

# T. D. n° 1

## Analyse de la variance à un facteur

### 1 Concepts nouveaux

1. Analyse de la variance
2. Modèle d'analyse de la variance à un facteur
3. Équation de décomposition de la variance
4. Tableau d'analyse de la variance
5. Homogénéité des variances (homoscédasticité versus hétéroscédasticité)
6. Comparaisons multiples de moyennes
7. Contraste linéaire
8. Coefficient  $\eta^2$
9. Puissance
10. Transformations normalisantes

**Tous les tests statistiques que vous effectuerez seront réalisés à  $\alpha = 5\%$  sauf mention contraire.**

#### Exercice 1 Questions de compréhension des concepts.

1. À quelle question permet de répondre une analyse de la variance à un facteur, qui par la suite sera notée ANOVA I ?
2. Expliquez intuitivement ce qu'est une équation d'analyse de la variance et comment elle va être utilisée pour définir le test de Fisher, noté  $F$ , sous-jacent à l'ANOVA I ?
3. Pouvez-vous donner une justification « mathématique » de la pertinence de la statistique  $F$  de l'ANOVA I ( $F = s_F^2/s_R^2$ ) pour tester l'égalité des moyennes des différentes modalités du facteur ?
4. Si nous utilisons une ANOVA I pour comparer deux moyennes, le test  $F$  réalisé est-il comparable à un des test  $t$  que vous avez étudié pour comparer deux moyennes. Si oui, lequel ? Quelle est la relation exacte entre les deux statistiques de test ? Perdons-nous quelque chose à faire une ANOVA I plutôt qu'un test  $t$  ?
5. Que mesure  $\eta^2$  ? Voyez-vous un lien avec une mesure semblable déjà définie dans un de vos cours de Licence ?
6. Quelles sont les hypothèses statistiques sous-jacentes au modèle d'ANOVA I ?
7. Citez deux tests pour vérifier l'homogénéité des variances quand vous réalisez une ANOVA I.

8. À quoi cela peut-il servir de transformer la réponse d'un modèle tel que le modèle d'ANOVA I puisse être utilisé ?
9. Dans des comparaisons multiples de moyennes en ANOVA I, pourquoi appliquons-nous des corrections de type Bonferroni ?
10. Quelle est la différence entre une méthode de comparaison à priori et une méthode de comparaison à posteriori ?
11. Définissez ce qu'est un contraste linéaire. À quel type de question permet-il de répondre en ANOVA I ?
12. À quelle question permettent de répondre les tests de Newman-Keuls et de Tukey ? En quoi sont-ils identiques et différents ? Lequel est le plus conservateur ?

**Exercice 2 Exercice issu du livre de Howell, Méthodes statistiques pour sciences sociales, De Boeck, 2008.**

Dans le but d'étudier l'effet de la méthode de mémorisation utilisée sur la rétention de mots, Eysenck (1974) demande à 50 sujets (entre 50 et 65 ans) d'étudier 27 mots différents. Les 50 sujets sont répartis dans 5 conditions expérimentales (10 sujets par condition) :

1. addition : les sujets doivent compter le nombre de lettres de chaque mot ;
2. rimes : les sujets doivent trouver une rime pour chaque mot ;
3. adjectifs : les sujets doivent donner un adjectif à attacher à chaque mot ;
4. images : les sujets doivent se former une image de chaque mot ;
5. intentionnel : les sujets sont informés qu'ils doivent tenter de retenir la liste de mots.

Les résultats des sujets (nombre de mots retenus sur les 27) se trouvent dans le fichier Exemple1 – Howell.MTW qui est sur le site :

<http://www-irma.u-strasbg.fr/mmaumy/enseignement/index.html>

1. Identifiez la variable dépendante et le facteur.
2. Écrivez les hypothèses statistiques du chercheur.
3. Calculez les moyennes et les écart-types de chaque modalité du facteur. Puis représentez les données sous la forme de box plot. Sur base de ces statistiques descriptives, quelle première conclusion pourriez-vous tirer ?
4. Écrivez le modèle de l'ANOVA à un facteur explicitement.
5. Donnez, sur base des résultats des sujets, les valeurs numériques des estimations de  $\mu$ , des  $\alpha_i$  et des  $\varepsilon_{ij}$ .
6. Réalisez une ANOVA I sur ces données. Pour cela, il faut vérifier que les conditions d'applications sont vérifiées. Quelle est la p-valeur du test ? Ce test est-il significatif ? Quelle conclusion pouvez-vous en tirer ?

7. S'il existe une différence entre les méthodes utilisées, utilisez une méthode de test appropriée pour classer ces méthodes en fonction de leurs efficacités.

**Exercice 3 Exercice issu du livre de Howell, Méthodes statistiques pour sciences sociales, De Boeck, 2008.**

Un chercheur, qui s'inspire des travaux de Conti et Musty (1984), s'intéresse à l'effet d'un ingrédient actif de la marijuana, le THC (tétrahydrocannabinol), sur l'activité motrice du rat. Il réalise une expérience où il a trois groupes de 5 rats. Il injecte au premier groupe un placebo, groupe témoin, au second groupe 0,1 g de THC, au troisième 0,5 g et au quatrième groupe 1 g et observe l'activité motrice du rat durant les 10 minutes après l'injection. Voici les résultats qu'il obtient :

Témoin	Dose = 0,1 g	Dose = 0,5 g	Dose = 1 g
30	60	71	33
27	42	50	78
52	48	38	71
38	52	59	58
20	28	65	35
26	93	58	35
8	32	74	46
41	46	67	32
49	63	61	
49	44		

1. Vérifiez que les conditions d'application de l'ANOVA I sont vérifiées.
2. Construisez, avec l'aide de MINITAB, le tableau d'analyse de la variance pour ces données.
3. Le produit a-t-il un effet significatif sur l'activité motrice du rat ? Écrivez explicitement les hypothèses nulle et alternative de votre test et justifiez votre réponse.
4. Calculez  $\eta^2$ . Que concluez-vous ?
5. Comparez, avec l'aide de MINITAB, deux à deux les trois traitements pour en déduire un classement des moyennes.
6. Réalisez, avec l'aide de MINITAB, un autre test adéquat pour comparer (en un bloc) les groupes qui ont reçu l'ingrédient actif et le groupe qui ne l'a pas reçu.

**Exercice 4 Exercice issu du livre de Howell, Méthodes statistiques pour sciences sociales, De Boeck, 2008.**

Foa, Rothbaum, Riggs et Murdock (1991) ont réalisé une étude qui évalue 4 types différents de thérapies appliquées à des victimes de viol. Les sujets du groupe

« Thérapie anti-stress » (14 sujets) ont reçu des instructions sur la manière de faire face au stress. Les sujets du groupe « Exposition prolongée » (10 sujets) ont repassé les événements de façon répétée dans leur tête. Les sujets du groupe « Guidance de soutien » (11 sujets) ont appris une technique générale de résolution de problèmes. Enfin, le groupe « Témoin liste d'attente » (10 sujets) n'a reçu aucune thérapie. Les données figurent ci-dessous, la variable dépendante étant l'estimation de la gravité à une série de symptômes.

Thérapie anti-stress	Exposition prolongée	Guidance de soutien	Témoin
3; 13; 13; 8; 11; 9; 12 7; 16; 15; 18; 12; 8; 10	18; 6; 21; 34; 26; 11; 2 5; 5; 26	24; 14; 21; 5; 17; 17 23; 19; 7; 27; 25	12; 30; 27; 20 17; 23; 13; 28 12; 13

1. Avant de réaliser une ANOVA I sur ces données, les chercheurs se sont demandés si les conditions d'application de l'ANOVA I sont vérifiées. Pouvez-vous répondre aux chercheurs? Pour répondre à leur question, vous utiliserez MINITAB pour réaliser les tests. Quelle conclusion pouvez-vous tirer du test d'homogénéité des variances? Justifiez.
2. Que conseillerez-vous aux chercheurs sur la base de ces résultats?
3. À cette étape, deux solutions peuvent être proposées.
  1. La première est d'appliquer une procédure de Welch (lorsque les échantillons ont des variances hétérogènes et qu'ils sont de taille inégale). Cette procédure s'est révélée efficace lorsque les échantillons proviennent de populations normales. Nous renvoyons au livre de Howell pour de plus amples renseignements sur cette procédure. De plus, MINITAB ne calculant pas cette statistique, nous abandonnerons cette solution. Et par conséquent, nous nous retournerons vers la seconde.
  2. La seconde est de faire une transformation des données. Citez quelles sont les transformations que vous connaissez. Quelle est celle que vous choisissez ici?
4. À partir de cette question, vous travaillerez non plus sur les données brutes mais les données transformées. Vérifiez que les conditions d'application du modèle de l'ANOVA I sont vérifiées. Le chercheur va pouvoir désormais réaliser une analyse de la variance à un facteur.
5. Donnez le tableau de l'analyse de la variance à l'aide MINITAB.
6. Calculez  $\eta^2$ .
7. Tracez un graphe indiquant les moyennes des quatre groupes.
8. Que concluez-vous?