

# L'ANALYSE DE LA COVARIANCE

## (LE POINT DE VUE DU PRATICIEN)

Carlos LOPEZ

Institut de L'Elevage – Service Biométrie  
[carlos.lopez@inst-elevage.asso.fr](mailto:carlos.lopez@inst-elevage.asso.fr)

*L'article présenté est un exemple des notes techniques rédigées à l'attention des zootechniciens responsables de la conception, de la mise en place et de l'analyse des essais en alimentation animale. L'objectif de ce type de document est de présenter de façon simple et la plus intuitive possible à des chercheurs et des ingénieurs utilisateurs de la statistique mais non statisticiens des méthodes qu'ils peuvent être amenés à utiliser ou dont ils peuvent avoir à interpréter les résultats. L'accent est mis plus sur leurs limites d'utilisation que sur les développements statistiques de ces méthodes.*

### 1. Introduction

Commençons tout d'abord par examiner 4 exemples de dispositifs expérimentaux ..

1<sup>er</sup> exemple: Un organisme agricole veut comparer les caractéristiques d'un nouveau régime alimentaire sur l'accroissement pondéral de porcelets par rapport à un régime de référence. Le zootechnicien sait par ailleurs que la prise de poids dépend en partie de poids initial du porcelet. Il souhaite, en conséquence, mesurer finement la part du gain de poids qui est due spécifiquement au poids initial.

2<sup>ème</sup> exemple: Dans une expérience de comparaison des rendements de six variétés de maïs, le technicien constate a posteriori qu'il y a une grande variabilité parcellaire de la densité des plants. Il soupçonne cette hétérogénéité d'être à l'origine des résultats non significatifs qu'il observe sur les rendements moyens variétaux et voudrait donc pouvoir prendre en compte la densité dans ces tests.

3<sup>ème</sup> exemple: Dans une étude sur les effets de 3 types de fertilisation

(A: témoin ; B: Azote ; C: Azote + Phosphore)

sur la croissance en circonférence de peupliers l'expérimentateur veut tenir compte de la compétition entre arbres. Il définit pour cela comme indice de compétition la surface du sol potentiellement disponible pour la croissance de l'arbre, critère qu'il veut utiliser pour estimer la part de la croissance due en propre à la fertilisation.

4<sup>ème</sup> exemple: Une étude porte sur la comparaison de la production de lait de 2 groupes de vaches en début de lactation soumises chacune à un des 2 types d'alimentation suivant:

(A: Ensilage de maïs ; B: Ensilage de maïs et d'herbe)

L'expérimentateur désire également contrôler l'effet du rang de lactation (primipare / multipare) sur la production. Il veut enfin tenir compte de la production pré-expérimentale dans la comparaison des 2 types d'alimentation.

#### **Qu'ont en commun ces 4 dispositifs expérimentaux ?**

Les 4 protocoles ont pour objectif le test de l'effet éventuel d'un facteur, le facteur étudié (que l'on appelle de façon générale le "traitement") sur les moyennes d'une variable d'intérêt, la "réponse". La technique statistique adéquate est ici l'analyse de la variance.

Protocole	Réponse	Traitement
1/	Gain de poids	Régime alimentaire
2/	Rendement	Variété
3/	Circonférence	Apport fertilisant
4/	Production lait	Alimentation

(le 4<sup>ème</sup> protocole est un peu plus complexe dans la mesure où l'on veut tenir compte a priori d'un effet 'rang de lactation').

Mais ces 4 dispositifs ont autre chose en commun .....

Dans les 4 cas l'expérimentateur veut tenir compte de l'effet **spécifique** d'une variable, (une "covariable"), sur la réponse.

Nous allons définir plus loin ce que l'on entend par 'spécifique'.

Ainsi dans le premier protocole l'expérimentateur veut-il tenir compte de la covariable poids initial. Dans la deuxième étude la covariable est la densité de semis. C'est la surface au sol dans le troisième dispositif et la production pré-expérimentale dans l'essai alimentation.

Protocole	Réponse	Traitement	Covariable
1/	Gain de poids	Régime alimentaire	poids initial
2/	Rendement	Variété	densité
3/	Circonférence	Apport fertilisant	surface sol
4/	Production lait	Alimentation	prod. pré-expé

## 2. Quelle conséquence la prise en compte d'une covariable a-t-elle sur l'analyse du dispositif?

Commençons par regarder, en prenant comme exemple le protocole alimentaire sur les porcelets, ce que l'on fait si on n'introduit pas de covariable. (on suppose que le poids initial est mesuré en kg et le gain en grammes). On pose alors le modèle d'analyse de variance suivant :

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad (1)$$

où  $Y_{ij}$  désigne le gain de poids du  $j^{\text{ème}}$  porcelet soumis au  $i^{\text{ème}}$  régime,  $\mu_i$  le gain moyen identique pour tous les porcelets soumis au régime  $i$  et  $\varepsilon_{ij}$  l'écart autour de ce gain moyen spécifique au  $j^{\text{ème}}$  porcelet soumis au régime  $i$ , de variance  $\sigma^2$ .

Comment est modifié le modèle (1) lorsque l'on veut prendre en compte la covariable  $X$  ?

Une troisième composante de  $Y_{ij}$  apparaît alors dans le modèle:

$$Y_{ij} = \mu_i^* + \beta*(X_{ij} - X_{..}) + \varepsilon_{ij}^* \quad (2)$$

Où  $X_{ij}$  est le poids initial (en kg) du porcelet  $ij$ ,

$X_{..}$  est le poids initial moyen sur l'ensemble des 2 régimes,

et  $\beta$  est la pente (en g/kg) de la régression  $Y$  sur  $X$  intra-régime.

•  $\beta*(X_{ij} - X_{..})$  représente la part, dans le gain de poids  $Y_{ij}$ , qui est due au poids initial. Plus précisément un écart de poids de  $(X_{ij} - X_{..})$  kg pour le porcelet  $ij$  induit, du fait de l'incidence du poids initial sur le gain, un accroissement moyen du gain de  $\beta$  [g/kg]\* $(X_{ij} - X_{..})$  [kg].

- Il est important de comprendre que cette part du gain est attribuable au poids initial, elle n'est pas due, en particulier, au régime auquel est soumis le porcelet mais **spécifiquement** au poids. Elle traduit simplement le fait, si par exemple la pente  $\beta$  est positive, que les plus gros porcelets ont tendance à grossir plus en moyenne que les plus petits **indépendamment** du régime qu'on leur donne.
- La composante  $\mu_i^*$  représente le gain moyen du régime  $i$  "corrige" de la part du gain due au poids initial. On l'appelle le gain moyen du régime  $i$  "ajusté" de l'effet du poids initial, (ou "moyenne ajustée du poids"). Elle est différente du terme  $\mu_i$  du modèle (1) qui, on l'a vu, est le gain moyen "brut".
- L'écart  $\varepsilon_{ij}^*$  est également différent de celui du modèle (1) car il est aussi ajusté de l'effet du poids initial. On suppose que, comme pour le modèle (1), la variance des écarts  $\varepsilon_{ij}^*$  ( $\sigma^{*2}$ ), variance des écarts du modèle (2), est constante.

Le test de l'effet régime consiste alors à tester l'égalité des moyennes ajustées  $\mu_i^*$ .

Le modèle (2) est un modèle d'analyse de covariance.

*Bon tout cela c'est très bien mais à quoi cela sert-il de compliquer le modèle (1)?*

N'oublions pas que l'objectif principal de l'expérimentation est de mettre en évidence un effet éventuel du facteur étudié, (par exemple ici une différence significative des gains moyens entre les régimes alimentaires).

En règle générale un expérimentateur cherche toujours à avoir un dispositif le plus puissant possible.

S'il y a, a priori, plusieurs moyens d'augmenter la puissance, accroître le nombre de répétitions, choisir un niveau du risque de première espèce a plus élevé, travailler sur des données plus homogènes c'est-à-dire avoir la variance des écarts la plus petite possible, il n'y en a qu'un seul a posteriori, diminuer la variance résiduelle.

C'est précisément ce qui se passe en analyse de la covariance

Si le poids initial est corrélé significativement au gain de poids, (à régime fixé), alors la variance résiduelle du modèle de covariance (2) est plus petite que la variance résiduelle du modèle (1).

$$\sigma^{*2} < \sigma^2$$

La diminution de la variance des écarts est d'autant plus importante que la corrélation entre le poids (X) et le gain (Y) intra-régime est forte.

**En expérimentation l'analyse de la covariance est principalement une méthode permettant d'accroître la puissance du test du facteur étudié.**

Graphiquement cela se comprend très bien. Illustrons-le à partir d'un exemple caricatural.

Supposons que nos deux groupes de porcelets soient constitués de 3 animaux chacun.

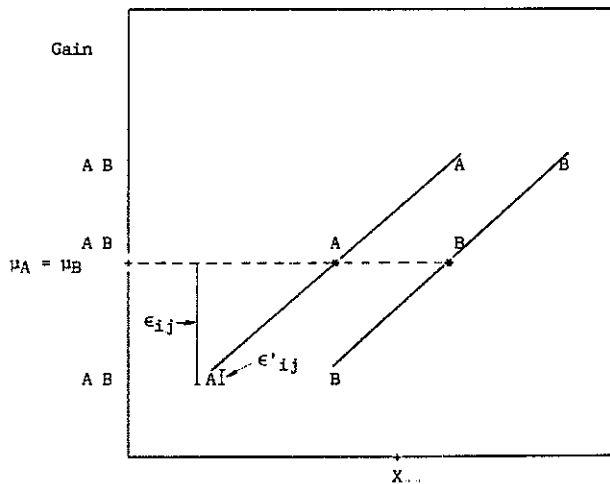
On observe les résultats suivants :

Régimes	A			B		
Gain (kg)	5.0	6.0	6.5	5.0	6.0	6.5
Poids initial. (kg)	30.0	35.0	40.0	35.0	40.0	44.0

Les gains de poids sont rigoureusement identiques pour les 2 régimes. On ne peut absolument pas discriminer les 2 groupes.

Représentons maintenant les gains en fonction du poids initial.

Tout change, on observe maintenant en effet *qu'à poids initial fixé* le régime A donne toujours un gain supérieur à celui du régime B. On a une parfaite discrimination entre les 2 groupes.



L'analyse de la covariance permet de tester rigoureusement cette impression visuelle de discrimination. Elle compare les gains moyens entre les 2 régimes *pour des porcelets ayant un même poids initial* (on dit "à poids initial fixé").

L'écart  $\varepsilon_{ij}^*$  utilisé dans le test est ici l'écart entre le gain  $Y_{ij}$  et le gain moyen prédit par la droite de régression dans le régime n<sup>o</sup>i *pour un poids initial de  $X_{ij}$  kg*. Il est plus petit en moyenne que l'écart  $\varepsilon_{ij}$  du modèle (1) qui est l'écart entre  $Y_{ij}$  et le gain moyen du groupe i.

On voit bien que plus forte est la corrélation entre le gain et le poids initial, plus petit, en moyenne, est l'écart  $\varepsilon_{ij}^*$  par rapport à  $\varepsilon_{ij}$  donc plus petite est la variance  $\sigma^{*2}$  par rapport à  $\sigma^2$  et mieux sont discriminés les groupes.

Le test du facteur "Traitement" est le but principal de l'étude mais signalons également que l'expérimentateur peut être aussi intéressé par l'effet spécifique de la covariable X sur la réponse.

Ainsi le zootechnicien voudra-il savoir quel accroissement moyen du gain peut-il attendre par unité de variation du poids indépendamment du régime donné. La réponse lui est fournie par l'estimation de la pente  $\beta$ .

### 3. Sur les conditions d'utilisation de l'analyse de la covariance

Trois conditions doivent être remplies :

1. Il faut que la **liaison** entre la covariable X et la réponse Y soit **linéaire**.
2. Il ne faut pas qu'il y ait un effet du facteur étudié sur la covariable X.

#### Pas d'effet du 'traitement' sur $X^*$

Car alors les moyennes ajustées peuvent très bien être non interprétables comme on le verra plus loin. Il faut en particulier essayer d'éviter de prendre comme covariable des variables dont la mesure est connue en même temps que celle de la réponse ou, plus généralement, dont les valeurs sont mesurées après l'application des traitements.

Le deuxième protocole est, à ce titre, un bon exemple car si la densité moyenne des plants d'une variété de maïs, *mesurée a posteriori*, est caractéristique de la variété alors les

rendements moyens ajustés de la densité ne sont pas interprétables car ils correspondent à une densité moyenne que n'atteint presque aucune variété.

Une situation où l'on est sûr qu'il ne peut y avoir de relation causale entre le "Traitement" et la covariable X est celle où cette dernière est mesurée avant la randomisation, (on parle alors de variable pré-expérimentale).

3. La liaison entre la réponse (Y) et la covariable (X) ne doit pas dépendre du "Traitement".

Sur notre exemple il faut que la pente de la régression du gain sur le poids initial soit la même d'un régime à l'autre.

On peut dire aussi, de façon symétrique, que la différence de gain entre les deux régimes est la même quelque soit le poids initial.

autrement dit:

**Il ne faut pas qu'il y ait d'interaction entre le 'Traitement' et la covariable (X) sur la réponse (Y).**

### **Comment vérifier ces conditions?**

1. Linéarité de la liaison Y / X

Une représentation des nuages de points (X, Y) pour chaque niveau du facteur étudié avec vérification visuelle de la linéarité peut souvent suffire.

On peut si on a des relations paraboliques ou asymptotiques marquées procéder à des transformations de la covariable X. Faisons remarquer toutefois que, dans ce cas et si on soupçonne a priori ce type de relations, une prise en compte de la covariable par la définition de blocs sur les niveaux de celle-ci aurait été une démarche plus judicieuse.



## 2. Absence d'effet "Traitement" sur la covariable X

Il suffit de tester l'effet du facteur "Traitement" sur la covariable X par une analyse de variance, (dans l'exemple des régimes sur porcelets on compare les deux poids initiaux moyens des 2 régimes).

Si l'effet est significatif alors que la covariable est "post-expérimentale" l'analyse de la covariance est déconseillée, si la covariable est "pré-expérimentale" alors il n'aurait pas fallu commencer l'essai dans ces conditions (d'où l'intérêt de faire cette analyse de variance avant de démarrer l'essai).

## 3. Absence d'interaction "Traitement"\*X sur la réponse

On peut tester le parallélisme des droites rigoureusement. On peut aussi se contenter d'une représentation des droites de régression et vérifier visuellement le parallélisme des droites.

Si on ne peut considérer les droites comme parallèles alors l'interprétation d'un effet du facteur étudié (à covariable X fixée) devient plus difficile puisque cet effet dépend du niveau de la covariable. Ainsi, dans l'exemple des porcelets, la différence de gain moyen entre les deux régimes dépendrait du poids initial des porcelets.

## 4. Notions de moyennes ajustées

Une notion importante dans l'analyse des dispositifs expérimentaux est celle de moyenne ajustée. Nous l'avons définie comme étant la quantité  $\mu_i^*$  du modèle (2) d'analyse de covariance. C'est le gain moyen des porcelets soumis au régime i "corrige" de l'effet du poids initial.

Essayons d'expliquer cette notion.

Si on ne tient pas compte du poids initial des porcelets, c'est-à-dire si l'on analyse le modèle (1), le test de l'effet régime se fait en comparant les gains moyens :  $Y_1$   $Y_2$  observés sur les deux groupes de porcelets.

Mais les porcelets soumis au régime  $i$  n'ont pas tous le même poids initial, appelons  $X_i$  la moyenne de ces poids. Si le gain dépend du poids initial alors on peut déduire du modèle (2) la part dans le gain moyen  $Y_i$  observé dans le groupe  $i$  qui est due au régime  $i$  et celle qui est due au poids initial.

Plus précisément, on a la relation suivante:

$$Y_i = \mu_i^* + \beta^*(X_i - X_{..}) + \varepsilon_i^* \quad (3)$$

où  $\mu_i^*$  est le gain moyen dans le groupe  $i$  ajusté du poids initial,

$\beta^*(X_i - X_{..})$  est la part de  $Y_i$  due au poids initial moyen  $X_i$  du groupe,

$\varepsilon_i^*$  est la moyenne des écarts  $\varepsilon_{ij}^*$

On en déduit immédiatement une estimation du gain moyen ajusté du poids:

$$\hat{Y}_i = Y_i - b^*(X_i - X_{..}) \quad (4)$$

où  $b$  est l'estimation de la pente  $\beta$ .

L'interprétation de la formule (3) est alors claire.

Supposons que la liaison entre le poids initial et le gain de poids soit positive, (ce qui est plausible), c'est à dire que les porcelets plus gros que la moyenne sur l'ensemble des 2 régimes ( $X_{..}$ ) ont des gains en moyenne plus importants que les porcelets les moins gros.

Alors si les porcelets du groupe  $i$  sont plus gros que la moyenne, (donc d'une quantité  $(X_i - X_{..})$ ), leur gain moyen est sur-estimé d'une quantité (positive) égale à  $\beta^*(X_i - X_{..})$  et donc pour avoir le gain de poids "corrigé" de l'excès de poids  $(X_i - X_{..})$  il faut retrancher au gain moyen observé  $Y_i$  cette quantité  $\beta^*(X_i - X_{..})$ .

Comme la pente  $\beta$  est inconnue on l'estime à partir des données, (soit  $b$  son estimation), on obtient alors l'estimation du gain ajusté donné en (4).

On voit bien comment joue la correction quand on prend en compte le poids

$X_i > X_{..}$	$\Rightarrow$	$\hat{Y}_i < Y_i$
Porcelets trop gros		ajusté < observé

De la même façon si le groupe  $i$  est constitué de porcelets moins gros que la moyenne, ( $X_i < X_{..}$ ) alors leur gain de poids est sous-estimé d'une quantité  $\beta \cdot (X_i - X_{..})$ . On obtient une estimation du gain de poids corrigé du poids initial en retranchant  $\beta \cdot (X_i - X_{..})$  au gain moyen observé.

$X_i < X_{..}$	$\Rightarrow$	$\hat{Y}_i > Y_i$
Porcelets moins gros		ajusté > observé

Reprenons l'exemple caricatural du paragraphe 2.

Les moyennes brutes observées sont rigoureusement égales

$$Y_{1..} = Y_{2..} = 5.8 \text{ kg}$$

alors que les porcelets du régime A sont moins gros que ceux du régime B:

A: $X_{1..} = 35 \text{ kg}$	B: $X_{2..} = 39.7 \text{ kg}$	(A+B): $X_{..} = 37.3 \text{ kg}$
------------------------------	--------------------------------	-----------------------------------

Les porcelets de A sont donc a priori handicapés dans leur possibilité de gain et ce indépendamment du fait qu'on leur donne le régime A, (avec B cela serait pareil).

La pente de la régression "Gain / Poids" intra-régime est environ de 0.16 (kg de gain par kg de poids initial). Le "manque à gagner" du groupe A par rapport à un poids initial moyen général de 37.3 kg est donc de  $0.16 \cdot (35 - 37.3) = -0.4 \text{ kg}$

Le gain moyen du groupe A corrigé de l'effet du poids est alors:  $5.8 - (-0.4)$  soit  $6.2$  kg

On "recale" en quelque sorte les porcelets soumis au régime A à un niveau de poids moyen égal au poids moyen général sur l'ensemble des 2 régimes.

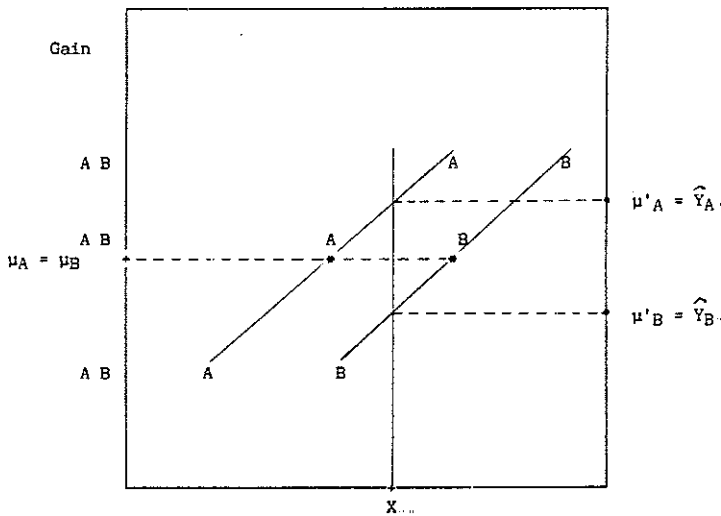
De la même façon le gain moyen du groupe B est sur-estimé d'une quantité égale à  $0.4$  kg ( $0.16 \cdot (39.7 - 37.3)$ ) due au sur-poids initial des porcelets B.

Le gain moyen du groupe B corrigé de l'effet du poids initial est donc :  $5.8 \text{ kg} - 0.4 \text{ kg} = 5.4 \text{ kg}$

On retrouve quantitativement ce que l'on avait observé graphiquement:

Le régime A donne un meilleur gain moyen que le régime B à **poids initial fixé** :

$$A : \hat{Y}_1 = 6.2 \text{ kg} \quad B : \hat{Y}_2 = 5.4 \text{ kg}$$



### *Des moyennes ajustées non interprétables...*

Ces corrections des moyennes est la seconde propriété de l'analyse de la covariance après la réduction de la variance résiduelle. C'est une qualité secondaire, la justification de l'utilisation de la méthode doit rester, dans le contexte expérimental, la réduction de la variance résiduelle.

On remarque d'ailleurs que cette correction ne joue pas lorsque la moyenne de la covariable est identique dans les groupes:

$$X_1 = X_2 = X \implies \hat{Y}_i = Y_i$$

C'est complètement en accord avec l'interprétation que l'on a donné plus haut de ces moyennes ajustées.

On remarquera le rôle joué par le poids initial moyen général  $X_{..} = 37.3$  kg. La correction (4) ramène en fait tous les groupes à un même niveau de la covariable X: le niveau moyen général  $X_{..}$ .

Ceci est fondamental car c'est ce qui explique que parfois les moyennes ajustées peuvent n'avoir aucun sens.

Revenons pour illustrer cela sur l'expérimentation Maïs.

L'expérimentateur a constaté, a posteriori, une forte hétérogénéité sur les densités moyennes de plants entre variétés. La correction sur les rendements moyens ramène ceux-ci à un niveau de densité correspondant à la densité moyenne générale sur l'ensemble des 6 variétés de maïs.

Les rendements moyens ajustés peuvent alors ne correspondre à rien si la densité est une caractéristique variétale car la quasi-totalité sinon la totalité des variétés peut très bien ne jamais avoir cette densité moyenne.

On peut même obtenir des résultats complètement incohérents avec un classement des moyennes ajustées inverse de celui des moyennes brutes !!

## 5. Modèles plus compliqués

La prise en compte d'une covariable a été présentée jusqu'à présent dans le cas d'une expérimentation particulièrement simple, le dispositif expérimental à un facteur étudié. En fait le problème de l'ajustement sur une covariable est un problème très général qui peut se poser pour des dispositifs plus compliqués. Si le principe de la démarche reste le même l'analyse statistique et son interprétation peuvent être plus ardues.

On peut ainsi être amené à prendre en compte une caractéristique ayant une incidence sur la réponse analysée, la race sur la croissance d'animaux par exemple ou la taille de la portée dans le cas de l'étude sur les porcelets.

le modèle se complique un peu avec l'introduction dans le modèle (2) d'un terme supplémentaire  $\pi_j$  correspondant au nouveau facteur étudié ou contrôlé :

$$Y_{ij} = \mu + \alpha_i + \pi_j + \beta_*(X_{ij} - X_{..}) + \varepsilon_{ij} \quad (5)$$

où  $\mu$  représente le gain moyen général sur l'ensemble des I "Traitements" et des J modalités du second facteur introduit dans le modèle,

$\alpha_i$  représente l'écart de gain par rapport à  $\mu$  imputable au régime i,

$\pi_j$  représente l'écart de gain par rapport à  $\mu$  imputable à la modalité j second facteur,

$\beta_*(X_{ij} - X_{..})$  est toujours l'écart de gain par rapport à  $\mu$  dû à l'écart de poids ( $X_{i.} - X_{..}$ ),

L'analyse, plus compliquée au niveau des procédures de calculs, reste dans son interprétation fondamentalement la même avec les mêmes contraintes.

## 6. Méthodes alternatives: index, blocs

Une méthode assez couramment utilisée et facile d'emploi est la méthode des **index**. Elle correspond dans son esprit à un ajustement comme le fait l'analyse de la covariance mais sur une transformation de la réponse établie a priori.

On définit à partir de la réponse Y et de la covariable X une fonction - réponse (l'index I) sur lequel sera testé l'effet du facteur étudié.

$$I = f(X, Y) \quad (6)$$

Ainsi dans l'étude sur porcelets l'expérimentateur peut-il être intéressé par le gain relatif:

$$I = Y / X$$

On peut imaginer des fonctions - réponses plus compliquées. Le point important est que l'index ait une interprétation agronomique, zootechnique, ... claire. La liaison posée ainsi **a priori** entre Y et X n'est peut être pas forcément la meilleure (c'est-à-dire celle qui décrit le mieux les données de l'expérimentation) mais ce n'est pas grave, c'est son interprétation qui prime.

Dans l'étude n° 4 sur la production laitière on pourrait de même définir l'index : "variation de la production par rapport à la lactation précédente",  $I = Y - X$

et poser le modèle d'analyse de variance suivant sur I :

$$I_{ij} = \mu + \alpha_i + \pi_j + \varepsilon_{ij} \quad (7)$$

où  $\alpha_i$  représente l'effet de l'alimentation et  $\pi_j$  représente l'effet du rang de lactation sur la variation de la production.

Cet index prend comme modèle de variation de Y en fonction de X un modèle linéaire de pente  $\beta=1$

$$Y_{ij} = I_{ij} + 1 \cdot X_{ij} \quad (8)$$

Faire une analyse de variance sur I c'est "presque comme" faire une analyse de covariance sur Y en supposant connue et égale à 1 la pente de la régression de Y sur X.

Il n'y a, là non plus, aucune raison que la "bonne" pente soit égale à 1, le modèle (8) n'est probablement pas le meilleur modèle mais, par contre, il présente deux avantages:

- $I = Y - X$  a une interprétation zootechnique plus claire que  $(Y - b \cdot X)$ ,
- La pente 1 étant supposée connue, on ne perd pas un degré de liberté à l'estimer.

La principale méthode alternative à l'analyse de la covariance est le **dispositif en blocs complets de Fisher** (les blocs étant construits à partir d'un certain nombre de facteurs et de

covariables dont on veut contrôler les effets afin d'accroître la puissance du dispositif, race, caractéristiques génétiques, poids pré-expérimentaux, ...).

Cette méthode impose évidemment, si l'on veut que la covariable X soit prise en compte dans la construction des blocs, de connaître les valeurs de X en début d'expérimentation. Elle est donc un peu plus contraignante au niveau du dispositif à mettre en place.

En contrepartie elle est utilisable quel que soit le type de liaison entre Y et X (linéaire ou non linéaire) et les calculs sont plus simples. La méthode des blocs possède en outre un avantage qui peut s'avérer décisif notamment lorsque les variables réponses sont issues d'une évaluation subjective (évaluation visuelle d'une hauteur d'herbe dans une parcelle, de l'état d'engraissement d'un animal, ...). Une éventuelle dérive de la mesure n'induit pas de biais si les évaluations sont faites bloc par bloc, elle est absorbée dans l'effet bloc.

## 7. Alors pratiquement que faire ?

- Si la covariable est mesurée a posteriori <sup>(1)</sup>, (comme par exemple les densités de plants), l'alternative est ici entre la méthode des index et l'analyse de la covariance. C'est alors au responsable de l'expérimentation de choisir la méthode d'analyse en gardant à l'esprit les qualités et les limites de chacune d'elles notamment au niveau de l'interprétation des moyennes ajustées.

- Si la covariable est mesurée a priori, (par exemple le poids initial des porcelets), le choix se fera entre l'analyse de la covariance (ou éventuellement les index) et l'analyse en blocs de Fisher suivant le type de liaison entre Y et X.

- Si la liaison Y / X est linéaire "forte" (c'est-à-dire une corrélation de l'ordre de 0.60 au moins) on peut utiliser l'analyse de la covariance,

- Si la liaison est non linéaire ou linéaire mais avec une corrélation 'moyenne' il est préférable de poser un modèle en blocs.

Remarque importante : En règle générale et quel que soit le modèle posé (bloc, covariance ou index), il faut **éviter de prendre en compte des covariables qui sont faiblement liées**



à la réponse. On ne diminue pas significativement les écarts résiduels tout en perdant des degrés de liberté pour estimer les paramètres des facteurs et covariables introduits. On peut même avoir parfois des variances résiduelles plus grandes.

Là comme ailleurs il faut agir avec discernement, essayer d'avoir des modèles les plus simples possibles en utilisant toute l'information pertinente et ..... uniquement l'information pertinente.

---

*(1) : mesurée a posteriori n'implique pas nécessairement qu'il existe une relation causale entre la covariable X et le facteur étudié; ainsi si l'on veut étudier l'effet de deux traitements A et B sur la réponse Y d'un système enzymatique chez le rat on peut prendre en compte le poids du foie de l'animal si la réponse est liée à ce critère. La valeur du poids ne peut évidemment pas être connue avant le sacrifice de l'animal et il n'y a aucune raison que les poids moyens des foies dans les deux groupes soient liés au traitement.*

