

GM – PROBABILITÉS ET STATISTIQUE – SÉRIE 11

En salle

Exercice 1 Une chercheuse s'intéresse aux mérites relatifs des régimes par rapport aux médicaments hypocholestérolémiant. Pour chacun des 65 sujets qui ont commencé l'étude avec un cholestérol élevé, elle enregistre le niveau total de cholestérol sanguin (en mg par décilitre) après 6 mois de participation à l'étude. Les patients sont répartis en 5 groupes : un groupe de contrôle (C) qui reçoit un placebo, un groupe de régime végétarien (V), un groupe de régime pauvre en graisses (PG), un groupe à faible dose des médicaments (DF) et un groupe avec une dose élevée de médicaments (DE). Les résumés des données :

Groupe	Moyenne	Écart-type	n_{groupe}
C	240	1.22	25
V	225	1.18	10
PG	230	1.10	10
DF	215	1.02	10
DE	200	1.11	10

Tableau d'ANOVA :

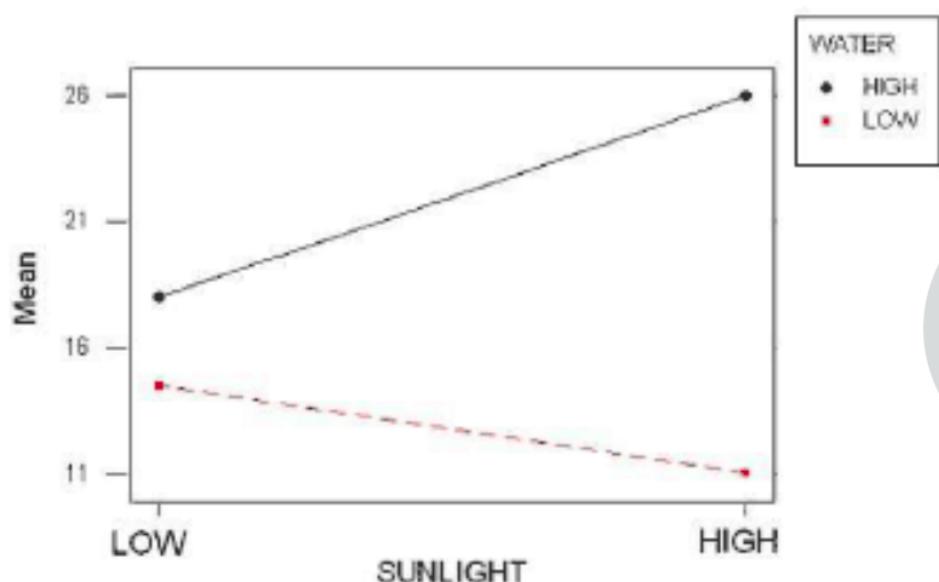
Source	df	SC	CM	F	p
Traitements					4.8×10^{-11}
Erreur	60		80		
Total		11681.4			

- Quelles suppositions doivent être remplies pour obtenir une p -valeur valide en utilisant ANOVA ?
- Compléter le tableau.
- Sur la base de ces données, existe-t-il des preuves qu'au moins l'une des moyennes des groupes diffère(nt) ? Justifier votre réponse en effectuant un test d'hypothèses approprié. Énoncer les hypothèses nulle et alternative. Quelle est votre conclusion, si $\alpha = 0.01$? Interpréter le résultat.
- Quel pourcentage de la variabilité est expliqué par la différence entre les traitements ?
- Calculer les statistiques de test t_{obs} pour les 10 tests de comparaison (chaque paire).
- Quelles paires de moyennes sont différentes au niveau (nominal - c'est-à-dire sans ajustement) de signification $\alpha = 0,01$?
- Selon la méthode Bonferroni, trouver le seuil pour chaque test si vous voulez un seuil global de $\alpha = 0,01$. Quelles paires de moyennes sont différentes au niveau de signification (global) $\alpha = 0,01$?

Exercice 2 La quantité de soleil et l'arrosage influent-ils sur la croissance des géraniums ? La croissance des plantes de 16 plantes est mesurée en centimètres. Chaque combinaison de lumière du soleil et d'eau (élevé/faible) comporte 4 plantes ; p. ex., la combinaison d'eau à haut niveau avec un ensoleillement élevé comprend 4 plantes d'une longueur de 21 à 30 cm. Le tableau d'ANOVA (partiellement complet) se trouve ci-dessous, ainsi qu'un graphique d'interaction.

Source	df	SS	MS	F	$P(> F)$
Water	(a)	342.2	(g)	(k)	0.000365 ***
Sunlight	(b)	20.2	(h)	(l)	0.256272
(n)	(c)	132.2	(i)	(m)	0.010152 *
Error	(d)	(f)	(j)		
Total	(e)	665.6			

Interaction Plot - LS Means for YIELD



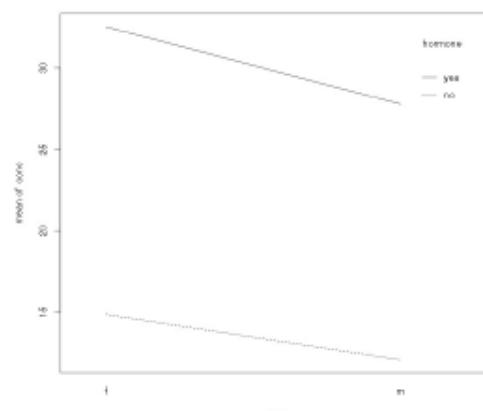
- (a)-(n) Compléter le tableau.
- (o) Écrire le modèle **complete** (théorique), y compris toutes les suppositions du modèle.
- (p) La quantité (niveau) d'arrosage affecte-t-elle la croissance des géraniums en pot ? Réaliser le test d'hypothèse pertinent en veillant à rédiger chacune des 5 étapes, et *interpréter vos résultats* (utiliser $\alpha = 0,05$).
- (q) La quantité (niveau) de lumière solaire affecte-t-elle la croissance des géraniums en pot ? Réaliser le test d'hypothèse pertinent en veillant à rédiger chacune des 5 étapes, et *interpréter vos résultats* (utiliser $\alpha = 0,05$).
- (r) L'effet du niveau d'ensoleillement dépend-il du niveau d'arrosage ? Réaliser le test d'hypothèse pertinent en veillant à rédiger chacune des 5 étapes, et *interpréter vos résultats* (utiliser $\alpha = 0,05$).
- (s) Est-il judicieux de supprimer la variable « sunlight » du modèle ? **Expliquer.**

À domicile

Exercice 1 Nous voulons tester si la concentration plasmatique de calcium (mg/100 ml) des oiseaux mâles et femelles est affectée par le traitement hormonal. Les sorties de R de l'expérience sont données ci-dessous. Vous pouvez supposer que le design est équilibrée.

```
interaction.plot(sex,hormone,conc)
```

```
Df Sum Sq Mean Sq F value    Pr(>F)
sex        1   70.3   70.3   3.733   0.0713 .
hormone    1 1386.1 1386.1  73.585 2.22e-07 ***
sex:hormone 1   4.9    4.9   0.260   0.6170
Residuals  16  301.4   18.8
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1
```



- (a) Combien d'oiseaux y a-t-il dans l'étude ? Combien de mâles ? Combien de femelles ?
- (b) Combien de traitements hormonaux existe-t-il ?
- (c) Combien d'oiseaux y a-t-il dans chaque groupe d'hormone \times sexe \times ?
- (d) Combien de termes d'interaction peuvent être estimés ? Comment vous le savez ?
- (e) Écrire le modèle **complète** (théorique), y compris toutes les suppositions du modèle.
- (f) Existe-t-il une preuve graphique d'un effet d'interaction ? **Expliquer**.
- (g) Écrire les 3 tests d'hypothèses possibles qui peuvent être effectués dans cette expérience. Quelles hypothèses NULLES sont REJETÉES au niveau $\alpha = 10\%$?
- (h) En utilisant les résultats de la partie (g), écrire le modèle final (théorique).

Exercice 2 Cet exemple est tiré d'une déclaration de Texaco, Inc. au sous-comité de la pollution de l'air et de l'eau du comité des travaux publics du Sénat américain, le 26 juin 1973. M. John McKinley, président de Texaco, a cité un filtre automobile développé par Associated Octel Company comme étant efficace pour réduire la pollution.

Cependant, des questions ont été soulevées concernant les effets des filtres sur les performances des véhicules, la consommation de carburant, la contre-pression des gaz d'échappement et le silencieux. Concernant la dernière question, il a fait référence à un jeu de données prouvant que les propriétés de silencieux du filtre Octel étaient au moins égales à celles des silencieux standard.

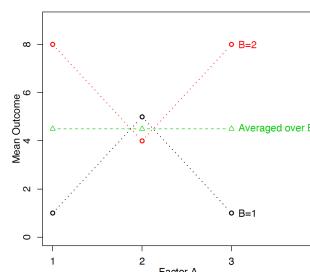
Il s'agit d'une expérience dans laquelle le traitement *type de filtre* avec les niveaux *standard* et *octel* est attribué aléatoirement aux unités expérimentales (voitures). Trois types de voitures sont utilisés : petite, moyenne et grande. Le résultat est la variable quantitative (continue) *bruit*.

Six voitures sont testées pour chaque combinaison *taille (size)* \times *type de filtre*. Un tableau d'ANOVA partiel est présenté ci-dessous :

Source	df	SS	MS	F	p
Taille			13026		< 2e-16 ***
Filtre		1056			0,00373 ***
		804			0,000158 ***
Erreur			65		
		29874			

- (a) Compléter le tableau.
- (b) Un modèle d'interaction ou un modèle additif serait-il plus approprié dans cette situation ? Expliquer.
- (c) Quel pourcentage de la variabilité totale est expliqué par le modèle choisi dans la partie (b) ?

Exercice 3 Un exemple typique d'interaction statistiquement significative avec des effets PRINCIPAUX statistiquement NON significatifs est celui où, par exemple, il y a trois niveaux du facteur *A* et deux niveaux du facteur *B*, et le schéma des effets des variations du facteur *A* est tel que les moyennes sont en forme de V pour un niveau de *B* et en forme de V inversé pour l'autre niveau de *B*. Le test de l'effet principal pour *A* consiste à vérifier si les résultats moyens aux trois niveaux de *A*, calculés sur les deux niveaux de *B*, sont équivalents :



Expliquer pourquoi la non-significativité des effets principaux est trompeuse ici.