

Travaux Pratiques Analyse Statistique et R 01

Analyses de la variance (ANOVA) sur R

Alison PATOU

patou.alison@gmail.com

7/02/2020

Introduction

Dans ce TP nous allons effectuer des tests d'analyse de la variance (ANOVA) à un ou deux facteurs.

Exercice 1

Info: Nous travaillons sur le **traitement_veaux**

Le transport d'animaux d'élevage implique une succession de manipulations et de confinements qui, inévitablement, sont responsables de stress.

L'objectif de cette étude est d'examiner le comportement des animaux lors de transports de longue durée. Pour mesurer le niveau de stress, nous avons calculé le pourcentage de temps que les animaux passent couchés. La fatigue des animaux (et donc le stress) est d'autant plus grande que le temps passé couché est important.

L'expérimentation a consisté à observer 18 veaux transportés de France en Italie, la durée du voyage étant de 29 heures.

Trois traitements ont été proposés durant la pause :

- Traitement 1 : les veaux ne reçoivent ni eau ni aliment à la pause.
- Traitement 2 : les veaux reçoivent de l'eau et sont alimentés à la pause par deux abreuvoirs.
- Traitement 3 : les veaux reçoivent de l'eau et sont alimentés à la pause par cinq abreuvoirs.

Pour enregistrer les comportements des animaux, la bétailière est équipée de cameras et chaque animal est individualisé par un signe distinctif sur la peau. Les cassettes vidéo ont été dépouillées à l'aide d'un logiciel. Les données sont disponibles dans le tableau ci-dessous.

Traitement 1	17,40	20,00	26,70	31,70	35,80	47,80
Traitement 2	14,65	37,22	37,73	43,61	46,07	47,40
Traitement 3	18,76	19,49	27,19	45,42	53,20	61,27

Question 1

- Charger le jeu de données traitement_veaux
Déterminer la structure du dataset

Question 2

- Déterminer le stress moyen ainsi que la variance moyenne pour chacun des traitements..

Question 3

- Réaliser une ANOVA et afficher le résultat

Question 4

- Quelles sont les conditions d'utilisation du modèle d'ANOVA
Montrer graphiquement (et avec explications en commentaire) que les 3 conditions sont respectées.

Question 5

- Montrer maintenant à l'aide des différents tests (Bartlett, Shapiro-Wilk et Dubin-Watson) que les conditions sont vérifiées.

Question 6

- Que pouvez-vous conclure sur l'ANOVA ?

Exercice 2

On a relevé le gain moyen de poids par jour sur des nourrissons à l'issu d'un traitement d'injection de vitamine D de 30 jours selon leur catégorie de poids à la naissance :

	P1	P2	P3
Traitement1	1,26	1,21	1,19
Traitement2	1,29	1,23	1,23
Traitement3	1,38	1,27	1,22

Question 1

- Créer le dataset dont vous aurez besoin : rentrer les données sous forme d'un tableau de données brutes c'est à dire avec 3 colonnes : traitement, PoidsNaissance et valeurs.
Le plus simple est de créer les trois vecteurs, puis de les concaténer, de renommer les colonnes et de bien les convertir en table.

Question 2

- Réaliser une Anova à deux facteurs

Question 3

- Vérifiez que les conditions d'utilisation d'Anova sont respectées

Question 4

- Que pouvez-vous en conclure

Question 5

- Nous allons maintenant effectuer une comparaison sans le terme d'interaction. Nous utiliserons pour cela le '+' dans la fonction d'ANOVA sur R :
`Valeur~Traitement+PoidOrigine`

Question 6

- Que pouvez-vous en dire ?

Exercice 3

Un tisseur s'intéresse à la teneur en pH au cœur de la teinture d'une bobine de coton.

Ces bobines sont fabriquées à partir de 3 lignes de production : L1, L2 et L3, et à partir de 3 solutions chimiques pour la teintures qui sont élaborées dans 5 sortes de citernes : C1, C2, C3, C4 et C5, chaque citerne pouvant alimenter n'importe quelle ligne de production.

Il s'aperçoit d'une disparité entre la teneur en pH entre les différentes bobines et il ne sait pas s'il doit mettre cette disparité au compte d'un effet "type de ligne de production" ou au compte d'un effet "type de citerne", ou au compte des deux effets.

Pour étudier cela de plus près, il prélève deux bobines et récupère les résultats suivant :

	C1	C2	C3	C4	C5
L1	5.5	6.2	5.4	5.6	6.3
	5.3	6.2	5.2	5.4	6.0
L2	5.5	6.4	5.4	5.4	6.0
	5.3	6.2	5.4	5.4	6.0
L3	5.6	6.0	5.3	5.6	6.3
	5.2	6.2	5.1	5.5	6.1

Question 1

- Créer le jeu de données grâce aux lignes de commande R suivantes :

```
pH = c(5.5, 6.2, 5.4, 5.6, 6.2, 5.3, 6.2, 5.2, 5.4, 6.0, 5.5, 6.4, 5.4, 5.4, 6.0, 5.3, 6.2,  
5.4, 5.4, 6.0, 5.6, 6.0, 5.3, 5.6, 6.3, 5.2, 6.2, 5.1, 5.5, 6.1)  
ligne = factor(rep(c("L1", "L2", "L3"), rep(10, 3)))  
citerne = factor(rep(c("C1", "C2", "C3", "C4", "C5"), 6))  
coton = data.frame(pH, ligne, citerne)
```

Question 2

- Est-ce qu'il y a un effet du facteur citerne?

Question 3

- Est-ce qu'il y a un effet du facteur ligne?

Question 4

- Est-ce qu'il y a une interaction entre les deux facteurs citerne et ligne?

Exercice 4

Nous avons relevé le pourcentage d'absorption du fer, suite à l'injection de 4 vitamines sur 6 races de chiens différents :

	VitA	VitB	VitC	VitD
R1	64	78	75	55
R2	72	91	93	66
R3	68	97	78	49
R4	77	82	71	64
R5	56	85	63	70
R6	95	77	76	68

Question 1

- Créer le dataset dont vous aurez besoin en vous aidant des lignes de commande de la question 1 de l'exercice précédent.

Question 2

- Quel type d'analyse de la variance devez-vous utiliser ? ANOVA 1 facteur ou ANOVA 2 facteurs ? Pourquoi ?
Effectuer l'analyse de la variance.

Question 3

- Vérifiez que les conditions d'utilisation d'Anova sont respectées

Question 4

- La race a-t-elle un effet significatif sur la quantité d'absorption du fer ?
Ecrivez explicitement les hypothèses H0 et H1 de votre modèle.

Question 5

- Le type de vitamine injecté a-t-il un effet significatif sur la quantité d'absorption du fer ?
Ecrivez explicitement les hypothèses H0 et H1 de votre modèle.

Question 6

- Analyser vos résultats d'ANOVA, que pouvez vous en conclure ?