

Psy1004 – Section 8:

Analyse de variance à un facteur

- Plan du cours:

  - Varia
  - Retour sur intra/TP2
  - 8.0: Les besoins d'une bonne expérience
    - a) Besoin au niveau de la méthode scientifique
    - b) Implications au niveau de la méthode statistique
  - 8.1: Logique derrière l'ANOVA
- Plan du cours:

  - 8.2: Calcul d'une ANOVA
    - a) Répartition de la somme des carrés
    - b) Répartition des degrés de liberté
    - c) Tableau récapitulatif
  - 8.3: Test  $F$ .

Disponible sur: <http://mapageweb.umontreal.ca/cousined/home/course/PSY1004>

---

---

---

---

---

---

---

---

Varia

■ ?

---

---

---

---

---

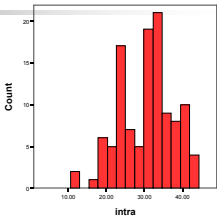
---

---

---

Retour sur intra (1/3)

- Moyenne: 28 / 49 (soit 57%)
- Écart type: 11.0
- Asymétrie: -1.2
- Examen très dur, car je dois discriminer.



---

---

---

---

---

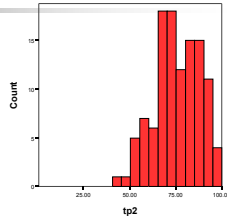
---

---

---

## Retour sur TP2

- Moyenne: 75.6
- Écart type: 11.9
- Asymétrie: -0.23




---

---

---

---

---

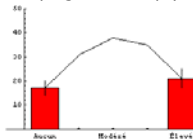
---

---

---

## 8.0: Les besoins d'une bonne expérience a) au niveau de la méthode scientifique (1/2)

- Exemple: effet du stress sur la mémoire.
  - Vous effectuez un test de mémoire standard (50 mots sont lus, puis le participant doit en rappeler le plus grand nombre possible).
  - Pour la moitié des participants, vous leur donnez un haut niveau de stress en leur disant que s'ils rapportent moins de 30 mots, ils seront exclus du programme de psychologie (pas éthique).
  - Résultats:



- Conclusion: "Le stress n'a aucun effet significatif sur la mémoire ( $t(78) = 1.05, p > .05$ )."

---

---

---

---

---

---

---

---

## 8.0: Les besoins d'une bonne expérience a) au niveau de la méthode scientifique (2/2)

- Les résultats sont-ils bons? oui
- Les résultats sont-ils valides? non!
  - parce que seulement deux niveaux de stress ont été explorés, dont un particulièrement extrême...
  - Pour vraiment décider si le stress a un effet sur la mémoire, il aurait fallu explorer plusieurs niveaux de stress.
  - Par exemple, avoir cinq groupes:
    - groupe 1: Il s'agit d'une pratique
    - groupe 2: Ce test de mémoire compte pour 10% de votre session
    - groupe 3: Ce test de mémoire compte pour 25% de votre session
    - groupe 4: Ce test compte pour 50% de votre session
    - groupe 5: Le résultat de ce test sera la note finale dans le bulletin
- Solution au niveau de la méthode expérimentale:
  - Toujours faire une expérience dans laquelle plusieurs niveaux de "difficulté" sont explorés.
- Exemples:
  - Impact de l'alcool sur la qualité de la conduite (0, .02, .04, .06, .08, ...)
  - Dosage idéal de patch pour arrêter de fumer (1 par jour, 2 par jour, ...)

---

---

---

---

---

---

---

---

## 8.0: Les besoins d'une bonne expérience

### b) Implications pour la méthode statistique

Supposons que:

nous ayons décidé de tester 5 niveaux (k) de stress (disons 1 à 5) →

Nous voulons savoir si le stress affecte la performance (i.e. peu importe le niveau particulier):

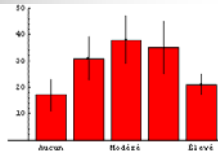
- $H_0: \mu = \mu$  pour tous les niveaux  $i, j$ .
- $H_1: \mu \neq \mu$  pour au moins une paire  $i, j$ .

Pour vérifier  $H_0$  avec un test  $t$ , il faudrait comparer le niveau 1 avec le niveau 2, le niveau 1 avec le niveau 3, ... le niveau 4 avec le niveau 5

- Ceci implique un très grand nombre de comparaisons  $\binom{5}{2} = 10$
- Le risque d'erreur  $\alpha$  augmente considérablement (alors qu'on n'est pas censé le dépasser dans une décision).

■ Solution au niveau de la méthode statistique:

- L'ANOVA (analyses of variance): un seul test qui teste simultanément k moyennes pour voir s'il existe n'importe quel deux niveaux qui diffèrent.
- L'ANOVA est un test paramétrique sur les moyennes (deux ou plus).




---

---

---

---

---

---

---

---

---

---

## 8.1: Logique derrière l'ANOVA

### a) Les sources de variances (1/4)

- Exemple de données pour k=3 niveaux: (Il y a trois sujets par niveaux, soit un échantillon total de 9: petit)

Il y a de la **variance intra-groupe**. Cette variance reflète "l'erreur expérimentale"

Idem. Seule "l'erreur expérimentale" est en jeu ici.

Idem. Les trois variances intra-groupes devraient être similaires

Variance entre les moyennes des groupes: **Variance intergroupe**. Si les trois groupes sont similaires (pas d'effet du stress), la variance entre les groupes devraient refléter elle aussi l'erreur expérimentale.

		Facteur A (e.g. Stress)		
		1	2	3
Données brutes (n = 3)	1	3	3	5
	2	5	4	5
	3	7	5	8
Moyenne		5	4	6
Écart type		2	1	1.73
Moyenne globale:		5		

Note: Ici, on parle en terme de variance, mais la même logique s'applique en terme d'écart type.

---

---

---

---

---

---

---

---

---

---

## 8.1: Logique derrière l'ANOVA

### a) Les sources de variances (2/4)

- Exemple de données pour k=3 niveaux: (Ici, il y a un effet du stress)

Il y a de la **variance intra-groupe**. Cette erreur expérimentale affecte les données d'un groupe de +ou- 2 (et ce, peu importe le groupe).

Variance intergroupe. Cette variance affecte les moyennes de plus ou moins 3-4. La variabilité entre les groupes est supérieure à la variabilité dans un groupe (n'importe lequel). Il y a donc autre chose que de l'erreur expérimentale en jeu ici: **Effet du facteur Stress**.

		Facteur A (e.g. Stress)		
		1	2	3
Données brutes (n = 3)	1	3	3	10
	2	5	4	10
	3	7	5	13
Moyenne		5	4	11
Écart type		2	1	1.73
Moyenne globale:		6.66		

Note: penser en terme de variabilité, pas en terme de variance

---

---

---

---

---

---

---

---

---

---

## 8.1: Logique derrière l'ANOVA

### b) nomenclature (3/4)

- Pour ne pas confondre les différents éléments, nous allons utiliser ces indices:

- $I$  : "étant donné"
- $S$  : "entre les sujets"
- $A_1$  : le niveau 1 du facteur A (i.e. le niveau de stress 1).
- $V_{S|A_1}$  : Variabilité entre les sujets étant donné qu'ils sont dans le niveau de stress 1, i.e. entre les scores brutes du groupe 1 par rapport à la moyenne du groupe 1.
- $V_{S|A}$  : Variabilité entre les sujets étant donnée qu'ils sont dans un groupe (le 1, le 2 ou le 3), i.e. entre les scores brutes par rapport à leurs moyennes respectives.
- $V_A$  : Variabilité entre les groupes, i.e. entre les moyennes par rapport à la moyenne globale.
- $V_T$  : Variabilité totale, i.e. entre les scores brutes et la moyenne globale.

---

---

---

---

---

---

---

---

## 8.1: Logique derrière l'ANOVA

### c) idée fondatrice de l'ANOVA (4/4)

- S'il n'y a que de l'erreur expérimentale (pas d'effet du stress), alors la variance intergroupe ( $V_A$ ) devrait être la même que la variance intra-groupe ( $V_{S|A}$ ).

- Car la variance intra-groupe reflète l'erreur expérimentale
- Car la variance intergroupe reflète l'erreur expérimentale:
- Les deux doivent donc avoir à peu près la même ampleur;
- Autrement dit:  $V_A \approx V_{S|A}$  ou encore:

$$\frac{V_A}{V_{S|A}} \approx 1; \frac{\text{Erreur expérimentale}}{\text{Erreur expérimentale}} \approx 1$$

- S'il y a un effet du facteur Stress, alors la variance entre les groupes (intergroupes  $V_A$ ) devrait être supérieurs à la variance intra-groupe  $V_{S|A}$ :  $\frac{V_A}{V_{S|A}} > 1$ ;  $\frac{\text{Erreur exp.} + \text{Effet de A}}{\text{Erreur expérimentale}} > 1$

---

---

---

---

---

---

---

---

## 8.2: Calcul d'une ANOVA à la main

### a) Répartition de la somme des carrés

- Rappel: la variance se calcule par:  $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

**Degré de liberté (dl):**  
pondère la somme des carrés par le nombre d'observation - 1

**Somme des carrés (SC):** La somme des écarts à la moyenne au carré (SEC)

SC intra-groupe				
1	2	3		
score	score	score	SC <sub>S A1</sub>	SC <sub>S A2</sub>
3	4	5	1	1
5	0	4	0	5
7	4	5	1	8
5	4	6	4	4
<b>X</b>				
<b>SC</b>	8	2	6	16
<b>total</b>				

SC <sub>total</sub>	
score	SC <sub>total</sub>
3	4
5	0
7	4
3	4
4	1
5	0
5	0
5	0
8	9
<b>X</b>	
<b>SC</b>	22

Remarquez que:

$$SC_{S|A} + SC_A = SC_T$$

Ce sera toujours vrai.

SC intergroupe			
moyenne	SEC <sub>X</sub>	nbre sujet	SC <sub>A</sub>
5	0	3	0
4	1	3	3
6	1	3	3
<b>X</b>			
<b>SC</b>	5	6	6

---

---

---

---

---

---

---

---

## 8.2: Calcul d'une ANOVA à la main

### b) Répartition des degrés de liberté

- Nous avons trois types de SC; il faut les pondérer par le nombre adéquat de degrés de liberté pour obtenir les variances correspondantes ( $k=3$  niveaux de stress,  $n=3$  sujets par groupe):
    - $dl_{S|A} = k(n-1)$  ex précédent: 6
    - $dl_A = k-1$  2
    - $dl_T = k n - 1$  8
- Remarquez que:  $dl_{S|A} + dl_A = dl_T$   
Ce sera toujours vrai.
- Nous pouvons alors diviser les SC par les dl respectifs, pour obtenir la moyenne des écarts au carré (MEC – CM):
- |                       |              |      |
|-----------------------|--------------|------|
| $SC_{S A} / dl_{S A}$ | $= CM_{S A}$ | 2.66 |
| $SC_A / dl_A$         | $= CM_A$     | 3    |
| $SC_T / dl_T$         | $= CM_T$     | 2.75 |
- Dans les faits, CM est une variance, mais comme il s'agit d'une variance répartie d'une façon spéciale, on l'appelle plus neutrement le "carré moyen"

---

---

---

---

---

---

---

---

## 8.2: Calcul d'une ANOVA à la main

### c) Tableau récapitulatif

Les résultats d'une ANOVA sont souvent présentés dans un tableau récapitulatif:

	SC	dl	CM	F
Stress (A)	6	2	3	1.13
Erreur (S A)	16	6	2.66	
Total	22	8		

➔ Détails

- Stress: i.e. le facteur intergroupe, qui distingue les groupes, avec 3 niveaux dans notre exemple (d'où 2 dl).
- Erreur: notre estimé de l'erreur expérimentale, i.e. de la variabilité uniquement dû à des différences entre les sujets d'un même groupe; il s'agit de la variance intra-groupe (trois groupes à 3 sujets donne 3 fois (3-1) dl, soit 6).
- Total: Uniquement dans le but de vérifier que les SC et dl sont bien réparties.
- F (parfois appelé ratio, ou encore ratio F), le rapport entre CM intergroupe et CM intra-groupe.

---

---

---

---

---

---

---

---

## Pourquoi fait-on tout ça?

On veut savoir si les moyennes entre les différents niveaux de stress diffèrent.

Plus précisément, on recherche une différence, peu importe laquelle.

Pour ce faire, on regarde si la variabilité entre les groupes (variabilité intergroupe) surpasse la variabilité à l'intérieur d'un groupe (variabilité intra-groupe).

La variabilité à l'intérieur d'un groupe est un indicateur de l'erreur expérimentale (propre à tout échantillonnage).

Le  $CM_A$  estime la variabilité intergroupe (égal 3),

Le  $CM_{S|A}$  estime la variabilité intra-groupe (égal 2.66).

Le ratio  $CM_A$  sur  $CM_{S|A}$  doit être proche de 1 si les groupes ne diffèrent pas. Nous avons obtenu 1.13. Est-ce que ce nombre diffère de 1?

	Facteur A (e.g. Stress)		
	1	2	3
Données	3	3	5
brutes	5	4	5
( $n = 3$ )	7	5	8
Moyenne	5	4	6
Écart type	2	1	1.73

---

---

---

---

---

---

---

---

### 8.3: Test F.

■ Hypothèses:

- $H_0: \mu_i = \mu_j$ , pour tous les niveaux  $i, j$  (i.e.  $F = 1$ )
- $H_1: \mu_i \neq \mu_j$ , pour au moins deux niveaux  $i, j$  (i.e.  $F > 1$ )

■ Seuil:

- 5% à moins d'indication contraire

■ Test:

- Rejet de  $H_0$  si  $F = \frac{CM_A}{CM_{S/A}} > s(\alpha)$  (l'ex.: 1.13 pour 2,6 dl)

- $s(\alpha)$  lu sur une table F (distribution de Fisher) avec deux degrés de libertés: degré de liberté de A et degré de liberté de S/A:  $F(d_A, d_{S/A})$ . (Seul la table pour  $\alpha = 5\%$  fournie dans les notes)

■ Calcul et conclusion:

- "Il n'existe pas de différence significative entre les moyennes ( $F(2,6) = 1.13, p > .05$ ). Autrement dit, le stress n'affecte pas la mémoire dans notre tâche." (données fictives)

PSY1004-043 - Section 8 p. 16

### Exemple 1

Soit une étude sur l'effet de la musique sur les palpitations cardiaques. La musique (du classique) a été manipulée suivant le volume en décibel (30, 60, 100, et 200 dB). Les palpitations cardiaques ont été mesurées sur 10 sujets dans chaque groupe sur 30 secondes.

- a et b) Quelles sont les hypothèses? le seuil?
- c) Voici le tableau des ANOVA, complétez-le.

	SC	dl	CM	F
Volume (A)	156	xxx	xxx	xxx
Erreur (S/A)	48	xxx	xxx	
Total	xxx	xxx		

- d) Concluez:

PSY1004-043 - Section 8 p. 17

### Exemple 2

Soit une étude sur l'effet de la difficulté d'une tâche de coordination motrice sur le nombre d'heure à pratiquer la tâche. La tâche est de faire une routine avec un yo-yo. 8 routines ont été sélectionnées (classée par des champions en yo-yo de la plus facile à la plus dure). On mesure combien d'heure la personne pratique la tâche avant d'abandonner le yo-yo.

- a et b) Quelles sont les hypothèses? le seuil?
- c) Voici le tableau des ANOVA, complétez-le.

	SC	dl	CM	F
difficulté (A)	115.5	xxx	xxx	xxx
Erreur (S/A)	xxx	120	5.5	
Total	xxx	xxx		

- d) Concluez:

PSY1004-043 - Section 8 p. 18



### Exemple 3

→ Détails

Effet de la privation de sommeil sur le QI. On teste 6 niveaux de privation (18 heures, 24 heures, 30 heures, 36 heures, 42 heures, et 48 heures). Il y a 6 participants dans chaque situations.

- a et b) Quelles sont les hypothèses? le seuil?
- c) Voici le listing de SPSS, complétez le tableau d'ANOVA.

	SC	df	CM	F
privation (A)				
Erreur (S/A)				
Total				

- d) Concluez:

---

---

---

---

---

---

---

---