Examen partiel

14 décembre 2005
***
Corrigé

Exercice 1. Analyse sensorielle de trois chocolats

1. Le modèle s'écrit, en notant $y_{i,j,k}$ la $k$ème note obtenue le jour j pour le chocolat i :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \alpha\beta_{i,j} + \epsilon_{i,j,k}, \quad i = 1\dots 3,\ j = 1\dots 2,\ k = 1\dots k(i,j)$$

   avec les hypothèses suivantes pour les résidus :

   $\forall\ (i,j,k)\ \mathcal{L}(\epsilon_{i,j,k} = \mathcal{N}(0,\sigma))$ et $Cov(\epsilon_{i,j,k}, \epsilon_{r,s,t}) = 0$ si $(i,j,k) \neq (r,s,t)$

   Une interaction *Jour* $\times$ *Chocolat* signifie qu'un même chocolat n'est pas apprécié de la même façon les 2 jours.
2. Le vecteur des paramètres peut être estimé par moindres carrés.
3. On procède à l'ANOVA (en stockant les résidus afin de pouvoir vérifier s'ils vérifient les hypothèses) à l'aide de la fonction modèle linéaire généralisé et on obtient les résultats suivant :

```
Modèle linéaire généralisé : note en fonction de jour; chocolat

Facteur      Type Niveaux  Valeurs
jour         fixe       2  1 2
chocolat     fixe       3  1 2 3

Analyse de la variance pour note

Source          DL     SC séq   SC ajust   CM ajust       F      P
jour             1    26,3511    25,6301    25,6301   84,45  0,000
chocolat         2     1,3894     1,2021     0,6010    1,98  0,152
jour*chocolat    2     0,7662     0,7662     0,3831    1,26  0,294
Erreur          39    11,8364    11,8364     0,3035
Total           44    40,3431
```

Le test de l'homoscédasticité des résidus à l'aide d'un test de Levenne n'est pas significatif :

```
Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 0,613
P :                   0,691
```

Le test de normalité à l'aide d'un test de Shapiro-Wilk (car N=45) n'est pas significatif :

```
Valeur de P (approximatif) : > 0,1000
R:                             0,9885
W-test pour la normalité
```

On va donc faire également un test de Bartlet (plus puissant que Levenne à cause de la normalité) :

```
Test de Bartlett (loi normale)

Statistique du test : 2,643
P :                   0,755
```

Ce test n'est pas significatif ce qui confirme bien l'homoscédasticité des résidus.

L'étude du protocole expérimental nous assure de l'indépendance des variables $\epsilon_{i,j,k}$.

Puisque les conditions d'application de l'ANOVA sont bien respectées, on peut déduire des résultats reportés dans le tableau ci-dessus qu'en considérant le modèle introduit à la première question il n'y a, au seuil $\alpha = 5$ %, ni effet *chocolat* $(0,152 > 0,05)$ ni effet de l'interaction $(0,294 > 0,05)$.

4. La moyenne des notes est de $4,676$. Les moyennes par chocolat sont :

Variable	chocolat	N	Moyenne	EcarType
note	1	15	4,507	0,982
	2	15	4,507	0,807
	3	15	5,013	1,041

Pour obtenir les moyennes ajustées $(\hat{\mu} + \hat{\alpha}_1, \hat{\mu} + \hat{\alpha}_2, \hat{\mu} + \hat{\alpha}_3)$ on fait appel à la technique apprise dans le cours, basée sur la régression linéaire multiple, afin d'estimer les coefficients $\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$. Cette estimation des paramètres est encore obtenue à l'aide du modèle linéaire généralisé (pensez à afficher les résultats les plus détaillés). Les calculs renvoient les résultats ci-dessous :

Terme	Coef	Er-T coef	T	P
Constante	4,58763	0,08801	52,13	0,000
jour				
1	0,80874	0,08801	9,19	0,000
chocolat				
1	0,2124	0,1238	1,72	0,094
2	-0,2043	0,1215	-1,68	0,101
jour*chocolat				
1 1	0,0713	0,1238	0,58	0,568
1 2	-0,1921	0,1215	-1,58	0,122

On en déduit que :

$$\hat{\mu} = 4,588, \hat{\alpha}_1 = 0,212, \hat{\alpha}_2 = -0,204, \hat{\alpha}_3 = -0,008$$

et que les moyennes ajustées valent :

$$\hat{\mu} + \hat{\alpha}_1 = 4,800, \hat{\mu} + \hat{\alpha}_2 = 4,384, \hat{\mu} + \hat{\alpha}_3 = 4,580$$

Par conséquent c'est le chocolat 1 qui est le préféré. Le déséquilibre des données fait que la moyenne d'un chocolat n'est plus égale à sa moyenne ajustée. Ainsi contrairement à ce que la moyenne des notes semblaient indiquer ce n'est pas le chocolat 3 qui est le préféré.

5. $H_0 \ : \ \alpha_1 = 0$ contre $H_1 \ : \ \alpha_1 \neq 0$. La statistique du test est $\frac{\hat{\alpha}_1}{\hat{\sigma}_{\hat{\alpha}_1}}$ et suit, sous l'hypothèse nulle, une loi de Student avec autant de degrés de libertés que ceux de la variance résiduelle (i.e. 39). La probabilité critique associée à cette valeur est de $0,094$, ce chocolat n'est donc pas significativement plus apprécié que les autres.

Exercice 2. Rendemment de blé

1. Le modèle s'écrit, en notant $y_{i,j,k,l}$ le rendement de la $l$ème parcelle cultivée avec l'irrigation i, le fongicide j et la dose d'engrais k :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{i,j} + \alpha\gamma_{i,k} + \beta\gamma_{j,k} + \epsilon_{i,j,k,l}$$
$$\text{avec } i = 1\ldots 2,\ j = 1\ldots 2,\ k = 1\ldots 3,\ l = 1\ldots 2$$

avec les hypothèses suivantes pour les résidus :

$$\forall\ (i,j,k,l)\ \mathcal{L}(\epsilon_{i,j,k,l} = \mathcal{N}(0,\sigma)) \text{ et } Cov(\epsilon_{i,j,k,l}, \epsilon_{r,s,t,u}) = 0 \text{ si } (i,j,k,l) \neq (r,s,t,u)$$

2. Les contraintes imposées par le modèle donnent les relations suivantes :

$$\sum_{i=1}^{I} \hat{\alpha}_i = 0 \text{ donc } \hat{\alpha}_{\text{avec irrigation}} = -\hat{\alpha}_{\text{sans irrigation}} = 4,8922$$

$$\sum_{j=1}^{J} \beta_j = 0 \text{ donc } \hat{\beta}_{\text{avec fongicide}} = -\hat{\alpha}_{\text{sans fongicide}} = 3,8917$$

$$\sum_{k=1}^{K} \gamma_k = 0 \text{ donc } \hat{\gamma}_{\text{engrais important}} = -\hat{\gamma}_{\text{engrais moyen}} - \hat{\gamma}_{\text{engrais faible}} = 11,2751$$

$$\forall i \ \sum_{k=1}^{K} \widehat{\alpha\gamma}_{i,k} = 0 \text{ donc } \widehat{\alpha\gamma}_{\text{ss irr,eng imp}} = -\widehat{\alpha\gamma}_{\text{ss irr,eng faible}} - \widehat{\alpha\gamma}_{\text{ss irr,eng moy}} = 0,7960$$

$$\forall k \ \sum_{i=1}^{I} \widehat{\alpha\gamma}_{i,k} = 0 \quad \text{donc} \quad \widehat{\alpha\gamma}_{\text{ss irr,eng imp}} = -\widehat{\alpha\gamma}_{\text{av irr,eng imp}} = -0,7960$$

$$\widehat{\alpha\gamma}_{\text{ss irr,eng moy}} = -\widehat{\alpha\gamma}_{\text{av irr,eng moy}} = 0,9467$$
$$\widehat{\alpha\gamma}_{\text{ss irr,eng faib}} = -\widehat{\alpha\gamma}_{\text{av irr,eng faib}} = -0,1507$$

Les autres coefficients étaient renseignés dans le tableau fourni avec l'énoncé.

3. Test de significativité de l'interaction *irrigation* $\times$ *engrais* :

$$H_0 \ : \ \forall (i,k), \alpha\gamma_{i,k} = 0 \qquad H_1 \ : \ \forall (i,k), \alpha\gamma_{i,k} \neq 0$$

La statistique de test utilisée est :

$$F = \frac{CM_{\text{interaction}}}{CM_{\text{résiduelle}}}$$

Ici $ddl_{\text{interaction}} = (I-1)(K-1) = 2$ et $ddl_{\text{résiduelle}} = n - 1 - (I-1) - (J-1) - (K-1) - (I-1)(K-1) = 24 - 1 - 1 - 2 - 1 - 2 = 17$. Sous l'hypothèse nulle et une hypothèse de normalité des données, F suit une loi de Fisher à 2 et 17 degrés de liberté. Décision : soit on compare f_{obs} avec le quantile $1-\alpha$ de la loi de Fisher à 2 et 17 degrés de liberté, soit on compare la $p-$valeur avec le seuil α du test. Ici la $p-$valeur vaut $0,7535$, on accepte l'hypothèse nulle au seuil de $\alpha = 5$ %. On considère donc qu'il n'y a pas d'interaction entre le facteur *irrigation* et le facteur *engrais*.

4. La probabilité critique du test dur l'effet *engrais* est inférieure à $0,0005$, on en déduit que le test est significatif : il y a un effet engrais. On se demande si le coefficient $\gamma_{\text{eng imp}}$ est significativement différent de 0 au seuil $\alpha = 5$ %. Pour cela on teste l'hypothèse $H_0 \ : \gamma_{\text{eng imp}} = 0 \qquad H_1 \ : \gamma_{\text{eng imp}} \neq 0$. La statistique du test est $\dfrac{\hat{\gamma}_{\text{eng imp}}}{\hat{\sigma}_{\hat{\gamma}_{\text{eng imp}}}}$ et suit, sous l'hypothèse nulle, une loi de Student avec autant de degrés de libertés que ceux de la variance résiduelle (i.e. 17). La probabilité critique associée à cette valeur est $< 0,0005$ donc $< \alpha = 0,05$. Par conséquent les rendements observés pour un apport important d'engrais sont significativement supérieurs (en moyenne de 11,2751) au rendemment moyen.

5. Le modèle retenu est le modèle sans interaction :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{i,j,k,l}, \quad i = 1\ldots 2, \ j = 1\ldots 2, \ k = 1\ldots 3, \ l = 1\ldots 2$$

avec les hypothèses suivantes pour les résidus :

$$\forall \ (i,j,k,l) \ \mathcal{L}(\epsilon_{i,j,k,l} = \mathcal{N}(0,\sigma)) \text{ et } Cov(\epsilon_{i,j,k,l}, \epsilon_{r,s,t,u}) = 0 \text{ si } (i,j,k,l) \neq (r,s,t,u)$$

Les estimations des paramètres des effets principaux sont identiques dans le modèle avec ou sans interaction car les données sont équilibrées (cf cours sur les plans d'expérience).

$$\hat{Y}_{\text{av irr,av fon,eng imp}} = \hat{\mu} + \hat{\alpha}_{\text{av irr}} + \hat{\beta}_{\text{av fon}} + \hat{\gamma}_{\text{eng imp}} = 88,68 \text{ qx/ha.}$$