

# Régression linéaire et analyse de la variance

Licence MIASHS - L3

Année 2015 - 2016

Simplice Dossou-Gbété - Jonathan Jung



## TD n° 2 : Analyse de la variance

**Exercice 1.** Un fabricant soupçonne que des lots de poudre de phosphate de calcium livrés par son fournisseur diffèrent de manière significative du point de vue de la teneur en calcium. Nous disposons d'un grand nombre de lots dans le dépôt. Cinq de ces lots sont prélevés au hasard pour analyse. Un chimiste réalise six déterminations sur chaque lot et obtient les résultats suivants pour la teneur en calcium (% pds) :

Lot 1	Lot 2	Lot 3	Lot 4	Lot 5
23.46	23.59	23.51	23.28	23.29
23.48	23.46	23.64	23.40	23.46
23.56	23.42	23.46	23.37	23.37
23.39	23.49	23.52	23.46	23.32
23.40	23.50	23.49	23.39	23.38
23.58	23.55	23.62	23.50	23.48

- Préciser le facteur de classification (lots ou type de détermination).
- Le plan d'expériences est-il équilibré ?
- Représenter graphiquement les données.
- Écrire le modèle d'analyse de la variance à un facteur en fonction de la réponse précédente.
- Préciser les hypothèses relatives à la méthode.
- Calculer les estimations des paramètres du modèle d'analyse de la variance en fonction de la contrainte d'identifiabilité choisie.
- Les données présentent-elles une évidence en faveur de la différence entre les lots ?

**Exercice 2.** On souhaite comparer trois traitements, notés A, B et C contre l'asthme. Le traitement B est un nouveau traitement mis en compétition avec les traitements classiques A et C. Les patients sont affectés à un traitement par tirage au sort parmi ceux venant consulter dans un centre de soin. On mesure sur chaque patient la durée, en jours, séparant de la prochaine crise d'asthme. Les mesures sont reportées dans le tableau ci-dessous

Traitement A	26	27	35	36	38	41	42	45	50	65							
Traitement B	29	42	44	44	45	48	48	52	56	56	58	58	60	61	63	63	69
Traitement C	26	26	30	30	33	36	38	38	39	46	47	51	51	56	75		

- Quelle est la population de référence pour cette étude ?
- Quelle est la variable de classification et quelle est la variable réponse ? Quel est le nombre de modalités de la variable de classification ? Comment appelle-t-on encore la variable de classification ?
- Le plan d'expérience est-il équilibré ?
- Écrire la matrice du modèle d'analyse de la variance et proposer une contrainte d'identifiabilité.
- Calculer les estimations des paramètres de la moyenne.
- Quelle méthode statistique peut-on utiliser pour conclure à l'existence d'une différence entre les traitements ? Préciser les hypothèses techniques pour la mise en œuvre de cette méthode.

- (g) Si les hypothèses techniques sont vérifiées, pourrait-on conclure à l'existence d'une différence entre les traitements au seuil associé à une probabilité d'erreur de première espèce de 0.05 ?
- (h) Quelle méthode statistique peut-on utiliser pour conclure à l'existence d'une différence entre le traitement A et le traitement B ? Mettre en œuvre de cette méthode.
- (i) Comparer les traitements 2 à 2 par la méthode de Bonferroni.
- (j) Pourriez-vous dire que les données sont compatibles avec les hypothèses ?

**Exercice 3.** Afin de tester l'hypothèse que la consommation de caféine facilite l'apprentissage on a comparé les scores à un examen de trois groupes d'étudiants : le groupe 1 boit une tasse, le groupe 2 boit 2 tasses et le groupe 3 boit 3 tasses de café. Leurs scores à l'examen sont les suivants

Matière	Groupe 1	Groupe 2	Groupe 3
Mathématiques	50	48	57
Informatique	42	47	59
Physique	53	65	48
Chimie	45	59	46
Mécanique	55	51	45

- (a) Quelle est la variable de classification et quelle est la variable réponse ? Quel est le nombre de modalités de la variable de classification ? Comment appelle-t-on encore la variable de classification ?
- (b) Quelle est la population de référence pour cette étude ?
- (c) Le plan d'expérience est-il équilibré ?
- (d) Représenter graphiquement les scores obtenus dans les trois groupes ainsi que leurs moyennes respectives.
- (e) Écrire le modèle d'analyse de la variance à un facteur.
- (f) Formuler l'hypothèse nulle de ce problème.
- (g) Estimer les paramètres du modèle d'analyse de la variance en précisant la contrainte d'identifiabilité utilisé pour cette estimation.
- (h) Les données sont-elles en faveur du rejet de cette hypothèse nulle ?
- (i) Préciser les hypothèses que doivent vérifier le mécanisme de génération des données pour la réalisation de ce test d'hypothèse.
- (j) Calculer des valeurs résiduelles ainsi que les valeurs résiduelles standardisées.
- (k) Analyser les résidus en vue de diagnostics sur les hypothèses de l'analyse de variance. Peut-on dire que les données ne contredisent pas l'hypothèse de normalité ?