



Faculté de psychologie

Pôle Humanités

Liens entre deux variables nominales (analyse d'indépendance) Tests du Chi²

Module HPS3-3A

GALHARRET J-M,
Laboratoire de Mathématiques Jean Leray
Faculté de Psychologie.

Variable nominale

Rappels

- Une variable nominale (ou catégorielle ou qualitative) est une variable dont les modalités sont des catégories :

Tableau de contingence

Effectifs croisés entre les modalités des variables nominales

On va compter le nombre d'individus correspondants au croisement des modalités des deux variables :

		Année d'étude			
		L1	L2	L3	TOTAL
Reprise d'étude	OUI	5 %	5 %	5 %	5 %
	NON	95 %	95 %	95 %	95 %
	Total	100 %	100 %	100 %	100 %

		Année d'étude			
		L1	L2	L3	TOTAL
Reprise d'étude	OUI	33,3 %	33,3 %	33,3 %	100 %
	NON	33,3 %	33,3 %	33,3 %	100 %
	Total	33,3 %	33,3 %	33,3 %	100 %

Hypothèse H0

Indépendance

L'indépendance entre les variables X et Y va s'exprimer de la façon suivante : soit i une modalité de X et j une modalité de Y :

$$\mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i) \times \mathbb{P}(Y = j)$$

Par exemple si le fait d'avoir repris ses études et la réussite au semestre 1 sont indépendants on a :

$$\mathbb{P}(\textit{Reprise} = \textit{OUI}, \textit{Reussite} = \textit{OUI}) = \mathbb{P}(\textit{Reprise} = \textit{OUI}) \times \mathbb{P}(\textit{Reussite} = \textit{OUI})$$

c'est à dire ici

$$\mathbb{P}(\textit{Reprise} = \textit{OUI}, \textit{Reussite} = \textit{OUI}) = \frac{10}{420} = 2.3 \%$$

$$\mathbb{P}(\textit{Reprise} = \textit{OUI}) \times \mathbb{P}(\textit{Reussite} = \textit{OUI}) = \frac{260}{420} \times \frac{25}{420} = 3.7 \%$$

Effectifs théoriques

Sous l'hypothèse d'indépendance

Effectifs observés

		Réussite au semestre 1		
		OUI	NON	Total
Reprise d'étude	OUI	10	15	25
	NON	250	150	400
	Total	260	165	425

Effectifs théoriques

		Réussite au semestre 1		
		OUI	NON	Total
Reprise d'étude	OUI	15.3	9.7	25
	NON	244.7	155.3	400
	Total	260	165	425

$$n_{th}(OUI, OUI) = \frac{260 \times 25}{425} = 15.3$$

Statistique de test

		Réussite au semestre 1		
		OUI	NON	Total
Reprise d'étude	OUI	10	15	25
	NON	250	150	400
	Total	260	165	425

		Réussite au semestre 1		
		OUI	NON	Total
Reprise d'étude	OUI	15.3	9.7	25
	NON	244.7	155.3	400
	Total	260	165	425

$$\begin{aligned}
 \chi^2 &= \sum_{\text{cellules}} \frac{(n_{\text{obs}} - n_{\text{th}})^2}{n_{\text{th}}} \\
 &= \frac{(10 - 15.3)^2}{15.3} + \frac{(15 - 9.7)^2}{9.7} + \frac{(250 - 244.7)^2}{244.7} + \frac{(150 - 155.3)^2}{155.3} \\
 &= 5.02
 \end{aligned}$$

Table du Chi2

On lit dans la table du Chi2 sur la ligne $ddl = (I - 1) \times (J - 1)$ où I est le nombre de modalités de la première variable et J le nombre de modalités de la deuxième variable. On choisit $I \leq J$

$$ddl = 1 \times 1 = 1$$

ddl	p_value				
	0,1	0,05		0,01	0,001
1	2,706	3,841		6,635	10,83
2	4,605	5,991		9,210	13,82
3	6,251	7,815		11,34	16,27
4	7,779	9,488		13,28	18,47
5	9,236	11,07		15,09	20,52
6	10,64	12,59		16,81	22,46
7	12,02	14,07		18,48	24,32
8	13,36	15,51		20,09	26,12
9	14,68	16,92	5.02	21,67	27,88
10	15,99	18,31		23,21	29,59

Taille d'effet

On définit le V de Cramer :

$$V = \sqrt{\frac{\chi^2}{N \times (I - 1)}}$$

Les normes de Cohen sont identiques à celles pour la corrélation.

V	[0.1,0.3[[0.3,0.5[[0.5,1]
Effet	Faible	Modéré	Fort

$$V = \sqrt{\frac{5.02}{420 \times 1}} = .109$$

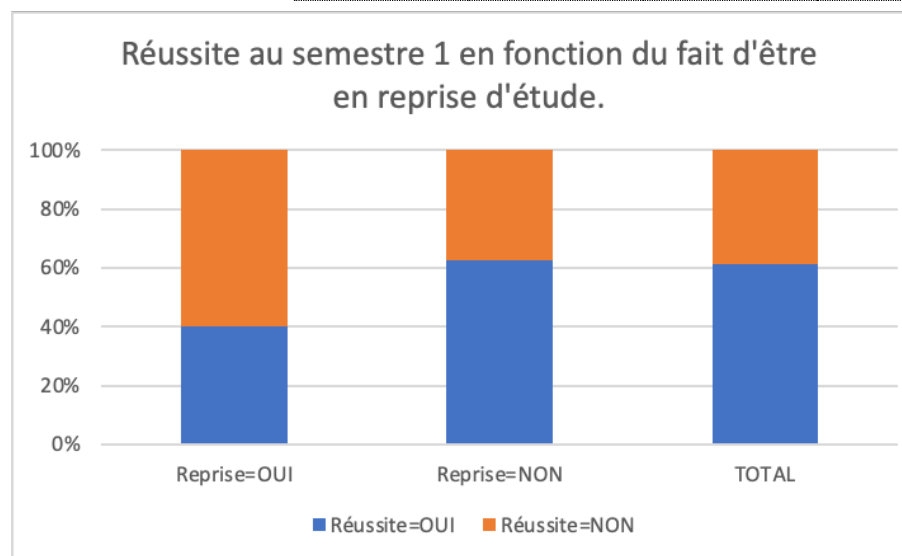
Autre vision de l'indépendance

Profil lignes

On va calculer la proportion d'étudiants en reprise d'étude qui a réussi, celle qui ne l'est pas et la proportion d'étudiants qui a réussi.

		Réussite au semestre 1		
		OUI	NON	Total
Reprise d'étude	OUI	40 %	60 %	100 %
	NON	62,5 %	37,5 %	100 %
	Total	61,2 %	38,8 %	100 %

		Réussite au semestre 1		
		OUI	NON	Total
Reprise d'étude	OUI	10	15	25
	NON	250	150	400
	Total	260	165	425



Les proportions observées sur les reprise d'études (40 % ,60%) est-elle significativement différente de celles des non reprises d'études (62.5 % ,37.5%) ?

Conclusion normes APA

Un test du Chi2 d'indépendance a permis de mettre en évidence un lien significatif entre la réussite au premier semestre et le fait d'être en reprise d'étude,

$$\chi^2(1) = 5.02, p < .05$$

Plus précisément, sur l'échantillon, seuls 40% des reprises d'études ont réussi leur semestre 1 alors qu'ils sont 62.5 % pour les autres.

Exemple 2

- On se pose une nouvelle question : peut-on considérer qu'il est plus facile de reprendre ces études en L1 qu'en L2 et L3 ?
- Comment procéder ?
- On pose la question suivante aux étudiants en L1, en L2 et en L3 : avez-vous réussi votre année ?
- On a deux variables nominales : réussite (OUI/NON), année d'étude (L1,L2,L3).
- On obtient les réponses ci-dessous, parmi les étudiants en reprise d'étude :

	L1	L2	L3	TOTAL
Oui	30	20	5	55
Non	40	30	30	100
TOTAL	70	50	35	155

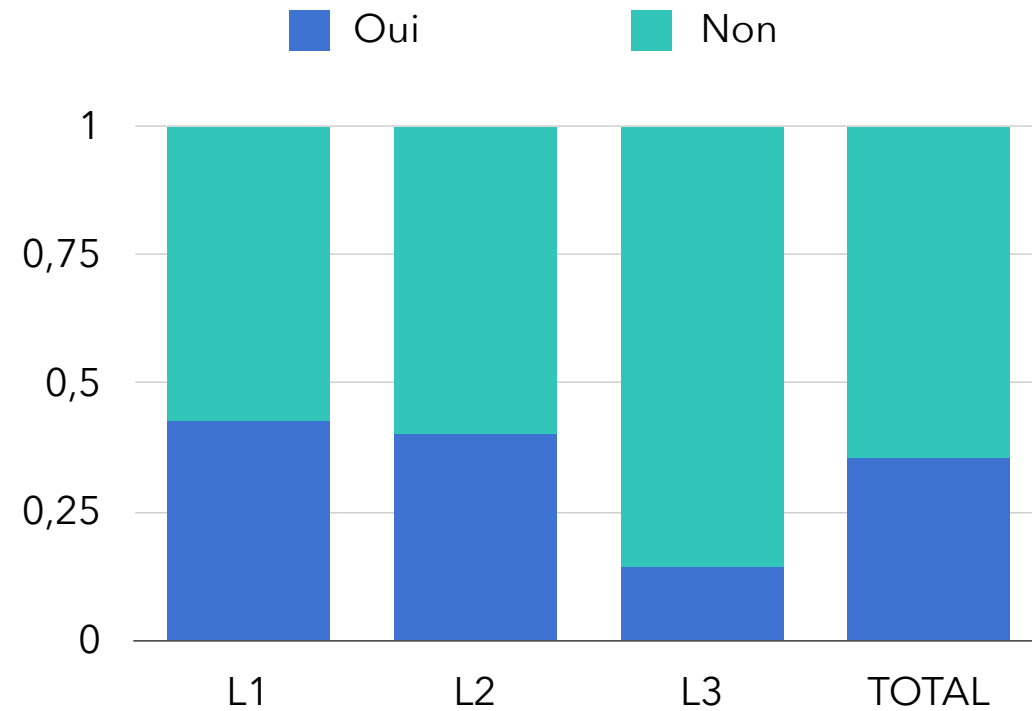


Table des profils colonnes

	L1	L2	L3	TOTAL
Oui	42,9 %	40,0 %	14,3 %	35,5 %
Non	57,1 %	60,0 %	85,7 %	64,5 %
TOTAL	100,0 %	100,0 %	100,0 %	100,0 %

Réalisation du Test

Tableau des effectifs observés

	L1	L2	L3	TOTAL
Oui	30	20	5	55
Non	40	30	30	100
TOTAL	70	50	35	155

Tableau des effectifs théoriques

	L1	L2	L3	TOTAL
Oui	24,8	17,7	12,4	54,9
Non	45,2	32,3	22,6	100,1
TOTAL	70	50	35	155

$$\chi^2 = 8,53$$

$$V = \sqrt{\frac{8,53}{1 \times 135}} = 0,25$$

Un test du Chi2 a permis de mettre en évidence une différence significative de réussite chez étudiants en reprise d'étude selon l'année durant laquelle ils ont repris leurs études, $\chi^2(2) = 8.53$, $p < .05$, $V = 0.25$. Plus précisément, les étudiants en reprise d'étude réussissent mieux lorsqu'ils reprennent en L1 et L2 qu'en L3, ce lien est faible d'après les standards de Cohen.

Test du Chi² de Mac Némar

Conditions d'application

- On considère une variable Y binaire qui est mesurée deux fois sur les mêmes individus (groupes appariés).
- Typiquement on va l'appliquer dans une situation du type : on pose une question d'intention (réponse OUI/NON) avant et après une expérimentation. On s'intéresse au individus qui ont changé d'avis. On teste l'hypothèse :

Y1\Y2	0	1
0	n ₁	n ₂
1	n ₃	n ₄

$$H_0 : p_{0/1} = p_{1/0}$$

$$\chi^2 = \frac{(n_2 - n_3)^2}{n_2 + n_3}$$

$ddl = 1$ pour le Test de Mac-Némar

Exemple

On considère la situation suivante : on veut voir l'effet du nouveau design d'un smartphone sur l'intention d'achat.

- Pour ce faire on pose la question à 1000 personnes de l'achat de ce nouveau smartphone avant que celui-ci est été dévoilé 450 déclarent vouloir l'acheter.
- Une fois le smartphone dévoilé on pose de nouveau la même question 500 personnes déclarent vouloir l'acheter dont 100 ont changé d'avis positivement (c'est à dire qu'ils n'avaient pas l'intention a priori de l'acheter et l'ont une fois le smartphone dévoilé)

AV\AP	OUI	NON	TOTAL
OUI	400	50	450
NON	100	400	550
TOTAL	500	500	1000

$$\chi_{MN}^2 = \frac{(100 - 50)^2}{100 + 50} = 16.7$$

Conclusion

Un test de Mac-Némar a permis de mettre en évidence qu'il y a significativement plus d'individus qui ont changé d'avis positivement que négativement,
 $\chi_{MN}^2(1) = 16.7, p < .001.$