chaque groupe.
 - diminue la variance et augmente la puissance des tests en enlevant des sources de variabilité.
 - peut augmenter la variance si le plan n'est pas orthogonal et l'homogénéisation insuffisante.
 - accroît la complexité de l'analyse.

COMMENT GAGNER DE LA PRÉCISION STATISTIQUE ?

- Augmenter le nombre de répétitions :
 - la variance des estimateurs diminue proportionnellement au nombre d'observations si les répétitions sont réalisées dans les mêmes conditions,
 - mais la variance résiduelle peut augmenter si on est obligé de prendre une population plus hétérogène.
Exemple : tests sur des rats élevés dans un même laboratoire vs rats obtenus par différents fournisseurs. Ou bien nécessité de mener l'expérience sur plusieurs champs, plusieurs centres.
- Contrôle local (ou blocking). On regroupe les unités expérimentales en fonctions de critères ou facteurs : on essaie d'avoir la population la plus homogène possible à l'intérieur de chaque groupe.
 - diminue la variance et augmente la puissance des tests en enlevant des sources de variabilité.
 - peut augmenter la variance si le plan n'est pas orthogonal et l'homogénéisation insuffisante.
 - accroît la complexité de l'analyse.

COMMENT GAGNER DE LA PRÉCISION STATISTIQUE ?

- Augmenter le nombre de répétitions :
 - la variance des estimateurs diminue proportionnellement au nombre d'observations si les répétitions sont réalisées dans les mêmes conditions,
 - mais la variance résiduelle peut augmenter si on est obligé de prendre une population plus hétérogène.
Exemple : tests sur des rats élevés dans un même laboratoire vs rats obtenus par différents fournisseurs. Ou bien nécessité de mener l'expérience sur plusieurs champs, plusieurs centres.
- Contrôle local (ou blocking). On regroupe les unités expérimentales en fonctions de critères ou facteurs : on essaie d'avoir la population la plus homogène possible à l'intérieur de chaque groupe.
 - diminue la variance et augmente la puissance des tests en enlevant des sources de variabilité.
 - peut augmenter la variance si le plan n'est pas orthogonal et l'homogénéisation insuffisante.
 - accroît la complexité de l'analyse.

COMMENT GAGNER DE LA PRÉCISION STATISTIQUE ?

- Augmenter le nombre de répétitions :
 - la variance des estimateurs diminue proportionnellement au nombre d'observations si les répétitions sont réalisées dans les mêmes conditions,
 - mais la variance résiduelle peut augmenter si on est obligé de prendre une population plus hétérogène.
Exemple : tests sur des rats élevés dans un même laboratoire vs rats obtenus par différents fournisseurs. Ou bien nécessité de mener l'expérience sur plusieurs champs, plusieurs centres.
- Contrôle local (ou blocking). On regroupe les unités expérimentales en fonctions de critères ou facteurs : on essaie d'avoir la population la plus homogène possible à l'intérieur de chaque groupe.
 - diminue la variance et augmente la puissance des tests en enlevant des sources de variabilité.
 - peut augmenter la variance si le plan n'est pas orthogonal et l'homogénéisation insuffisante.
 - accroît la complexité de l'analyse.

Deuxième partie II

TECHNIQUES DE BLOCAGE (BLOCKING)

EXEMPLE 2.1 CHOIX DU NOMBRE DE BLOCK

PROBLÈME

On dispose de p lots d'animaux de taille n . On teste un médicament et on observe la réponse x . A taille de l'expérience np fixée comment choisir n et p pour avoir la meilleure réponse ?

RÉPONSE

La réponse s'écrit $x_{ij} = \mu + a_i + \varepsilon_{ij}$. On note $\sigma_A^2 = \text{Var}(a_i)$ la variabilité inter lots et $\sigma^2 = \text{Var}(\varepsilon_{ij})$ la variabilité intra lot, on a :

$$\text{Var}(\bar{X}) = \frac{\sigma_A^2}{p} + \frac{\sigma^2}{np}$$

est minimum pour $n = 1$, i.e. p maximal. On diminue ainsi l'impact de la variabilité inter lot.

REMARQUE

On peut avoir des contraintes expérimentales, par exemple :

- *1 lot correspond à une portée qui oblige à avoir une taille de lot > 1 .*
- *On compte le nombre de bactéries à l'intérieur du canal d'une dent au microscope électronique. On fait un premier zoom sur plusieurs zones (zones primaires), puis un deuxième zoom plus fort en plusieurs endroits (zones secondaires) de la zone. L'expérimentateur préfère choisir 3 zones primaires et 4 secondaires car c'est plus simple. Le plan optimal est 12 zones primaires et 1 secondaire. On peut trouver un compromis : 6 zones primaires et 2 secondaires.*

EXEMPLE 2.2 : BLOCAGE SPATIAL, EXPÉRIENCE EN CHAMP

On souhaite tester 5 variétés de blés. On découpe un champ en parcelles. Pour contrôler les variations de fertilité, on regroupe les parcelles en blocs homogènes.

Bloc 1	1	2	3	4	5
--------	---	---	---	---	---

Bloc 2	3	1	5	2	4
--------	---	---	---	---	---

⋮

Bloc b	4	3	1	5	2
--------	---	---	---	---	---

DÉFINITION

On peut définir formellement un plan d'expérience comme une application qui affecte à chaque parcelle j du bloc i un traitement ;

$$(i, j) \mapsto d(i, j)$$

Modèle d'analyse : ANOVA à deux facteurs croisés

$$y_{i,j} = \beta_i + \tau_{d(i,j)} + \varepsilon_{ij}$$

En notation vectorielle :

$$Y = B\beta + T_d \tau + \varepsilon_{ij}$$

Ici les effets traitements représentent le **paramètre d'intérêt**. Le facteur bloc est un **paramètre de nuisance**.

DÉFINITION

Le plan est **complet équilibré** si

- Chaque traitement apparaît au moins une fois par bloc (complet)
- Chaque traitement apparaît le même nombre de fois dans chaque bloc (équilibré).

POURQUOI CHOISIR UN PLAN COMPLET ÉQUILIBRÉ

- C'est le plan qui a la plus faible matrice de variance pour le paramètre d'intérêt.
- C'est le plan qui est universellement optimal parmi tous les plans possibles.
- Le caractère aléatoire ou fixe des blocs ne va pas influencer l'analyse sur les traitements.
- C'est un plan orthogonal : en fait, mathématiquement, cela se traduit par le fait que les sous-espaces engendrés par les colonnes de A et B sont orthogonaux.

A RETENIR

L'orthogonalité est la première propriété que l'on cherchera à obtenir. C'est d'elle que découle les propriétés d'optimalité du plan d'expérience.

POURQUOI CHOISIR UN PLAN COMPLET ÉQUILIBRÉ

- C'est le plan qui a la plus faible matrice de variance pour le paramètre d'intérêt.
- C'est le plan qui est universellement optimal parmi tous les plans possibles.
- Le caractère aléatoire ou fixe des blocs ne va pas influencer l'analyse sur les traitements.
- C'est un plan orthogonal : en fait, mathématiquement, cela se traduit par le fait que les sous-espaces engendrés par les colonnes de A et B sont orthogonaux.

A RETENIR

L'orthogonalité est la première propriété que l'on cherchera à obtenir. C'est d'elle que découle les propriétés d'optimalité du plan d'expérience.

POURQUOI CHOISIR UN PLAN COMPLET ÉQUILIBRÉ

- C'est le plan qui a la plus faible matrice de variance pour le paramètre d'intérêt.
- C'est le plan qui est universellement optimal parmi tous les plans possibles.
- Le caractère aléatoire ou fixe des blocs ne va pas influencer l'analyse sur les traitements.
- C'est un plan orthogonal : en fait, mathématiquement, cela se traduit par le fait que les sous-espaces engendrés par les colonnes de A et B sont orthogonaux.

A RETENIR

L'orthogonalité est la première propriété que l'on cherchera à obtenir. C'est d'elle que découle les propriétés d'optimalité du plan d'expérience.

POURQUOI CHOISIR UN PLAN COMPLET ÉQUILIBRÉ

- C'est le plan qui a la plus faible matrice de variance pour le paramètre d'intérêt.
- C'est le plan qui est universellement optimal parmi tous les plans possibles.
- Le caractère aléatoire ou fixe des blocs ne va pas influencer l'analyse sur les traitements.
- C'est un plan orthogonal : en fait, mathématiquement, cela se traduit par le fait que les sous-espaces engendrés par les colonnes de A et B sont orthogonaux.

A RETENIR

L'orthogonalité est la première propriété que l'on cherchera à obtenir. C'est d'elle que découle les propriétés d'optimalité du plan d'expérience.

POURQUOI CHOISIR UN PLAN COMPLET ÉQUILIBRÉ

- C'est le plan qui a la plus faible matrice de variance pour le paramètre d'intérêt.
- C'est le plan qui est universellement optimal parmi tous les plans possibles.
- Le caractère aléatoire ou fixe des blocs ne va pas influencer l'analyse sur les traitements.
- C'est un plan orthogonal : en fait, mathématiquement, cela se traduit par le fait que les sous-espaces engendrés par les colonnes de A et B sont orthogonaux.

A RETENIR

L'orthogonalité est la première propriété que l'on cherchera à obtenir. C'est d'elle que découle les propriétés d'optimalité du plan d'expérience.

PROPRIÉTÉS D'UN PLAN ORTHOGONAL

- L'estimation des effets traitements se fait comme si les blocs n'existaient pas \leftrightarrow analyse simplifiée.
- L'estimation des effets traitements n'est pas modifiée si on enlève l'effet bloc du modèle (car non significatif).
- L'impact du "blocage" se fait essentiellement sur l'estimation de la variance résiduelle (dénominateur du test de student ou de Fisher). Attention : si on omet l'effet bloc dans le modèle, la variance résiduelle apparente augmente (biais positif) et donc certains effets des traitements peuvent devenir non significatifs alors qu'ils sont significatifs.
- Les tests de types 1 et 3 sont identiques et avec les bons degrés de liberté pour chacun des effets (caractérisation de l'orthogonalité).

PROPRIÉTÉS D'UN PLAN ORTHOGONAL

- L'estimation des effets traitements se fait comme si les blocs n'existaient pas \leftrightarrow analyse simplifiée.
- L'estimation des effets traitements n'est pas modifiée si on enlève l'effet bloc du modèle (car non significatif).
- L'impact du "blocage" se fait essentiellement sur l'estimation de la variance résiduelle (dénominateur du test de student ou de Fisher). Attention : si on omet l'effet bloc dans le modèle, la variance résiduelle apparente augmente (biais positif) et donc certains effets des traitements peuvent devenir non significatifs alors qu'ils sont significatifs.
- Les tests de types 1 et 3 sont identiques et avec les bons degrés de liberté pour chacun des effets (caractérisation de l'orthogonalité).

PROPRIÉTÉS D'UN PLAN ORTHOGONAL

- L'estimation des effets traitements se fait comme si les blocs n'existaient pas \leftrightarrow analyse simplifiée.
- L'estimation des effets traitements n'est pas modifiée si on enlève l'effet bloc du modèle (car non significatif).
- L'impact du "blocage" se fait essentiellement sur l'estimation de la variance résiduelle (dénominateur du test de student ou de Fisher). Attention : si on omet l'effet bloc dans le modèle, la variance résiduelle apparente augmente (biais positif) et donc certains effets des traitements peuvent devenir non significatifs alors qu'ils sont significatifs.
- Les tests de types 1 et 3 sont identiques et avec les bons degrés de liberté pour chacun des effets (caractérisation de l'orthogonalité).

PROPRIÉTÉS D'UN PLAN ORTHOGONAL

- L'estimation des effets traitements se fait comme si les blocs n'existaient pas \leftrightarrow analyse simplifiée.
- L'estimation des effets traitements n'est pas modifiée si on enlève l'effet bloc du modèle (car non significatif).
- L'impact du "blocage" se fait essentiellement sur l'estimation de la variance résiduelle (dénominateur du test de student ou de Fisher). Attention : si on omet l'effet bloc dans le modèle, la variance résiduelle apparente augmente (biais positif) et donc certains effets des traitements peuvent devenir non significatifs alors qu'ils sont significatifs.
- Les tests de types 1 et 3 sont identiques et avec les bons degrés de liberté pour chacun des effets (caractérisation de l'orthogonalité).

PROPRIÉTÉS D'UN PLAN ORTHOGONAL

- L'estimation des effets traitements se fait comme si les blocs n'existaient pas \leftrightarrow analyse simplifiée.
- L'estimation des effets traitements n'est pas modifiée si on enlève l'effet bloc du modèle (car non significatif).
- L'impact du "blocage" se fait essentiellement sur l'estimation de la variance résiduelle (dénominateur du test de student ou de Fisher). Attention : si on omet l'effet bloc dans le modèle, la variance résiduelle apparente augmente (biais positif) et donc certains effets des traitements peuvent devenir non significatifs alors qu'ils sont significatifs.
- Les tests de types 1 et 3 sont identiques et avec les bons degrés de liberté pour chacun des effets (caractérisation de l'orthogonalité).

CARACTÉRISATION D'UN PLAN ORTHOGONAL À 2 FACTEURS.

Une condition nécessaire et suffisante pour qu'un plan à 2 facteurs croisés soit orthogonal est :

$$n_{ij} = \frac{n_{i+} n_{+j}}{n_{++}} \quad (1)$$

où

- n_{ij} est le nombre d'apparitions conjointes des niveaux i et j des facteurs A et B.
- $n_{i+} = \sum_j n_{ij}$ = nombre de fois où le niveau i du facteur A apparaît dans le plan.
- $n_{+j} = \sum_i n_{ij}$ = nombre de fois où le niveau j du facteur B apparaît dans le plan.

REMARQUE

La condition (1) est équivalente à $\frac{n_{ij}}{n_{i'j}} = \frac{n_{i+}}{n_{i'+}}$

CAS DES PLANS EN BLOCS DE TAILLES ÉGALES

Dans le cas de plans en blocs de taille fixe, $n_{i+} = \text{cte}$. Un plan sera donc orthogonal si le nombre d'occurrences du traitement j dans chacun des blocs est le même et vaut n_{+j}/b ou b nombre de blocs. Les blocs sont composés de manière identique à une permutation intra-bloc près des unités statistiques.

Exemple avec 3 traitements :

bloc 1	1	1	2	2	3
bloc 2	2	2	1	3	1
bloc 3	1	2	3	1	2
bloc 4	3	2	2	1	1

CAS DES BLOCKS DE TAILLES INÉGALES

On veut tester trois régimes alimentaires sur des rats. On dispose de 3 portées de rats : deux de 6 rats et une de 9 rats. On peut construire un plan orthogonal

portée 1	1	1	2	2	3	3			
portée 2	1	1	2	2	3	3			
portée 3	1	1	1	2	2	2	3	3	3

LES PLANS EN BLOCS INCOMPLETS

Les PBIE sont historiquement issus des expériences agronomiques. On a b blocs, k parcelles et t traitements avec $t > k$.

On ne peut pas construire un plan complet équilibré.

On cherchera à construire un plan en blocs incomplets équilibrés PBIE (en anglais BIBD : balanced incomplete block design)

DÉFINITION

Un plan en b blocs de taille k est un PBIE si

- chaque traitement apparaît le même nombre de fois, noté r dans le plan (plan équirépété).
- chaque paire de traitements distincts apparaît λ fois ensemble (i.e. dans un même bloc) :

$$\lambda = \frac{r(k-1)}{I-1}$$

où $I = bk/r$

LES PLANS EN BLOCS INCOMPLETS

Les PBIE sont historiquement issus des expériences agronomiques. On a b blocs, k parcelles et t traitements avec $t > k$.

On ne peut pas construire un plan complet équilibré.

On cherchera à construire un plan en blocs incomplets équilibrés PBIE (en anglais BIBD : balanced incomplete block design)

DÉFINITION

Un plan en b blocs de taille k est un PBIE si

- chaque traitement apparaît le même nombre de fois, noté r dans le plan (plan équirépété).
- chaque paire de traitements distincts apparaît λ fois ensemble (i.e. dans un même bloc) :

$$\lambda = \frac{r(k-1)}{I-1}$$

où $I = bk/r$

EXEMPLE 2.3 : PBIE ($b = 10, k = 3, t = 5, r = 6, \lambda = 3$).

bloc 1	1	2	3
bloc 2	2	3	5
bloc 3	1	3	4
bloc 4	2	4	5
bloc 5	1	3	5
bloc 6	1	2	5
bloc 7	2	3	4
bloc 8	1	2	4
bloc 9	3	4	5
bloc 10	1	4	5

PROPRIÉTÉS DES PBIE

- Les PBIE sont universellement optimaux parmi tous les plans de même taille.
- Les PBIE ne sont pas des plans orthogonaux en général (sauf si complet).
- Il faudra tenir compte du caractère aléatoire ou non des blocs.

EXEMPLE 2.3 : PBIE ($b = 10, k = 3, t = 5, r = 6, \lambda = 3$).

bloc 1	1	2	3
bloc 2	2	3	5
bloc 3	1	3	4
bloc 4	2	4	5
bloc 5	1	3	5
bloc 6	1	2	5
bloc 7	2	3	4
bloc 8	1	2	4
bloc 9	3	4	5
bloc 10	1	4	5

PROPRIÉTÉS DES PBIE

- Les PBIE sont universellement optimaux parmi tous les plans de même taille.
- Les PBIE ne sont pas des plans orthogonaux en général (sauf si complet).
- Il faudra tenir compte du caractère aléatoire ou non des blocs.

EXEMPLE 2.3 : PBIE ($b = 10, k = 3, t = 5, r = 6, \lambda = 3$).

bloc 1	1	2	3
bloc 2	2	3	5
bloc 3	1	3	4
bloc 4	2	4	5
bloc 5	1	3	5
bloc 6	1	2	5
bloc 7	2	3	4
bloc 8	1	2	4
bloc 9	3	4	5
bloc 10	1	4	5

PROPRIÉTÉS DES PBIE

- Les PBIE sont universellement optimaux parmi tous les plans de même taille.
- Les PBIE ne sont pas des plans orthogonaux en général (sauf si complet).
- Il faudra tenir compte du caractère aléatoire ou non des blocs.

EXEMPLE 2.3 : PBIE ($b = 10, k = 3, t = 5, r = 6, \lambda = 3$).

bloc 1	1	2	3
bloc 2	2	3	5
bloc 3	1	3	4
bloc 4	2	4	5
bloc 5	1	3	5
bloc 6	1	2	5
bloc 7	2	3	4
bloc 8	1	2	4
bloc 9	3	4	5
bloc 10	1	4	5

PROPRIÉTÉS DES PBIE

- Les PBIE sont universellement optimaux parmi tous les plans de même taille.
- Les PBIE ne sont pas des plans orthogonaux en général (sauf si complet).
- Il faudra tenir compte du caractère aléatoire ou non des blocs.

EXEMPLE 2.3 : PBIE ($b = 10, k = 3, t = 5, r = 6, \lambda = 3$).

bloc 1	1	2	3
bloc 2	2	3	5
bloc 3	1	3	4
bloc 4	2	4	5
bloc 5	1	3	5
bloc 6	1	2	5
bloc 7	2	3	4
bloc 8	1	2	4
bloc 9	3	4	5
bloc 10	1	4	5

PROPRIÉTÉS DES PBIE

- Les PBIE sont universellement optimaux parmi tous les plans de même taille.
- Les PBIE ne sont pas des plans orthogonaux en général (sauf si complet).
- Il faudra tenir compte du caractère aléatoire ou non des blocs.

CONDITION D'EXISTENCE D'UN PBIE

On ne connaît pas de conditions suffisantes d'existence de PBIE.

CONDITIONS NÉCESSAIRES D'EXISTENCE

- $bk = tr = N$ où N nombre d'essais
- $r(k - 1) = \lambda(t - 1)$
- $t > k$ et $b > t$
- Si $\lambda = 1$ et $k = 3$, alors $t = 6i + 1$ ou $t = 6i + 3$, $i > 0$.
- Si $\lambda = 2$ et $k = 3$, alors
 - soit $r = 3i$, $v = 3i + 1$ et $b = (3i + 1)i$, $i > 1$.
 - soit $r = 3i + 2$, $t = 3(i + 1)$ et $b = (3i + 2)(i + 1)$, $i > 0$.
- Si $b = t$ et b et t sont pairs, alors $r - \lambda$ est une carré parfait (carré d'un entier).
- Si $b = t$ et b et t sont impairs, alors l'équation $x^2 = (k - \lambda)y^2 + (-1)^{\frac{t-1}{2}} z^2$ a une solution en nombre entiers avec x , y et z non tous nul.

COMPLÉMENTAIRE D'UN PBIE

A partir d'un PBIE(b, k, t, r, λ), on peut construire un PBIE complémentaire en remplaçant dans chaque bloc les traitements présents par les traitements manquants.

On a : $b_c = b$, $k_c = k - t$, $t_c = t$, $r_c = b - r$ et $\lambda_c = b - 2r + \lambda$.

PBIE OBTENU À PARTIR DU PBIE DE L'EXEMPLE 2.

$b=10$, $k=3$, $t=5$, $r=6$, $\lambda = 3$.

bloc 1	1	2	3
bloc 2	2	3	5
bloc 3	1	3	4
bloc 4	2	4	5
bloc 5	1	3	5
bloc 6	1	2	5
bloc 7	2	3	4
bloc 8	1	2	4
bloc 9	3	4	5
bloc 10	1	4	5



bloc 1	4	5
bloc 2	1	4
bloc 3	2	5
bloc 4	1	3
bloc 5	2	4
bloc 6	3	4
bloc 7	1	5
bloc 8	3	5
bloc 9	1	2
bloc 10	2	3

FAUT-IL UTILISER DES PLANS EN BLOCS ?

Les plans en blocs permettent de regrouper des unités expérimentales qui se ressemblent.

- Si les unités expérimentales ne sont pas raisonnablement similaires, on a toujours intérêt à les regrouper en blocs,
- mais utiliser des blocs de manière injustifiée peut diminuer la performance d'un plan d'expérience,
- ...sauf si on utilise un plan en blocs complets équilibrés.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

1 Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

2 Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

LES TROIS TYPES DE BLOCAGES

① Division naturelle discrète :

- jeunes animaux → portées,
- animaux ou humains → sexe,
- parcelles dans des rizières → canaux d'irrigations,
- processus industriel → lots,
- tests de consommateurs → testeurs, semaines
- expérience en laboratoire → technicien, paillasse,

② Division continue (nécessite de choisir des bornes pour homogénéiser) :

- expérience en champ → régions petites et compactes,
- animaux ou humains → âge, poids, état de santé,
- expériences sur une longue période → période de temps,
- essais cliniques → sévérité de la maladie.

1 Choix du blocage en fonction du management de l'essai :

- essais cliniques → docteurs, infirmières,
- essais multi-centrique → centres,
- expérimentation en champ → parcelles en longueur pour le passage des tracteurs,
- expérimentation en champ → certains traitements ne peuvent être appliqués que sur des larges parcelles,
- expérience en laboratoire → technicien, paillasse.

Souvent, il y a plusieurs variables de blocages simultanées, il faut utiliser des dispositifs plus complexes :

- carrés latins : deux facteurs de blocages croisés
 - expérience en champ → lignes et colonnes,
 - mesures répétées dans le temps → sujet, période.
- split plot : deux facteurs de blocages hiérarchiques.
 - sondage → département puis communes.
 - traitements pour les yeux → individu puis oeil.
- etc.

1 Choix du blocage en fonction du management de l'essai :

- essais cliniques → docteurs, infirmières,
- essais multi-centrique → centres,
- expérimentation en champ → parcelles en longueur pour le passage des tracteurs,
- expérimentation en champ → certains traitements ne peuvent être appliqués que sur des larges parcelles,
- expérience en laboratoire → technicien, paillasse.

Souvent, il y a plusieurs variables de blocages simultanées, il faut utiliser des dispositifs plus complexes :

- carrés latins : deux facteurs de blocages croisés
 - expérience en champ → lignes et colonnes,
 - mesures répétées dans le temps → sujet, période.
- split plot : deux facteurs de blocages hiérarchiques.
 - sondage → département puis communes.
 - traitements pour les yeux → individu puis oeil.
- etc.

① Choix du blocage en fonction du management de l'essai :

- essais cliniques → docteurs, infirmières,
- essais multi-centrique → centres,
- expérimentation en champ → parcelles en longueur pour le passage des tracteurs,
- expérimentation en champ → certains traitements ne peuvent être appliqués que sur des larges parcelles,
- expérience en laboratoire → technicien, paillasse.

Souvent, il y a plusieurs variables de blocages simultanées, il faut utiliser des dispositifs plus complexes :

- carrés latins : deux facteurs de blocages croisés
 - expérience en champ → lignes et colonnes,
 - mesures répétées dans le temps → sujet, période.
- split plot : deux facteurs de blocages hiérarchiques.
 - sondage → département puis communes.
 - traitements pour les yeux → individu puis oeil.
- etc.

- ① Choix du blocage en fonction du management de l'essai :
- essais cliniques → docteurs, infirmières,
 - essais multi-centrique → centres,
 - expérimentation en champ → parcelles en longueur pour le passage des tracteurs,
 - expérimentation en champ → certains traitements ne peuvent être appliqués que sur des larges parcelles,
 - expérience en laboratoire → technicien, paillasse.

Souvent, il y a plusieurs variables de blocages simultanées, il faut utiliser des dispositifs plus complexes :

- carrés latins : deux facteurs de blocages croisés
 - expérience en champ → lignes et colonnes,
 - mesures répétées dans le temps → sujet, période.
- split plot : deux facteurs de blocages hiérarchiques.
 - sondage → département puis communes.
 - traitements pour les yeux → individu puis oeil.
- etc.

- ① Choix du blocage en fonction du management de l'essai :
- essais cliniques → docteurs, infirmières,
 - essais multi-centrique → centres,
 - expérimentation en champ → parcelles en longueur pour le passage des tracteurs,
 - expérimentation en champ → certains traitements ne peuvent être appliqués que sur des larges parcelles,
 - expérience en laboratoire → technicien, paillasse.

Souvent, il y a plusieurs variables de blocages simultanées, il faut utiliser des dispositifs plus complexes :

- carrés latins : deux facteurs de blocages croisés
 - expérience en champ → lignes et colonnes,
 - mesures répétées dans le temps → sujet, période.
- split plot : deux facteurs de blocages hiérarchiques.
 - sondage → département puis communes.
 - traitements pour les yeux → individu puis oeil.
- etc.

- ① Choix du blocage en fonction du management de l'essai :
- essais cliniques → docteurs, infirmières,
 - essais multi-centrique → centres,
 - expérimentation en champ → parcelles en longueur pour le passage des tracteurs,
 - expérimentation en champ → certains traitements ne peuvent être appliqués que sur des larges parcelles,
 - expérience en laboratoire → technicien, paillasse.

Souvent, il y a plusieurs variables de blocages simultanées, il faut utiliser des dispositifs plus complexes :

- carrés latins : deux facteurs de blocages croisés
 - expérience en champ → lignes et colonnes,
 - mesures répétées dans le temps → sujet, période.
- split plot : deux facteurs de blocages hiérarchiques.
 - sondage → département puis communes.
 - traitements pour les yeux → individu puis oeil.
- etc.

- 1 Choix du blocage en fonction du management de l'essai :
 - essais cliniques → docteurs, infirmières,
 - essais multi-centrique → centres,
 - expérimentation en champ → parcelles en longueur pour le passage des tracteurs,
 - expérimentation en champ → certains traitements ne peuvent être appliqués que sur des larges parcelles,
 - expérience en laboratoire → technicien, paillasse.

Souvent, il y a plusieurs variables de blocages simultanées, il faut utiliser des dispositifs plus complexes :

- carrés latins : deux facteurs de blocages croisés
 - expérience en champ → lignes et colonnes,
 - mesures répétées dans le temps → sujet, période.
- split plot : deux facteurs de blocages hiérarchiques.
 - sondage → département puis communes.
 - traitements pour les yeux → individu puis oeil.
- etc.

Troisième partie III

NOTION DE PLANS OPTIMAUX

MATRICE D'INFORMATION ASSOCIÉE À UN PARAMÈTRE

On considère un modèle linéaire que l'on peut décomposer en un paramètre d'intérêt α et un paramètre de nuisance β .

$$Y = A\alpha + B\beta + \varepsilon.$$

DÉFINITION

On appelle **matrice d'information** (ou matrice de précision) associée au paramètre α la matrice :

$$C[\alpha] = A' \text{pr}_B^\perp A$$

REMARQUE

La matrice d'information est "l'inverse" de la matrice $\text{Var}(\hat{\beta})$ à un facteur σ^2 près. En fait, c'est une inverse généralisée.

- 1 La matrice d'information est directement liée au plan d'expérience choist.
- 2 On cherche donc le plan pour lequel la matrice d'information est la **plus grande possible**.
- 3 La façon dont on va déterminer le notion de "plus grand" va définir une notion d'optimalité.

- 1 La matrice d'information est directement liée au plan d'expérience choist.
- 2 On cherche donc le plan pour lequel la matrice d'information est **la plus grande possible**.
- 3 La façon dont on va déterminer le notion de "plus grand" va définir une notion d'optimalité.

- 1 La matrice d'information est directement liée au plan d'expérience choist.
- 2 On cherche donc le plan pour lequel la matrice d'information est **la plus grande possible**.
- 3 La façon dont on va déterminer le notion de "plus grand" va définir une notion d'optimalité.

OPTIMALITÉ UNIFORME

La notion d'optimalité uniforme est la plus forte des notions d'optimalité.

DÉFINITION

Soit \mathcal{D} un ensemble de plans réalisables. On dit qu'un plan d^* est universellement optimal pour l'effet α si

$$C_d[\alpha] \leq C_{d^*}[\alpha], \quad \forall d \in \mathcal{D}.$$

où $C_d[\alpha] \leq C_{d^*}$ signifie que $C_{d^*} - C_d[\alpha]$ est définie positive (**ordre de Loewner**). Autrement dit, pour toute fonction estimable $c'\alpha$,

$$\text{var}_{d^*}(c'\hat{\alpha}) \leq \text{var}_d(c'\hat{\alpha})$$

PROPOSITION

Un plan en blocs complets est uniformément optimal parmi tous les plans équirépétés.

PROPOSITION

Le dernier plan de pesées de l'exemple introductif est uniformément optimal parmi tous les plans de pesées pour chacun des poids.

REMARQUE

Un plan uniformément optimal est souvent lié à la notion d'orthogonalité. Dans le dernier de plan de pesées les effets p_0 , p_1 , p_2 et p_3 sont orthogonaux entre eux car la matrice de plan d'expérience

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix}$$

a ses colonnes orthogonales entre elles.

PROPOSITION

Un plan en blocs complets est uniformément optimal parmi tous les plans équirépartis.

PROPOSITION

Le dernier plan de pesées de l'exemple introductif est uniformément optimal parmi tous les plans de pesées pour chacun des poids.

REMARQUE

Un plan uniformément optimal est souvent lié à la notion d'orthogonalité. Dans le dernier de plan de pesées les effets p_0 , p_1 , p_2 et p_3 sont orthogonaux entre eux car la matrice de plan d'expérience

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix}$$

a ses colonnes orthogonales entre elles.

PROPOSITION

Un plan en blocs complets est uniformément optimal parmi tous les plans équirépétés.

PROPOSITION

Le dernier plan de pesées de l'exemple introductif est uniformément optimal parmi tous les plans de pesées pour chacun des poids.

REMARQUE

Un plan uniformément optimal est souvent lié à la notion d'orthogonalité. Dans le dernier de plan de pesées les effets p_0 , p_1 , p_2 et p_3 sont orthogonaux entre eux car la matrice de plan d'expérience

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix}$$

a ses colonnes orthogonales entre elles.

CRITÈRES D'OPTIMALITÉ

- Le problème de l'optimalité uniforme est qu'elle est souvent trop forte et inatteignable. Ceci est dû au fait que l'ordre de Loewner est un ordre partiel : deux matrices d'information ne sont pas toujours comparables.
- Idée : on va comparer les matrices d'information via un critère Φ à valeurs dans \mathbb{R} . On dira que le plan d^* est meilleur que le plan d pour le critère Φ si

$$\Phi(C_{d^*}[\alpha]) \leq \Phi(C_d[\alpha])$$

CRITÈRES D'OPTIMALITÉ

- Le problème de l'optimalité uniforme est qu'elle est souvent trop forte et inatteignable. Ceci est dû au fait que l'ordre de Loewner est un ordre partiel : deux matrices d'information ne sont pas toujours comparables.
- Idée : on va comparer les matrices d'information via un critère Φ à valeurs dans \mathbb{R} . On dira que le plan d^* est meilleur que le plan d pour le critère Φ si

$$\Phi(C_{d^*}[\alpha]) \leq \Phi(C_d[\alpha])$$

EXEMPLES CLASSIQUES DE CRITÈRES

- Critère de **D-optimalité** :

$$\Phi_D(C) = \sum_{i=1}^{t-1} \log(\lambda_i(C)) = \log(\text{"det"}(C)).$$

Minimiser Φ_D revient à minimiser le volume (dans l'espace des contrastes) de l'ellipsoïde de confiance sous l'hypothèse de normalité des erreurs.

- Critère de **A-optimalité** :

$$\Phi_A(C) = \sum_{i=1}^{t-1} \frac{1}{\lambda_i(C)} = \text{tr}(V).$$

Minimiser Φ_A revient à minimiser la variance moyenne des estimateurs des contrastes

EXEMPLES CLASSIQUES DE CRITÈRES

- Critère de **D-optimalité** :

$$\Phi_D(C) = \sum_{i=1}^{t-1} \log(\lambda_i(C)) = \log(\text{"det"}(C)).$$

Minimiser Φ_D revient à minimiser le volume (dans l'espace des contrastes) de l'ellipsoïde de confiance sous l'hypothèse de normalité des erreurs.

- Critère de **A-optimalité** :

$$\Phi_A(C) = \sum_{i=1}^{t-1} \frac{1}{\lambda_i(C)} = \text{tr}(V).$$

Minimiser Φ_A revient à minimiser la variance moyenne des estimateurs des contrastes

- Critère de **E-optimalité**

$$\Phi_E(C) = \left[\min_{i \in \{1, \dots, t-1\}} \lambda_i(C) \right]^{-1} = \max_{i \in \{1, \dots, t-1\}} \lambda_i(V).$$

Minimiser Φ_E revient à minimiser la variance maximale parmi les contrastes normés sur les traitements. Cela revient aussi, sous l'hypothèse de normalité des erreurs, à maximiser la puissance du test ($H_0 : \tau = 0$ contre $\|\tau\| = c > 0$).

- Critère de **MV-optimalité**

$$\Phi_{MV}(V) = \max_{i \neq j} \text{Var}(\hat{\tau}_i - \hat{\tau}_j)$$

Le critère MV consiste donc à minimiser la variance maximum des comparaisons deux à deux des traitements.

- Critère de **E-optimalité**

$$\Phi_E(C) = \left[\min_{i \in \{1, \dots, t-1\}} \lambda_i(C) \right]^{-1} = \max_{i \in \{1, \dots, t-1\}} \lambda_i(V).$$

Minimiser Φ_E revient à minimiser la variance maximale parmi les contrastes normés sur les traitements. Cela revient aussi, sous l'hypothèse de normalité des erreurs, à maximiser la puissance du test ($H_0 : \tau = 0$ contre $\|\tau\| = c > 0$).

- Critère de **MV-optimalité**

$$\Phi_{MV}(V) = \max_{i \neq j} \text{Var}(\hat{\tau}_i - \hat{\tau}_j)$$

Le critère MV consiste donc à minimiser la variance maximum des comparaisons deux à deux des traitements.

OPTIMALITÉ UNIVERSELLE

En analyse de la variance, lorsque les traitements jouent des rôles symétriques, on peut définir une notion d'optimalité plus forte que l'optimalité obtenue par des critères et moins forte que l'optimalité uniforme.

DÉFINITION

(Kiefer, 1975)

Un plan d^* est **universellement optimal** parmi une classe \mathcal{D} de plans s'il minimise tous les critères $\Phi(C_d)$ à valeur dans $\bar{\mathbb{R}}$ vérifiant :

- 1 Φ est invariant par permutation simultanée ligne-colonne,
- 2 $\Phi(tC)$ est décroissant en t pour $t > 0$,
- 3 $\Phi(C)$ est convexe.

REMARQUE

Un plan universellement optimal est également A-, D-, E- et MV-optimal.

PROPOSITION

Un plan en bloc complet équilibré est universellement optimal parmi tous les plans de même taille.

PROPOSITION

Un plan en bloc incomplet équilibré (PBIE) est universellement optimal parmi tous les plans de même taille.

Quatrième partie IV

RANDOMISATION

INTRODUCTION

Les expériences randomisées ont été introduite dans les années 1920 par Fisher en agronomie.

Exemple

On veut comparer t engrais ou variétés de plantes. On découpe le champs en parcelles (sans structure de blocs). Sur chaque parcelle, on applique un traitement.

PLAN SYSTÉMATIQUE

Un premier plan qui vient à l'esprit est un **plan systématique**.

1	2	3	1	2	3	1	2	3
---	---	---	---	---	---	---	---	---

Inconvénient : il peut y avoir une structure cachée de fertilité (due par exemple à un essai antérieur) qui corresponde à la structure du plan systématique. Cela peut favoriser un traitement.

PLAN AVEC PARCELLES TÉMOINS

- Une autre possibilité est d'utiliser des **parcelles témoins**. Par exemple

1	T	2	3	T	1	2	T	3
---	---	---	---	---	---	---	---	---

- On peut montrer que le nombre n_o optimal de témoins pour une expérience avec t traitements et r répétitions est :

$$n_o = r\sqrt{t}$$

Par exemple, si $t=4$, il faudra mettre deux fois plus de répétitions pour les témoins que pour les autres traitements.

- Inconvénient des témoins : les parcelles utilisées pour les témoins sont perdues pour les traitements. Cette méthode ancienne est maintenant peu utilisée.

NON VALIDITÉ DU MODÈLE STATISTIQUE

Afin de voir d'où vient le problème, écrivons le modèle statistique associé :

$$y_i = \mu + \tau_{d(i)} + u_i + \varepsilon_i, \quad i = 1, \dots, t.$$

où

- ε_i l'erreur de mesure.
- u_i correspond à la fertilité de la parcelle i .

Les u_i sont **fixes** mais **non connus**. On a :

- $\mathbb{E}(y_i) = \tau_{d(i)} + u_i$.
- $\text{Var}(y_i) = \sigma_e^2$

On voit que les variations de fertilité apporte un biais au modèle et l'ANOVA ne peut pas être validée.

RANDOMISATION TOTALE

Fisher a eu l'idée de laisser faire le hasard en utilisant un procédé de randomisation. On tire au sort les parcelles que l'on va affecter aux traitements.

Par exemple :

- On choisit au hasard sans remise 3 parcelles que l'on affecte au premier traitement.
- Puis on prend au hasard sans remise 3 parcelles parmi les restantes que l'on affecte au deuxième traitement.
- On affecte les trois dernières au troisième traitement.

CONSÉQUENCE DE LA RANDOMISATION

- l'effets u_i deviennent aléatoire.

- $\mathbb{E}(u_i) = \bar{u} = \frac{1}{n} \sum_{i=1}^n u_i.$

- $\text{Var}(u_i) = \frac{1}{n} \sum_i (u_i - \bar{u})^2$

- L'espérance peut être intégrée à la moyenne générale μ du modèle
- La variance de u_i peut être intégrée aux effets résiduels.
- $\mathbb{E}(y_i) = \mu + \tau_i$ et $\text{Var}(y_i) = \sigma_u^2 + \sigma_e^2$

La randomisation a transformé le biais en variance !

CONSÉQUENCE DE LA RANDOMISATION

- l'effets u_i deviennent aléatoire.

- $\mathbb{E}(u_i) = \bar{u} = \frac{1}{n} \sum_{i=1}^n u_i.$

- $\text{Var}(u_i) = \frac{1}{n} \sum_i (u_i - \bar{u})^2$

- L'espérance peut être intégrée à la moyenne générale μ du modèle
- La variance de u_i peut être intégrée aux effets résiduels.
- $\mathbb{E}(y_i) = \mu + \tau_i$ et $\text{Var}(y_i) = \sigma_u^2 + \sigma_e^2$

La randomisation a transformé le biais en variance !

CONSÉQUENCE DE LA RANDOMISATION

- l'effets u_i deviennent aléatoire.

- $\mathbb{E}(u_i) = \bar{u} = \frac{1}{n} \sum_{i=1}^n u_i.$

- $\text{Var}(u_i) = \frac{1}{n} \sum_i (u_i - \bar{u})^2$

- L'espérance peut être intégrée à la moyenne générale μ du modèle
- La variance de u_i peut être intégrée aux effets résiduels.
- $\mathbb{E}(y_i) = \mu + \tau_i$ et $\text{Var}(y_i) = \sigma_u^2 + \sigma_e^2$

La randomisation a transformé le biais en variance !

CONSÉQUENCE DE LA RANDOMISATION

- l'effets u_i deviennent aléatoire.

- $\mathbb{E}(u_i) = \bar{u} = \frac{1}{n} \sum_{i=1}^n u_i.$

- $\text{Var}(u_i) = \frac{1}{n} \sum_i (u_i - \bar{u})^2$

- L'espérance peut être intégrée à la moyenne générale μ du modèle
- La variance de u_i peut être intégrée aux effets résiduels.
- $\mathbb{E}(y_i) = \mu + \tau_i$ et $\text{Var}(y_i) = \sigma_u^2 + \sigma_e^2$

La randomisation a transformé le biais en variance !

CONSÉQUENCE DE LA RANDOMISATION

- l'effets u_i deviennent aléatoire.

- $\mathbb{E}(u_i) = \bar{u} = \frac{1}{n} \sum_{i=1}^n u_i.$

- $\text{Var}(u_i) = \frac{1}{n} \sum_i (u_i - \bar{u})^2$

- L'espérance peut être intégrée à la moyenne générale μ du modèle
- La variance de u_i peut être intégrée aux effets résiduels.
- $\mathbb{E}(y_i) = \mu + \tau_i$ et $\text{Var}(y_i) = \sigma_u^2 + \sigma_e^2$

La randomisation a transformé le biais en variance !

CONSÉQUENCE DE LA RANDOMISATION

- l'effets u_i deviennent aléatoire.

- $\mathbb{E}(u_i) = \bar{u} = \frac{1}{n} \sum_{i=1}^n u_i.$

- $\text{Var}(u_i) = \frac{1}{n} \sum_i (u_i - \bar{u})^2$

- L'espérance peut être intégrée à la moyenne générale μ du modèle
- La variance de u_i peut être intégrée aux effets résiduels.
- $\mathbb{E}(y_i) = \mu + \tau_i$ et $\text{Var}(y_i) = \sigma_u^2 + \sigma_e^2$

La randomisation a transformé le biais en variance !

CONSÉQUENCE DE LA RANDOMISATION

- l'effets u_i deviennent aléatoire.

- $\mathbb{E}(u_i) = \bar{u} = \frac{1}{n} \sum_{i=1}^n u_i.$

- $\text{Var}(u_i) = \frac{1}{n} \sum_i (u_i - \bar{u})^2$

- L'espérance peut être intégrée à la moyenne générale μ du modèle
- La variance de u_i peut être intégrée aux effets résiduels.
- $\mathbb{E}(y_i) = \mu + \tau_i$ et $\text{Var}(y_i) = \sigma_u^2 + \sigma_e^2$

La randomisation a transformé le biais en variance !

AUTRE EFFET DE LA RANDOMISATION

La randomisation évite un autre biais important : le **biais de sélection** par l'expérimentateur :

- On veut comparer 5 fongicides sur des rosiers. Malgré toute l'objectivité du monde, l'expérimentateur aura tendance à choisir en premier les roses les belles et donc de privilégier le traitement correspondant.
- En essai clinique le médecin aura tendance à privilégier le nouveau traitement sur les patients les plus atteints.
- En expérimentation ; on va utiliser le meilleur matériel pour le produit que l'on préfère.

La randomisation garantit l'objectivité de l'étude !

En essais cliniques aucune étude n'est validée si elle n'est pas randomisée en double aveugle (sauf impossibilité) : la randomisation évite le biais de sélection, le double aveugle évite le biais de traitement par le médecin et le biais psychologique du patient.

RANDOMISATION RESTREINTE

- La randomisation totale peut aboutir à une disposition indésirable. Par exemple, on peut avoir :

1	1	1	2	2	2	3	3	3
---	---	---	---	---	---	---	---	---

même si c'est avec une probabilité faible.

- Dans le même ordre d'idée, si on a une structure en bloc, on veut que la randomisation respecte cette structure.
- Il faut s'assurer que la randomisation valide le modèle.

RANDOMISATION D'UN PLAN EN BLOC

On considère un plan en b blocs de taille k . On peut conduire la randomisation suivante :

- 1 On construit un plan en bloc virtuel (non réalisé), par exemple un PBCE ou un PBIE.
- 2 On crée une permutation aléatoire des blocs.
- 3 A l'intérieur de chaque bloc, on crée une permutation aléatoire des parcelles

REMARQUE

- *Si le plan initial est un PBCE (resp. un PBIE), il le restera après randomisation.*
- *Les blocs devenant aléatoires, ils devront être traités comme des effets aléatoires dans l'analyse.*

EXEMPLE : RANDOMISATION D'UN PBIE

- 1 Construction d'un PBIE virtuel.

bloc 1	1	2	3
bloc 2	1	2	4
bloc 3	1	3	4
bloc 4	2	3	4

- 2 Permutation aléatoire des blocs : $1 \rightarrow 3$, $2 \rightarrow 2$, $3 \rightarrow 4$ et $4 \rightarrow 1$.

bloc 4	2	3	4
bloc 2	1	2	4
bloc 1	1	2	3
bloc 3	1	3	4

- 3 On permute aléatoirement les parcelles dans les blocs. On obtient le plan final :

2	4	3
4	1	2
4	3	1
1	2	3

NOTION GÉNÉRALE DE RANDOMISATION

On peut distinguer trois types de randomisations restreintes

- **Type I** : randomisation des parcelles. À partir d'un plan virtuel donné, on choisit aléatoirement une permutation des parcelles parmi un ensemble fixé de permutations possibles.
- **Type II** : randomisation des traitements. À partir d'un plan virtuel donné, on choisit aléatoirement une permutation des traitements parmi un ensemble fixé de permutations possibles.
- **Type III** : sélection aléatoire d'un plan. On choisit aléatoirement un plan parmi un ensemble fixé de plans.

REMARQUE

Notons que le type III est la forme la plus générale d'envisager la randomisation.

PROPRIÉTÉS D'UNE RANDOMISATION

Les trois principales propriétés recherchées pour une randomisation sont :

- **le non biais** : si l'on s'intéresse à l'estimation des contrastes des traitements, on désire que l'espérance (à travers la randomisation) des estimateurs des contrastes soit non biaisée.
- **la validité faible** : on cherche à "valider" le test de nullité des effets traitements, en imposant au numérateur et au dénominateur du test de Fisher d'être en moyenne égaux sous l'hypothèse nulle : on désire que, en l'absence d'effets traitements, l'espérance du carré moyen associé aux traitements soit égale à l'espérance du carré moyen associé à l'erreur.
- **la validité forte** : on cherche à "valider" le test correspondant à une sous-hypothèse du test précédent : on désire que, en l'absence d'effets traitements, l'espérance du carré moyen associé à tout sous-espace de l'espace des effets traitements soit égal à l'espérance du carré moyen associé à l'erreur.

PROPOSITION

Bailey & Rowley, 1987

- (i) La randomisation d'un plan en blocs complets à t traitements est faiblement valide si, pour chaque couple de parcelles de blocs différents, la probabilité que ces deux parcelles reçoivent le même traitement est de $1/t$.*
- (ii) La randomisation d'un plan en blocs complets à t traitements est fortement valide si, pour chaque couple de parcelles distinctes d'un même bloc, la probabilité que ces deux parcelles reçoivent un couple donné de traitements distincts est de $1/t(t-1)$; alors que, pour chaque couple de parcelles de blocs différents, la probabilité que ces deux parcelles reçoivent un couple donné de traitements est de $1/t^2$.*
- (iii) Une randomisation faiblement valide d'un plan en blocs complets à t traitements peut être rendue fortement valide en randomisant totalement les traitements.*

Cinquième partie V

PLANS FACTORIELS

PLANS FACTORIELS COMPLET

On considère maintenant des expériences avec plus de deux facteurs. Les facteurs peuvent être

- contrôlés (âge, sexe, traitement,...)
- d'homogénéisation (lots, centres, blocs...).

DÉFINITION

Un plan d'expérience factoriel est **complet** si on a au moins une observation pour chaque combinaison de niveaux des facteurs.

- Par exemple, un plan complet à 3 facteurs avec 3, 4 et 5 niveaux aura au minimum $3 \times 4 \times 5 = 120$ observations.
- Dans un plan complet, **toutes les interactions sont estimables.**

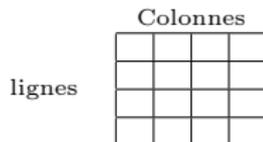
DÉFINITION

Un plan d'expérience factoriel complet est **équilibré** si on a le même nombre d'observations pour chaque combinaison de niveaux des facteurs.

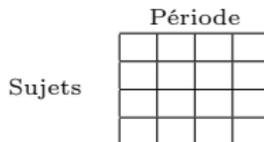
- Dans un plan complet équilibré, tous les effets (principaux et interactions) sont orthogonaux ; Les tests de type 1, 2 et 3 sont égaux.
- Un PBCE est une plan complet équilibré pour les effets blocs et traitements.

EXEMPLES DE PLANS FACTORIELS

- Dans une expérience en champ, on découpe le champs en parcelles suivant des lignes et des colonnes



- Dans une expérience en mesure répétée, on a les facteurs traitement, période et sujets



- On teste à la fois un type d'alimentation et une espèce de canard pour le gavage, ceci dans plusieurs exploitations. On a trois facteurs : deux facteurs d'intérêt, alimentation et canard et un facteur d'homogénéisation : exploitation.

CARRÉS LATINS

Problèmes :

- On a 3 facteurs ayant le même nombre de niveaux m .
- Un plan complet nécessiterait au moins m^3 expériences.
- On ne dispose que de m^2 expériences.
- On est obligé de sacrifier l'estimabilité de certains effets.
- On fait donc l'hypothèse qu'il **n'y a pas d'interaction** entre les effets et on veut que les effets principaux soient estimables entièrement.
- Sous cette hypothèse, on veut que les **effets principaux soient orthogonaux** (l'estimation des effets d'un facteur se font indépendamment de la présence ou non des autres facteurs : type 1=type 3).

Le **carré latin** répond à toutes ces contraintes.

DÉFINITION

Un carré latin est un tableau carré composé de nombre de 1 à m tel que chaque nombre apparaît une et une seule fois dans chaque ligne et chaque colonne. Les lignes correspondent aux niveaux du premier facteurs, les colonnes aux niveaux du deuxième facteur et les nombres aux niveaux du troisième facteur.

		facteur B				
	1	2	3	4	5	
facteur A	2	3	4	5	1	
	3	4	5	1	2	
	4	5	1	2	3	
	5	1	2	3	4	

EXEMPLE 5.1

On teste 5 types d'alimentation et 5 espèces de canard pour le gavage. On dispose de 5 volaillers recevant 5 canards. On fait l'hypothèse qu'il n'y a pas d'interactions entre le type de canard et l'alimentation. On peut utiliser le carré latin

		volailler				
		1	2	3	4	5
espèce	1	2	3	4	5	1
	2	3	4	5	1	2
	3	4	5	1	2	3
	4	5	1	2	3	4
	5	1	2	3	4	5

On pourra trouver des tables assez complètes de carrés latins dans

- Fisher R.A. et Yates F. (1963). Statistical tables for biological, agricultural and medical research. Edinburgh, Oliver and Boyd
- Pearson E.S. et Hartley H.O. (1966). Biometrika tables for statisticians (vol. I). Cambridge, University Press.

ANALYSE

```
proc glm data=canard;
class alimentation canard volailler;
model poids = alimentation canard volailler;
run;
quit;
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
alimentation	3	18934.00000	6311.33333	7.68	0.0177
canard	3	27980.00000	9326.66667	11.35	0.0069
volailler	3	1718.00000	572.66667	0.70	0.5872

Source	DF	Type III SS	Mean Square	F Value	Pr > F
alimentation	3	18934.00000	6311.33333	7.68	0.0177
canard	3	27980.00000	9326.66667	11.35	0.0069
volailler	3	1718.00000	572.66667	0.70	0.5872

On remarque les tests de type 1 et 3 sont identiques.

ANALYSE AVEC L'EFFET ALÉATOIRE "VOLAILLER"

L'effet poulailler est un paramètre d'homogénéisation, il doit donc être traité comme aléatoire (après randomisation).

```
proc mixed data=canard;
class alimentation canard volailler;
model poids = alimentation canard ;
random volailler;
run;quit;
```

Covariance Parameter Estimates	
Cov Parm	Estimate
volailler	0
Residual	738.89

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
alimentation	3	6	8.54	0.0138
canard	3	6	12.62	0.0053

Normalement, on devrait trouver exactement les mêmes résultats que précédemment pour les effets fixes, mais l'estimation à 0 du paramètre de variance associé à l'effet volailler a perturbé le test.

ANALYSE AVEC L'EFFET ALÉATOIRE "VOLAILLER", MÉTHODE MIVQUE 0..

En changeant la méthode d'estimation et de tests : MIVQUE0 au lieu de REML, on obtient les bons résultats.

```
proc mixed data=canard method=mivque0;
class alimentation canard volailler;
model poids = alimentation canard ;
random volailler;
run;quit;
```

Covariance Parameter Estimates	
Cov Parm	Estimate
volailler	0
Residual	822.00

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
alimentation	3	6	7.68	0.0177
canard	3	6	11.35	0.0069

la méthode MIVQUE0 tient compte de la structure particulière d'orthogonalité du plan.

MODÈLE AVEC INTERACTION

Que se passe-t-il si on rajoute une interaction canard*alimentation ?

```
proc glm data=canard;
class alimentation canard volailler;
model poids = alimentation canard volailler alimentation*canard;
run;
quit;
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
alimentation	3	18934.00000	6311.33333	.	.
canard	3	27980.00000	9326.66667	.	.
volailler	3	1718.00000	572.66667	.	.
alimentation*canard	6	4932.00000	822.00000	.	.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
alimentation	3	18934.00000	6311.33333	.	.
canard	3	27980.00000	9326.66667	.	.
volailler	0	0.00000	.	.	.
alimentation*canard	6	4932.00000	822.00000	.	.

On voit que l'effet canard*alimentation est partiellement confondu avec le troisième facteur. Par ailleurs, le plan est saturé.

MÊME CHOSE AVEC LA PROC MIXED

```
proc mixed data=canard method=mivque0;
class alimentation canard volailler;
model poids = alimentation | canard ;
random volailler;
run;quit;
```

Covariance Parameter Estimates	
Cov Parm	Estimate
volailler	38.2774
Residual	1.11E-12

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
alimentation	3	0	5.68E15	.
canard	3	0	8.4E15	.
alimentation*canard	9	0	6.66E14	.

L'interaction a récupéré les degrés de liberté pris par l'effet volailler en tant qu'effet fixe.

RANDOMISATION D'UN CARRÉ LATIN

Une randomisation valide d'un carré latin se fait de la façon suivante :

- 1 On permute aléatoirement les lignes du plan,
- 2 puis on permute aléatoirement les colonnes du plan.

Par exemple

1	2	3	4	5
2	3	4	5	1
3	4	5	1	2
4	5	1	2	3
5	1	2	3	4

→

4	1	3	2	5
1	3	5	4	2
5	2	4	3	1
3	5	2	1	4
2	4	1	5	3

EXEMPLE DE MESURE RÉPÉTÉE DANS LE TEMPS

On prend 4 patients sur lesquels on va donner 4 traitements différents pendant 4 jours. On a 3 effets : patient, période et traitement. On peut utiliser le dispositif :

	périodes			
patients	1	2	3	4
	2	3	4	1
	3	4	1	2
	4	1	2	3

Inconvénient : le traitement 2 est toujours précédé du traitement 1, 3 de 2, etc.

On peut laisser faire la randomisation ou utiliser un carré latin complet.

CARRÉS LATINS COMPLETS

DÉFINITION

Un **carré latin semi-complet** est un carré latin tel que chaque traitement est précédé par une et une seule fois par chaque autre traitement.

Exemple :

	périodes			
patients	1	2	3	4
	3	1	4	2
	2	4	1	3
	4	3	2	1

- Ce plan est optimal si on incorpore un effet de carry-over dans le modèle.
- Ce plan minimise l'impact du biais du carry-over si celui-ci n'est pas inclus dans l'analyse.
- La randomisation se fait en remplaçant la permutation des colonnes par une randimisation des traitements.

CARRÉ GRÉCO-LATIN

Contexte :

- On a m facteurs à m niveaux chacun.
- On suppose qu'il **n'y a pas d'interaction** entre les facteurs.
- Un plan complet nécessiterait m^4 niveaux. On ne dispose que de m^2 unités expérimentales.
- On veut que les effets principaux des facteurs soient orthogonaux entre eux.

DÉFINITION

Deux carrés latins de même taille sont **mutuellement orthogonaux** si en les superposant chaque couple de lettres latines et grecques apparaissent une et une seule fois. Le plan tableau résultant s'appelle **carré gréco-latin**.

$$\begin{array}{|c|c|c|c|} \hline a & b & c & d \\ \hline b & a & d & c \\ \hline c & d & a & b \\ \hline d & c & b & a \\ \hline \end{array} + \begin{array}{|c|c|c|c|} \hline \alpha & \beta & \gamma & \delta \\ \hline \gamma & \delta & \alpha & \beta \\ \hline \delta & \gamma & \beta & \alpha \\ \hline \beta & \alpha & \delta & \gamma \\ \hline \end{array} \longrightarrow \begin{array}{|c|c|c|c|} \hline a\alpha & b\beta & c\gamma & d\delta \\ \hline b\gamma & a\delta & d\alpha & c\beta \\ \hline c\delta & d\gamma & a\beta & b\alpha \\ \hline d\beta & c\alpha & b\delta & a\gamma \\ \hline \end{array}$$

- Les colonnes correspondent aux niveaux du premier facteur.
- Les lignes correspondent aux niveaux du deuxième facteur .
- Les lettres latines correspondent aux niveaux du troisième facteur.
- Les lettres grecques correspondent aux niveaux du quatrième facteur.

CARRÉS DE YOUDEN

Contexte :

- On dispose de t traitements, t sujets et k périodes ($k < t$).
- On veut construire un dispositif qui soit en bloc complet pour l'effet sujet et en bloc incomplet équilibré pour l'effet période.
- On a ainsi un plan où les trois facteurs en jeu sont orthogonaux
- Un tel dispositif s'appelle carré de **Youden** ou **carré latin incomplet**.
- Les conditions d'existence d'un tel dispositif sont les mêmes que celles d'un PBIE.

EXEMPLE DE CARRÉ DE YOUDEN

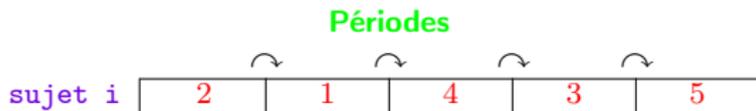
12	4	3	6
2	7	6	9
10	2	1	4
5	10	9	12
4	9	8	11
9	1	13	3
7	12	11	1
1	6	5	8
3	8	7	10
13	5	4	7
11	3	2	5
6	11	10	13
8	13	12	2

Sixième partie VI

PLANS EN CROSS-OVER

DÉFINITION

On cherche à comparer t traitements. Un plan en cross-over est un dispositif expérimental divisé en p périodes. Pour chaque période, chaque individu reçoit un traitement, en général différent du précédent.

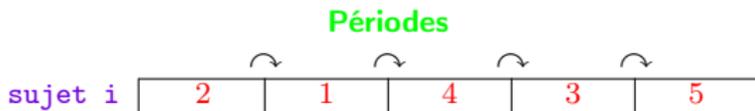


POURQUOI UTILISER UN PLAN EN CROSS-OVER ?

- En général la principale source de variabilité vient de l'effet sujet. Dans un cross-over, le sujet devient son propre témoin. Si le plan est orthogonal, l'effet sujet est éliminé pour l'analyse des effets traitements.
- Mais, on observe souvent un effet de carry-over (ou effet résiduel ou effet rémanent) qui doit être pris en compte dans l'analyse.
- On n'utilisera donc un cross-over que s'il y a une forte variabilité inter-individu.

DÉFINITION

On cherche à comparer t traitements. Un plan en cross-over est un dispositif expérimental divisé en p périodes. Pour chaque période, chaque individu reçoit un traitement, en général différent du précédent.

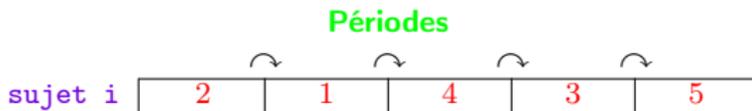


POURQUOI UTILISER UN PLAN EN CROSS-OVER ?

- En général la principale source de variabilité vient de l'effet sujet. Dans un cross-over, le sujet devient son **propre témoin**. Si le plan est orthogonal, l'effet sujet est éliminé pour l'analyse des effets traitements.
- Mais, on observe souvent un effet de **carry-over** (ou effet **résiduel** ou effet **rémanent**) qui doit être pris en compte dans l'analyse,
- On n'utilisera donc un cross-over que s'il y a une forte variabilité inter-individu.

DÉFINITION

On cherche à comparer t traitements. Un plan en cross-over est un dispositif expérimental divisé en p périodes. Pour chaque période, chaque individu reçoit un traitement, en général différent du précédent.

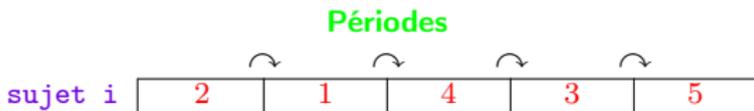


POURQUOI UTILISER UN PLAN EN CROSS-OVER ?

- En général la principale source de variabilité vient de l'effet sujet. Dans un cross-over, le sujet devient son **propre témoin**. Si le plan est orthogonal, l'effet sujet est éliminé pour l'analyse des effets traitements.
- Mais, on observe souvent un effet de **carry-over** (ou effet **résiduel** ou effet **rémanent**) qui doit être pris en compte dans l'analyse,
- On n'utilisera donc un cross-over que s'il y a une forte variabilité inter-individu.

DÉFINITION

On cherche à comparer t traitements. Un plan en cross-over est un dispositif expérimental divisé en p périodes. Pour chaque période, chaque individu reçoit un traitement, en général différent du précédent.

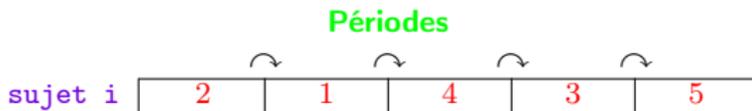


POURQUOI UTILISER UN PLAN EN CROSS-OVER ?

- En général la principale source de variabilité vient de l'effet sujet. Dans un cross-over, le sujet devient son **propre témoin**. Si le plan est orthogonal, l'effet sujet est éliminé pour l'analyse des effets traitements.
- Mais, on observe souvent un effet de **carry-over** (ou effet **résiduel** ou effet **rémanent**) qui doit être pris en compte dans l'analyse,
- On n'utilisera donc un cross-over que s'il y a une forte variabilité inter-individu.

DÉFINITION

On cherche à comparer t traitements. Un plan en cross-over est un dispositif expérimental divisé en p périodes. Pour chaque période, chaque individu reçoit un traitement, en général différent du précédent.



POURQUOI UTILISER UN PLAN EN CROSS-OVER ?

- En général la principale source de variabilité vient de l'effet sujet. Dans un cross-over, le sujet devient son **propre témoin**. Si le plan est orthogonal, l'effet sujet est éliminé pour l'analyse des effets traitements.
- Mais, on observe souvent un effet de **carry-over** (ou effet **résiduel** ou effet **rémanent**) qui doit être pris en compte dans l'analyse,
- On n'utilisera donc un cross-over que s'il y a une forte variabilité inter-individu.

MODÈLE D'ANALYSE AVEC EFFET DE CARRY-OVER

$$y_{i,j} = \mu + \pi_j + \beta_i + \tau_{d(i,j)} + \lambda_{d(i,j-1)} + \varepsilon_{ij}$$

avec

- i le numéro du sujet, j le numéro de la période,
- π_j l'effet période,
- $d(i, j)$ le traitement appliqué au sujet pendant la période j ,
- β_i l'effet du sujet i ,
- τ_k l'effet du sujet k ,
- λ_l l'effet rémanent du traitement l appliqué la période précédente,
- ε_{ij} sont des erreurs pas nécessairement non corrélées.

UN BON PLAN EN CROSS-OVER EST UN PLAN QUI DOIT :

- 1 éliminer l'effet période en équilibrant les traitements sur les périodes,
- 2 éliminer l'impact du carry-over en équilibrant les couples consécutifs (équilibre pour les voisinages),
- 3 éliminer l'effet sujet en utilisant un PBCE ou en minimiser l'impact en utilisant un PBIE.

REMARQUE

Les points 2 et 3 sont en fait contradictoires. On privilégiera 2 et on essaiera d'équilibrer pour les voisinages. Les recherches actuelles se portent sur des plans qui réalisent un compromis entre 2 et 3.

UN BON PLAN EN CROSS-OVER EST UN PLAN QUI DOIT :

- 1 éliminer l'effet période en équilibrant les traitements sur les périodes,
- 2 éliminer l'impact du carry-over en équilibrant les couples consécutifs (équilibre pour les voisinages),
- 3 éliminer l'effet sujet en utilisant un PBCE ou en minimiser l'impact en utilisant un PBIE.

REMARQUE

Les points 2 et 3 sont en fait contradictoires. On privilégiera 2 et on essaiera d'équilibrer pour les voisinages. Les recherches actuelles se portent sur des plans qui réalisent un compromis entre 2 et 3.

UN BON PLAN EN CROSS-OVER EST UN PLAN QUI DOIT :

- 1 éliminer l'effet période en équilibrant les traitements sur les périodes,
- 2 éliminer l'impact du carry-over en équilibrant les couples consécutifs (équilibre pour les voisinages),
- 3 éliminer l'effet sujet en utilisant un PBCE ou en minimiser l'impact en utilisant un PBIE.

REMARQUE

Les points 2 et 3 sont en fait contradictoires. On privilégiera 2 et on essaiera d'équilibrer pour les voisinages. Les recherches actuelles se portent sur des plans qui réalisent un compromis entre 2 et 3.

UN BON PLAN EN CROSS-OVER EST UN PLAN QUI DOIT :

- 1 éliminer l'effet période en équilibrant les traitements sur les périodes,
- 2 éliminer l'impact du carry-over en équilibrant les couples consécutifs (équilibre pour les voisinages),
- 3 éliminer l'effet sujet en utilisant un PBCE ou en minimiser l'impact en utilisant un PBIE.

REMARQUE

Les points 2 et 3 sont en fait contradictoires. On privilégiera 2 et on essaiera d'équilibrer pour les voisinages. Les recherches actuelles se portent sur des plans qui réalisent un compromis entre 2 et 3.

UN BON PLAN EN CROSS-OVER EST UN PLAN QUI DOIT :

- 1 éliminer l'effet période en équilibrant les traitements sur les périodes,
- 2 éliminer l'impact du carry-over en équilibrant les couples consécutifs (équilibre pour les voisinages),
- 3 éliminer l'effet sujet en utilisant un PBCE ou en minimiser l'impact en utilisant un PBIE.

REMARQUE

Les points 2 et 3 sont en fait contradictoires. On privilégiera 2 et on essaiera d'équilibrer pour les voisinages. Les recherches actuelles se portent sur des plans qui réalisent un compromis entre 2 et 3.

- Les plans les plus utilisés sont les **carrés latins semi-complets**. Ils réalisent les conditions 1 et 2. Ils respectent un équilibre pour les voisinages et sont universellement optimaux pour les plans de même taille parmi les plans équirépétés.
- Si on a beaucoup de sujet, on peut répliquer le carré latin.
- Si on a moins de périodes que de traitements, on peut utiliser un carré de Youden, qui a également de bonnes propriétés d'optimalité.
- une autre façon de construire des plans universellement optimaux est à partir d'une séquence de construire toutes les permutations possibles de traitements.

Avantage : facile à construire, plans optimaux.

Inconvénient : nécessite beaucoup de sujets.

- $t!$ sujets pour t traitements et t périodes.
- $A_t^k = \frac{t!}{(t-k)!}$ dans le cas général.

PLAN EN CROSS-OVER OBTENU PAR PERMUTATIONS

4 traitements sur 4 périodes et $4! = 24$ sujets

1	2	3	4
1	2	4	3
1	3	2	4
1	3	4	2
1	4	2	3
1	4	3	2
2	1	3	4
2	1	4	3
2	3	1	4
2	3	4	1
2	4	1	3
2	4	3	1
3	1	2	4
3	1	4	2
3	2	1	4
3	2	4	1
3	4	1	2
3	4	2	1
4	1	2	3
4	1	3	2
4	2	1	3
4	2	3	1
4	3	1	2
4	3	2	1

PLAN EN CROSS-OVER OBTENU PAR PERMUTATIONS

4 traitements sur 3 périodes et $4!/1! = 24$ sujets.

1	2	3
1	3	2
2	1	3
2	3	1
3	1	2
3	2	1
1	2	4
1	4	2
2	1	4
2	4	1
4	1	2
4	2	1
1	3	4
1	4	3
3	1	4
3	4	1
4	1	3
4	3	1
2	3	4
2	4	3
3	2	4
3	4	2
4	2	3
4	3	2

C'est un PBIE pour les sujets-traitements avec $r = 18$ et $\lambda = 12$.

CAS DU CROSS-OVER AB/BA

REMARQUE

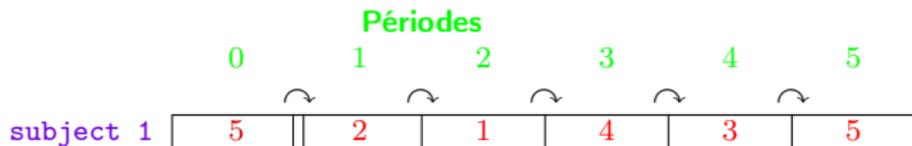
Les plans en cross-over à 2 traitements \times 2 périodes de type

$$\begin{array}{cc} A & B \\ B & A \end{array}$$

ne doivent pas être utilisés si on soupçonne un carry-over. Dans ce cas, il y a confusion entre les effets carry-over et traitement.

PLANS EN CROSS-OVER AVEC PRÉ-PÉRIODES

- pour éliminer l'effet de départ, on utilise parfois une **pré-période** sur laquelle on peut soit utiliser un traitement neutre ou un traitement actif. Soit on se sert de la pré-période comme **baseline**, soit on ne l'utilise pas dans l'analyse.



Le plan est **circulaire** si le traitement sur la pré-période est le même que celui sur la dernière période.

- Pour éliminer l'impact du carry-over, on peut utiliser des périodes intermédiaires sans traitement (**wash-out**), mais problème éthique ou de temps.

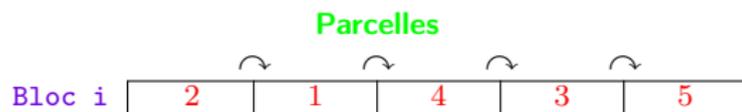
- Azaïs, Bailey et Monod (1993, Biometrics) ont proposé des plans circulaires optimaux lorsqu'il n'y a pas d'effet période.

$$\left[\begin{array}{c|ccccc} 5 & 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 5 & 3 & 1 & 4 \\ 1 & 3 & 5 & 2 & 4 & 1 \\ 5 & 4 & 3 & 2 & 1 & 5 \end{array} \right]$$

- On peut toujours utiliser la méthodes par toutes les permutations possibles.
- Azaïs (1987) a proposé une randomisation pour les plans circulaires. A partir d'un plan virtuel
 - On fait une permutation circulaire aléatoire des périodes,
 - On fait une permutation aléatoire des sujets,
 - On fait une permutation aléatoire des traitements.

CONTEXTE SPATIAL

Tous les plans en cross-over peuvent s'utiliser pour les effets de compétitions spatiales entre traitements au lieu d'effets rémanents.



Septième partie VII

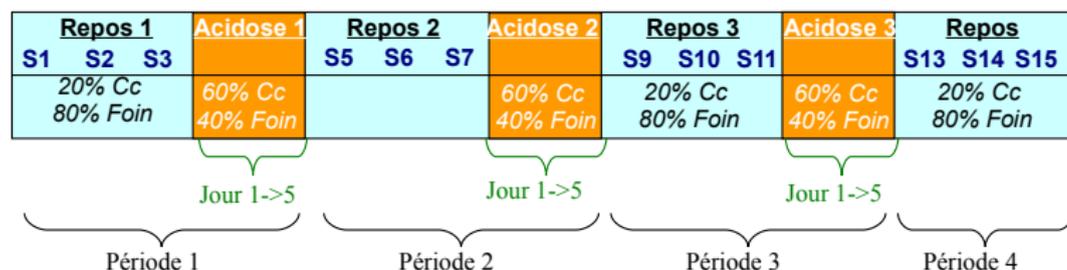
CAS CONCRETS

CAS N°1

- 3 lots de 12 vaches de 2 races (6 Holstein + 6 Montbéliardes dans chaque lot)
- 1 période préexpé de 3 semaines où toutes les vaches reçoivent le même régime : 1 mesure de la composition des laits individuels en fin de pré-expé.
- à l'issue de la période expé, les vaches des 3 lots reçoivent pendant 4 mois, 3 régimes différents : foin, herbe1 et herbe2.
- 3 périodes expé en Juin, Juillet et Août où 1/ des dosages de composition des laits individuels des vaches sont réalisés et 2/ des fromages sont fabriqués au cours de 3 journées consécutives avec le lait de mélange des 12 vaches de chaque lot (ainsi on a 3 périodes (juin, juillet et août) x 3 jours x 3 lots = 27 fromages).

CAS N°2

- 12 moutons
- 2 traitements (supplément levure vs placebo)
- 2 types d'alimentation (Acidogène / Repos)
- Tests comportementaux : 5 tests différents, réalisés une fois durant chaque challenges acidogènes, et une fois la semaine précédant ces challenges



S : semaine

CAS N°3

- 6 moutons canulés mâles castrés.
- Type de logement : en case individuelle au sol ou en case comportement. Les animaux sont placés en conditions constantes de lumière et de température.
- Lots expérimentaux : un unique lot de 6 + 2 moutons "secours" en régime alimentaire restreint.

AUGE (consommé)	Foin 022	Foin 022	Luzerne	Luzerne
RUMEN (introduit)	Foin 022	Luzerne	Foin 022	Luzerne
Traitements	①	②	③	④

CAS N°3 : PLAN D'EXPÉRIENCE PRÉVU

	<i>M1</i>	<i>M2</i>	<i>M3</i>	<i>M4</i>	<i>M5</i>	<i>M6</i>
<i>P1</i>	1	2	3	4	1	4
<i>P2</i>	2	4	1	3	3	2
<i>P3</i>	3	1	4	2	4	1
<i>P4</i>	4	3	2	1	2	3

CAS N°3 : PLAN RÉELLEMENT EFFECTUÉ

	M1	M2	M3	M4	M5	M6	M7
P1	<i>N</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>N</i>	<i>4</i>	
P2	2	4	1	3	3	2	
P3	3	1	4	2	4	1	
P4	<i>4</i>	<i>3</i>	<i>2</i>	<i>N</i>	<i>2</i>	<i>4</i>	
P4bis	4	3	2	1	2	3	4
P1bis		2	3	4	1	4	1

MODÈLE D'ANALYSE

$$y_{ij} = \text{traitement} + \text{jours-classe} + \text{jours-classe} \times \text{traitement} + a_i + b_{ij} \times \text{période} + e_{ij} + d$$

[traitement + jours-classe + jours-classe x traitement] = effets fixes

a_i = intercept qui est fct° de l'individu i (ex : M2 a tendance a mangé plus que les autres)

$[b_{ij} \times \text{période}]$ = effet aléatoire, possibilité d'une réponse différente de

l'individu statistique « mouton x traitement » en fct° des périodes de test.

e_{ij} = résiduel

Programme SAS

```
proc mixed data=_PROJ_Bilan_h1_SAS COVTEST;
  class MOUTONS TRAITEMENT JOURS_CLASSE PERIODE;
  model QI_H1 = JOURS_CLASSE TRAITEMENT JOURS_CLASSE*TRAITEMENT / ddfm=satterth
                                                    htype=3 SOLUTION ;
  repeated JOURS_CLASSE / sub=MOUTONS*TRAITEMENT type=ar(1);
  RANDOM INTERCEPT PERIODE / SUBJECT=MOUTONS*TRAITEMENT SOLUTION ;
  lsmeans JOURS_CLASSE TRAITEMENT JOURS_CLASSE *TRAITEMENT / pdiff adjust=TUKEY;
run;
```

CAS N°4

- 2 phases.
- Phase 1 : 6 lots de 4 vaches, 4 périodes + pré-période, 4 traitements.
- Phase 2 : 6 lots de 36 vaches, 4 périodes + pré-période, 6 traitements

Huitième partie VIII

ANALYSE D'UN CROSS-OVER AB/BA

CONTEXTE

- 2 traitements
- 2 périodes
- 2 groupes de patients
 - 1 groupe de taille n_1 reçoit A puis B
 - 1 groupe de taille n_2 reçoit B puis A
- Dispositif utilisé si on pense qu'il n'y a pas de carry-over.

MODÈLE SANS EFFET RÉMANENT

MODÈLE M1 (SANS EFFET RÉMANENT)

reponse = intercept + période + traitement + effet sujet + erreurs.

<i>période 1</i>	<i>période 2</i>
$y_{11k} = \mu + \pi_1 + \tau_A + s_{1k} + \varepsilon_{11k}$	$y_{11k} = \mu + \pi_2 + \tau_B + s_{1k} + \varepsilon_{12k}$
$y_{21k} = \mu + \pi_1 + \tau_B + s_{2k} + \varepsilon_{21k}$	$y_{22k} = \mu + \pi_2 + \tau_A + s_{2k} + \varepsilon_{22k}$

- les effets période, traitement et sujets sont orthogonaux (carré latin).
- L'analyse des traitements peut se faire avec l'effet sujet aléatoire ou fixe.
- Estimation des contrastes

$$\begin{aligned}
 \hat{\tau}_A - \hat{\tau}_B &= (y_{11\cdot} + y_{22\cdot}) - (y_{12\cdot} + y_{21\cdot}) \\
 &= \tau_A - \tau_B + (\varepsilon_{11\cdot} + \varepsilon_{22\cdot}) - (\varepsilon_{12\cdot} + \varepsilon_{21\cdot})
 \end{aligned}$$

ANALYSE AVEC EFFET RÉMANENT (CARRY-OVER)

MODÈLE M2 (AVEC EFFET RÉMANENT)

reponse = intercept + période + traitement + effet rémanent + sujet + erreurs.

période 1

$$y_{11k} = \mu + \pi_1 + \tau_A + s_{1k} + \varepsilon_{11k}$$

$$y_{21k} = \mu + \pi_1 + \tau_B + s_{2k} + \varepsilon_{21k}$$

période 2

$$y_{11k} = \mu + \pi_2 + \tau_B + \lambda_1 + s_{1k} + \varepsilon_{12k}$$

$$y_{22k} = \mu + \pi_2 + \tau_A + \lambda_2 + s_{2k} + \varepsilon_{22k}$$

- L'effet rémanent à deux ddl : il peut être considéré comme un facteur classique à trois niveaux : λ_0 , λ_1 et λ_2 avec la contrainte d'identifiabilité $\lambda_0 = 0$.

REPARAMÉTRISATION DE L'EFFET RÉMANENT

Posons

$$\begin{cases} \lambda_m = \frac{1}{2}(\lambda_1 + \lambda_2) \\ \lambda_d = \frac{1}{2}(\lambda_1 - \lambda_2) \end{cases}$$

On obtient

MODÈLE M3

<i>période 1</i>	<i>période 2</i>
$y_{11k} = \mu + \pi_1 + \tau_A + s_{1k} + \varepsilon_{11k}$	$y_{11k} = \mu + (\pi_2 + \lambda_m) + \tau_B + \lambda_d + s_{1k} + \varepsilon_{12k}$
$y_{21k} = \mu + \pi_1 + \tau_B + s_{2k} + \varepsilon_{21k}$	$y_{22k} = \mu + (\pi_2 + \lambda_m) + \tau_A - \lambda_d + s_{2k} + \varepsilon_{22k}$

- L'effet λ_m est confondu avec l'effet période.
- L'effet λ_d est inclu dans l'interaction traitement*période (ou groupe *période).
- On dira qu'il y a présence d'un carry-over si $\lambda_d = \frac{1}{2}(\lambda_1 - \lambda_2) = 0$.
- Si $\lambda_d = 0$ et $\lambda_m \neq 0$ signifie que l'effet rémanent est le même pour les deux traitements.

TEST DE L'EFFET λ_d

Tester l'effet λ_d est statistiquement équivalent à tester l'effet séquence (ou effet groupe) dans le modèle M4 :

MODÈLE M4

$$\text{réponse} = \mu + \text{période} + \text{trait} + \text{seq} + \text{sujet}(\text{seq}) + \varepsilon.$$

avec $\text{sujet}(\text{seq})$ en effet aléatoire.

<i>période 1</i>	<i>période 2</i>
$y_{11k} = \mu + \text{seq}_1 + \pi_1 + \tau_A + s_{1k} + \varepsilon_{11k}$	$y_{11k} = \mu + \text{seq}_1 + \pi_2 + \tau_B + s_{1k} + \varepsilon_{12k}$
$y_{21k} = \mu + \text{seq}_2 + \pi_1 + \tau_B + s_{2k} + \varepsilon_{21k}$	$y_{22k} = \mu + \text{seq}_2 + \pi_2 + \tau_A + s_{2k} + \varepsilon_{22k}$

- $\text{seq}_1 - \text{seq}_2 = 2\lambda_d = (\lambda_1 - \lambda_2)$
- l'effet sujet est emboîté dans l'effet séquence et donc dans l'effet rémanent.
- le test sur l'effet rémanent se fera donc en prenant comme source de variabilité l'effet sujet \leftrightarrow faible puissance. Test au niveau 10% préconisé.

ESTIMATION DE L'EFFET TRAITEMENT

La présence d'un effet carry-over dans les plans AB/BA pose des problèmes d'estimations et de tests : dans le modèle M3,

- soit on veut estimer sans biais par MCO le contraste sur les traitements, on ne doit se baser que sur la première période. On perd alors le bénéfice du cross-over, en particulier une grande variance de l'estimateur dû à l'effet sujet.

$$\hat{\tau}_A - \hat{\tau}_B = y_{11\cdot} - y_{21\cdot},$$

$$\mathbb{E}(\hat{\tau}_A - \hat{\tau}_B) = \tau_A - \tau_B,$$

$$\text{Var}(\hat{\tau}_A - \hat{\tau}_B) = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) (\sigma_s^2 + \sigma_e^2).$$

- soit on fait comme si l'effet rémanent était négligeable, on obtient alors un estimateur biaisé, mais avec une plus faible variance.

$$\hat{\tau}_A - \hat{\tau}_B = (y_{11\cdot} + y_{21\cdot}) - (y_{21\cdot} + y_{12\cdot}),$$

$$\mathbb{E}(\hat{\tau}_A - \hat{\tau}_B) = \tau_A - \tau_B - 2\lambda_d,$$

$$\text{Var}(\hat{\tau}_A - \hat{\tau}_B) = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \sigma_e^2.$$

Neuvième partie IX

QUELQUES MODÈLES D'ANOVA

MODÈLES HIÉRARCHIQUE À DEUX FACTEURS

On veut tester trois types de machines différentes :

- n_1 techniciens utiliseront la machine M_1
- n_2 techniciens utiliseront la machine M_2
- n_3 techniciens utiliseront la machine M_3

Chaque technicien utilisera la machine k jours. On note y_{ijk} la production observé par le technicien j sur la machine i le jour k . On a ainsi un modèle d'analyse de la variance à deux facteurs :

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk} \quad (1)$$

où

- α_i correspond à la performance de la machine i
- β_{ij} au "rendement moyen" du technicien j de la machine i
- ε_{ijk} correspond aux variations journalières du technicien (ij) .

C'est un cas particulier de plan incomplet, en effet,

	T_{11}	...	T_{1n_1}	T_{21}	...	T_{2n_2}	T_{31}	...	T_{3n_3}
M_1	×	...	×						
M_2				×	×			
M_3							×	×

une croix signifiant qu'il y a au moins une observation.

DÉFINITION

On dit que le plan hiérarchique est équilibré si $n_1 = n_2 = n_3 = \dots$

C'est un cas particulier de plan incomplet, en effet,

	T_{11}	...	T_{1n_1}	T_{21}	...	T_{2n_2}	T_{31}	...	T_{3n_3}
M_1	×	...	×						
M_2				×	×			
M_3							×	×

une croix signifiant qu'il y a au moins une observation.

DÉFINITION

On dit que le plan hiérarchique est équilibré si $n_1 = n_2 = n_3 = \dots$

FACTEUR EMBOÎTÉ

DÉFINITION

Chaque niveau du facteur technicien n'est présent qu'avec un seul niveau du facteur machine. On dit que le facteur technicien est emboîté (nested en anglais) dans le facteur machine.

DIAGRAMME DE HASSE

On note cette structure hiérarchique à l'aide du diagramme de Hasse :

$$\textit{machine} \longleftarrow \textit{technicien}$$

ou encore comme machine est emboîté dans l'intercept :

$$\textit{intercept} \longleftarrow \textit{machine} \longleftarrow \textit{technicien}$$

SYNTAXE SAS

Pour indiquer que le facteur technicien est emboîté dans machine, on utilise la syntaxe

```
rendement=machine technicien(machine)
```

mais on peut utiliser aussi

```
rendement= machine technicien*machine
```

ESTIMATION ET TESTS

- Cas où l'effet technicien est fixe :
 - il n'existe aucune fonction estimable associé à l'effet machine. Par exemple la commande lsmeans fait intervenir l'effet technicien.
 - On ne peut donc pas tester l'effet machine.

RÈGLE

En effet fixe, on ne peut pas estimer ou tester un effet s'il y a autre effet à droite du diagramme de Hasse.

- Si l'effet technicien est aléatoire, on peut estimer l'effet machine (en prenant en compte la variabilité des techniciens).

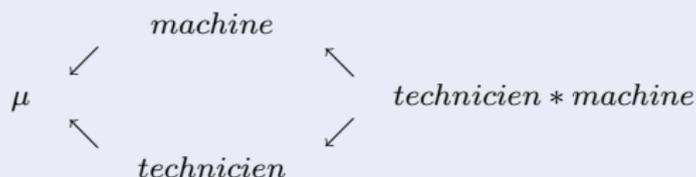
CAS DES FACTEURS CROISÉS AVEC INTERACTION

On suppose maintenant qu'un même technicien passe sur chacune des machines. On suppose également qu'il y a interaction. le modèle est :

$$\text{rendement} = \text{intercept} + \text{technicien} + \text{machine} + \text{technicien} * \text{machine}$$

DIAGRAMME DE HASSE CORRESPONDANT

l'interaction est emboîté dans les effets principaux :



CONSÉQUENCE

En présence d'interaction, on ne peut rien dire sur les facteurs principaux (estimation, tests). On est obligé de prendre en compte l'interaction.

CONSÉQUENCE

Si on retire les effets principaux dans le modèle, on ne change pas le modèle (seulement les sorties standard des logiciels).

RÈGLE

Quand dans un modèle d'ANOVA on enlève un effet alors qu'il est emboîté dans un autre, on ne change pas le modèle.

CONSÉQUENCE

Un modèle d'ANOVA ne dépend que des facteurs qui sont le plus à droite du diagramme de Hasse.

CONSÉQUENCE

En présence d'interaction, on ne peut rien dire sur les facteurs principaux (estimation, tests). On est obligé de prendre en compte l'interaction.

CONSÉQUENCE

Si on retire les effets principaux dans le modèle, on ne change pas le modèle (seulement les sorties standard des logiciels).

RÈGLE

Quand dans un modèle d'ANOVA on enlève un effet alors qu'il est emboîté dans un autre, on ne change pas le modèle.

CONSÉQUENCE

Un modèle d'ANOVA ne dépend que des facteurs qui sont le plus à droite du diagramme de Hasse.

CONSÉQUENCE

En présence d'interaction, on ne peut rien dire sur les facteurs principaux (estimation, tests). On est obligé de prendre en compte l'interaction.

CONSÉQUENCE

Si on retire les effets principaux dans le modèle, on ne change pas le modèle (seulement les sorties standard des logiciels).

RÈGLE

Quand dans un modèle d'ANOVA on enlève un effet alors qu'il est emboîté dans un autre, on ne change pas le modèle.

CONSÉQUENCE

Un modèle d'ANOVA ne dépend que des facteurs qui sont le plus à droite du diagramme de Hasse.

CONSÉQUENCE

En présence d'interaction, on ne peut rien dire sur les facteurs principaux (estimation, tests). On est obligé de prendre en compte l'interaction.

CONSÉQUENCE

Si on retire les effets principaux dans le modèle, on ne change pas le modèle (seulement les sorties standard des logiciels).

RÈGLE

Quand dans un modèle d'ANOVA on enlève un effet alors qu'il est emboîté dans un autre, on ne change pas le modèle.

CONSÉQUENCE

Un modèle d'ANOVA ne dépend que des facteurs qui sont le plus à droite du diagramme de Hasse.

Dixième partie X

TESTS

TEST D'UN MODÈLE CONTRE UN SOUS-MODÈLE

On suppose que les observations Y suivent le modèle linéaire \mathcal{M} défini par :

$$\begin{aligned}\mathcal{M} : \quad Y &= X\theta + \varepsilon \\ &= X_1\theta_1 + X_2\theta_2 + \varepsilon\end{aligned}$$

et on veut tester l'hypothèse, notée H_o : Y suit le modèle plus restrictif

$$\mathcal{M}_o : \quad Y = X_1\theta_1 + \varepsilon$$

Notation :

On note \hat{Y} et \hat{Y}_o les valeurs prédites de Y par les modèle \mathcal{M} et \mathcal{M}_o , i.e.

$$\mathcal{M} \quad \longrightarrow \quad \hat{Y} = P_X Y$$

$$\mathcal{M}_o \quad \longrightarrow \quad \hat{Y}_o = P_{X_1} Y$$

Décomposition des sommes des carrés :

$$\begin{aligned} \|Y\|^2 &= \|\hat{Y}_o\|^2 + \|(\hat{Y} - \hat{Y}_o)\|^2 + \|(Y - \hat{Y})\|^2 \\ &= SS(\theta_1) + SS(\theta_2|\theta_1) + SS_e(\theta_1, \theta_2) \end{aligned}$$

Statistique de test de Fisher :

$$F = \frac{MS(\theta_2|\theta_1)}{MS_e(\theta_1, \theta_2)} \sim \mathcal{F}[\text{ddl}(\theta_2|\theta_1) , \text{ddl}_e(\theta_1, \theta_2)]$$

TEST DANS L'ANOVA À 1 FACTEUR

Soit le modèle d'analyse de la variance à 1 facteur :

$$\mathcal{M} : \quad y_{ik} = \mu + \alpha_i + \varepsilon \quad \text{pour} \quad \begin{cases} i = 1, \dots, I \\ k = 1, \dots, n_i \end{cases}$$

Ici les tests de type I, II et III pour α consiste à tester le sous-modèle

$$\mathcal{M}_o : \quad y_{ik} = \mu + \varepsilon \quad \text{pour} \quad \begin{cases} i = 1, \dots, I \\ k = 1, \dots, n_i \end{cases}$$

contre le modèle \mathcal{M} .

ANOVA À 2 FACTEURS SANS INTERACTION : TYPE III

Soit le modèle

$$\mathcal{M} : y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad \text{pour} \quad \begin{cases} i = 1, \dots, I \\ j = 1, \dots, J \\ k = 1, \dots, n_{ij} \geq 1 \end{cases}$$

Test de type III pour α : c'est simplement le test du sous-modèle

$$\mathcal{M}_o : Y = \mu \mathbf{1}_n + B\beta + \varepsilon$$

contre le modèle :

$$\mathcal{M} : Y = \mu \mathbf{1}_n + A\alpha + B\beta + \varepsilon$$

ANOVA À 2 FACTEURS SANS INTERACTION : TYPE I

On suppose que l'on a entré d'abord α puis β dans "model" de la proc GLM, i.e. :

$$\text{model } y = \alpha \beta$$

Test de type I pour α

Revient à tester

$$\mathcal{M}_0 : Y = \mu \mathbf{1}_n + \varepsilon$$

contre

$$\mathcal{M}_1 : Y = \mu \mathbf{1}_n + A\alpha + \varepsilon$$

ATTENTION! : Le dénominateur du test de fisher n'est pas obtenu dans le modèle \mathcal{M}_1 mais dans le modèle complet \mathcal{M} .

Test de type I pour β

Il s'agit du même test que le test de type III pour β .

Principe général du test de type I

Ce test dépend de l'ordre dans lequel on rentre les facteurs dans "*model*" $y = \dots$ de la proc GLM. Pour tester un effet, on agit comme suit :

- 1) On prend \mathcal{M}_1 le modèle incluant cet effet et tous les autres effets placés avant.
- 2) On prend \mathcal{M}_o le modèle \mathcal{M}_1 auquel on a enlevé l'effet à tester.
- 3) On écrit la statistique F de Fisher correspondant au test du sous-modèle \mathcal{M}_o contre le modèle \mathcal{M}_1 .
- 4) On remplace le dénominateur par $MS_e(\mathcal{M})$, i.e. l'estimateur de σ^2 dans le modèle \mathcal{M} initial.

ANOVA À DEUX FACTEURS AVEC INTERACTION : TYPE III

Soit le modèle

$$\mathcal{M} : y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha * \beta)_{ij} + \varepsilon_{ijk} \quad \text{pour} \quad \begin{cases} i = 1, \dots, I \\ j = 1, \dots, J \\ k = 1, \dots, n_{ij} \geq 1 \end{cases}$$

Test de type III pour l'interaction

C'est le test du sous-modèle

$$\mathcal{M}_o : Y = \mu \mathbf{1}_n + A\alpha + B\beta + \varepsilon$$

contre le modèle :

$$\mathcal{M} : Y = \mu \mathbf{1}_n + A\alpha + B\beta + (A * B)(\alpha * \beta) + \varepsilon$$

Test de l'effet principal α

Tester le sous-modèle

$$\mathcal{M}_o : Y = \mu \mathbf{1}_n + B\beta + (A * B)(\alpha * \beta) + \varepsilon$$

contre le modèle

$$\mathcal{M} : Y = \mu \mathbf{1}_n + A\alpha + B\beta + (A * B)(\alpha * \beta) + \varepsilon$$

n'a pas de sens puisque l'effet principal α est emboîté dans l'interaction $\alpha * \beta$.

→ on teste en fait l'égalité des lsmeans (α) :

$$lsmeans(\alpha_1) = \dots = lsmeans(\alpha_p).$$

EN CONCLUSION

- Seul le test de type III peut-être interprété.
- Pour les effets qui sont à droite dans le diagramme de Hasse, il s'agit d'un modèle contre un sous-modèle.
- Pour les autres effets, le test de type III n'a en fait que peu d'intérêt.
- L'intérêt du test de type I réside dans sa comparaison avec le test de type III :
 - Si les degrés de libertés sont différents, il y a un problème de non-connexité des données ou de confusion partielle des effets. L'analyse doit en tenir compte.
 - Si les degrés de libertés sont égaux mais les Means Squares (MS) très différents, il y a un problème de déséquilibre fort des données.
- si les données sont équilibrées au sens facteurs croisés ou hiérarchiques, type I = type II = type III

BIBLIOGRAPHY



Bailey, R. A. (2000).
Lecture on designs of experiment.
Manuscript non publié.



Benoist, D., Tourbier, Y. & Germain-Tourbier, S.(1994).
Plans d'expériences : construction et analyse.
Lavoisier.



Dagnelie P. (2003).
Principes d'expérimentation : planification des expériences et analyse de leurs résultats.
P Gembloux, Presses agronomiques.



Shah, K. R. & Sinha, B. K. (1989).
Theory of optimal designs.
Lecture Notes in Statistics, 54. Springer-Verlag, New York.