Chapitre 10 : Test de Mac Nemar et Analyse Statistique Implicative

Jean-Claude Régnier

Université de Lyon - UMR 5191 ICAR
ENS-LSH 15, Parvis René Descartes BP 7000 69342 LYON cedex 07
jean-claude.regnier@univ-lyon2.fr

Résumé. Nous tentons de comparer, dans ce chapitre, les trois approches pour étudier des liens vraisemblables entre deux variables binaires, que sont : l'ASI, le test de Mac Nemar et le test d'indépendance fondé sur la mesure du $\chi 2$. Pour ce faire, nous faisons un retour sur des données issues de nos travaux passés dans le champ de la didactique des mathématiques.

1 Introduction

Comme nous l'avons déjà évoqué succinctement dans la première partie de cet ouvrage, la comparaison de deux séries successives de données binaires de type présence-absence ou échec-réussite relevées sur le même échantillon d'individus comme cela est le cas en ASI, peut aussi être réalisée à l'aide du test du χ^2 de Mac Nemar.

2 Le test de Mac Nemar

2.1 Un exemple introductif issu d'une situation dans le cadre de la didactique des mathématiques

Pour comparer la difficulté de deux épreuves de mathématiques passées par le même groupe d'individus (séries dites appariées), c'est cette approche qui avait été choisie à la fin des années 70 et début 80 par Jean-Claude Régnier (Régnier 1980 p. 62-75, 1983) dans ses travaux de DEA et thèse en didactique des mathématiques. Comme nous l'avons déjà présenté, l'information est résumée dans un tableau 2x2 dont nous donnons un exemple ci-dessous. Pour rester congruent au mode de présentation des recherches de liens génériquement notés $a \Rightarrow b$ dans le contexte ASI qui présuppose que $N(a) \leq N(b)$, nous représentons systématiquement la variable binaire a en ligne et la variable binaire b en colonne.

		Variable b = Épreuve Finale (item 113)		
		1 = Réussite	0 = Échec	**Total**
Variable a = Épreuve Initiale (item 101 et/ou item 201)	1 = Réussite	56	6	62
	0 = Échec	14	26	40
	Total	70	32	102

TAB. 1 - *Tableau extrait (Régnier 1983, p. 164-166) : items 101 et 201 -- item 113*

Pour mieux illustrer notre propos, nous rappelons succinctement le contenu de ces items qui relèvent du champ de la trigonométrie.

A partir d'angles **aigus** représentés graphiquement et fournis dans l'épreuve et non rapportés ici :

Item 101	Déterminer le sinus et le cosinus de l'angle (101) en utilisant le demi-cercle trigonométrique donné par report avec un calque ou par une construction.
Item 201	Déterminer le sinus et le cosinus de l'angle (201) en construisant un demi-cercle trigonométrique (unité 10 cm). Faire figurer la construction sur la feuille. Utiliser l'équerre pour construire les perpendiculaires.
Item 113	Déterminer le sinus et le cosinus de l'angle (113) à l'aide du demi-cercle trigonométrique fourni ou. par construction, sur la figure, d'un demi-cercle trigonométrique (unité 10 cm)

La question que nous nous posons alors est de savoir si les fréquences de réussite aux deux épreuves sont significativement différentes ou non.

2.2 Formalisation succincte du test de Mac Nemar

L'idée de Mac Nemar pour étudier ce type de lien entre les deux épreuves est qu'il est plus pertinent de ne prendre en compte que les discordances entre les deux épreuves c'est-à -dire le cas de réussite à l'une et d'échec à l'autre et son complémentaire. Dans le tableau ci-dessus, ce sont les deux effectifs 14 et 6 correspondant aux couples (A_Echec, B_Réussite) et (A_Réussite, B_Echec) qui sont considérés comme des informations majeures. Cette idée n'est pas rendue par le test du χ^2 d'indépendance que nous avons déjà évoqué précédemment (Partie 1 Chap. 1-6.6) en établissant la relation algébrique entre l'indice d'implication et la mesure du χ^2.

Si nous nous remettons dans le contexte de l'ASI, le tableau de référence est donc celui-ci :

		Variable b		
		1	0	**Total**
Variable a	1	$n(a \wedge b)$	$n(a \wedge \bar{b})$	$n(a)$
	0	$n(\bar{a} \wedge b)$	$n(\bar{a} \wedge \bar{b})$	$n(\bar{a})$
	Total	$n(b)$	$n(\bar{b})$	n

TAB. 2- *Tableau de contingence avec les notations ASI*

Dans l'hypothèse d'une équivalence entre les deux épreuves, la fréquence de ceux qui sont passés d'une réussite à un échec parmi ceux qui ont changé d'état est égale à la fréquence de ceux qui sont passés d'un échec à une réussite parmi ceux qui ont changé d'état, c'est à dire égale à 0,5. D'une certaine manière, cela revient à comparer une fréquence observée à une fréquence théorique de 0,5.

Mac Nemar a montré qu'il suffisait de prendre comme indice, la mesure suivante que nous nommerons χ^2 de Mac Nemar, $\chi^2_{MacNemar} = \frac{(n(\bar{a} \wedge b) - n(a \wedge \bar{b}))^2}{n(\bar{a} \wedge b) + n(a \wedge \bar{b})}$ dont la loi de probabilité est approximativement celle de la variable de Pearson χ^2 de degré de liberté 1

En résumé les 4 étapes de la démarche de ce test sont les suivantes :

- Étape 1 : formulation des hypothèses :

H_0 : symétrie des changements d'état entre les deux épreuves
H_1 : non-symétrie des changements d'état entre les deux épreuves

- Étape 2 : calcul de la valeur empirique du $\chi^2_{\text{(Mac Nemar)}}$
- Étape 3 : lecture de la valeur critique dans la table du χ^2 de Pearson de ddl=1 pour un risque α donné
- Étape 4 : décision statistique rejet ou non rejet de H_0

2.3 Retour à l'exemple présenté.

Dans le cas présenté, nous calculons la valeur empirique comme suit $\chi^2_{MacNemar} = \frac{(14-6)^2}{14+6} = \frac{8^2}{20} = 3{,}2$ et nous la confrontons à la valeur critique au niveau de risque α. Si nous choisissons un niveau de risque de 0,05, la valeur critique est alors de 3,84. Comme 3,2<3,84, nous ne rejetons pas l'hypothèse d'équivalence des deux épreuves que nous considérons comme telle avec un risque de 2ème espèce β inconnu.

Si nous revenons à la perspective de recherche de lien par rejet de l'indépendance en appliquant le test du χ^2 d'indépendance, nous trouvons une valeur empirique de 34,56 qui est très largement supérieure à la valeur critique de 3,84 pour un niveau de risque α=0,05 et même à la valeur critique 6,63 pour un niveau de risque α=0,01. Au sens du test du χ^2 d'indépendance, il existe donc un lien fort entre les deux variables.

Si nous nous plaçons dans la perspective de recherche de lien au sens de l'ASI, le calcul de l'intensité d'implication $\varphi_P(a,b)$ avec le modèle de Poisson et le calcul de l'intensité d'implication $\varphi_{BIN}(a,b)$ avec le modèle binomial

$\chi^2 = 34{,}56$		$\chi^2{}_{MC} = 3{,}2$		Intensités d'implication
a-b	b=1	b=0		$\varphi_P(a,b)$
a=1	56	6	62	0,9996
a=0	14	26	40	$\varphi_{BIN}(a,b)$
	70	32	102	0,9998

TAB. 3 - *analyse selon les trois perspectives: χ^2 d'indépendance, χ^2 Mac Nemar, ASI,*

Les valeurs qui figurent dans le tableau ci-dessus, indiquent un niveau de confiance en l'implication statistique (Réussir Item 101 et/ou Item 201)⇒(Réussir Item 113) supérieur à 0,99. Ce que nous avions pris en compte à l'époque par la mise en œuvre de l'indice d'implication que Régis Gras (1979) avait exposé dans sa thèse deux ans plus tôt. Ce que nous découvrons à ce jour, c'est qu'il y avait eu une erreur dans les références des tableaux, ce qui conduisait à prendre comme valeur $\chi^2_{MacNemar} = \frac{(56-26)^2}{56+26} = \frac{30^2}{82} = 10{,}97$ et nous avions alors appuyé le sens de la quasi-implication sur l'approche du test de Mac Nemar. Ce retour est donc l'occasion d'une rectification dans les analyses des données d'une thèse d'il y a plus d'un quart de siècle ! Nous étions satisfait de cette concordance entre les deux approches. Dans la conséquence de la confusion que nous avions faite dans les références des tableaux, le second cas où nous avions mis en œuvre cette comparaison relevait bien d'une situation de discordance entre les deux approches. Nous rapportons le tableau de contingence :

		Variable b = Épreuve Finale (item 114)		
		1 = Réussite	0 = Échec	**Total**
Variable a = Épreuve Initiale (item 102 et/ou item 204)	1 = Réussite	40	7	47
	0 = Échec	21	34	55
	Total	61	41	102

TAB. 4 - *Tableau extrait (Régnier 1983, p. 164-166) : items 102 et 204 -- item 114*

Pour mieux illustrer notre propos, nous rappelons succinctement ces items qui relèvent du champ de la trigonométrie.

A partir d'angles **obtus** représentés graphiquement et fournis dans l'épreuve et non rapportés ici :

Item 102	Déterminer le sinus et le cosinus de l'angle (102) en utilisant le demi-cercle trigonométrique donné par report avec un calque ou par une construction.
Item 204	Déterminer le sinus et le cosinus de l'angle (204) en construisant un demi-cercle trigonométrique (unité 10 cm). Faire figurer la construction sur la feuille. Utiliser l'équerre pour construire les perpendiculaires.
Item 114	Déterminer le sinus et le cosinus de l'angle (114) à l'aide du demi-cercle trigonométrique fourni ou. par construction, sur la figure, d'un demi-cercle trigonométrique (unité 10 cm)

Si nous revenons à la perspective de recherche de lien par rejet de l'indépendance en appliquant le test du χ^2 d'indépendance, nous trouvons une valeur empirique de 23,21 qui est encore très largement supérieure à la valeur critique de 3,84 pour un niveau de risque α=0,05 et même à la valeur critique 6,63 pour un niveau de risque α=0,01. Au sens du test du χ^2 d'indépendance, il existe donc un lien fort entre les deux variables.

Si nous nous plaçons dans la perspective de recherche de lien au sens de l'ASI, le calcul de l'intensité d'implication $\varphi_P(a,b)$ avec le modèle de Poisson et le calcul de l'intensité d'implication $\varphi_{BIN}(a,b)$ avec le modèle binomial

$\chi^2=23{,}21$		$\chi^2_{MC}=7$		Intensités d'implication
a-b	b=1	b=0		$\varphi_P(a,b)$
a=1	40	7	47	0,9983
a=0	21	34	55	$\varphi_{BIN}(a,b)$
	61	41	102	0,9993

TAB. 5 - *analyse selon les trois perspectives: χ^2 d'indépendance, χ^2 Mac Nemar, ASI,*

Les valeurs qui figurent dans le tableau ci-dessus, indiquent un niveau de confiance en l'implication statistique (Réussir Item 102 et/ou Item 204)$\Rightarrow$(Réussir Item 114) supérieur à 0,99. Ce que nous avions pris en compte à l'époque par la mise en œuvre de l'indice d'implication malgré la valeur empirique du χ^2 Mac Nemar erronée (3,45 au lieu de 7 !). Là nous nous étions étonné de la discordance des conclusions entre les deux approches. Nous avions alors tenté de rechercher diverses configurations de tableaux de contingences pour établir une comparaison.

2.4 Comparaison des approches : test de Mac Nemar, ASI et test d'indépendance

Nous avons répété cette recherche en essayant de produire un exemple pour chacun des huit cas possibles que nous identifions dans les tableaux suivant :

Au seuil de risque α	Test du χ^2 Mac Nemar		
	Décision…	Rejet de Ho	Non rejet de Ho
test du χ^2 d'indépendance	Rejet de Ho	Cas 1	Cas 2
	Non rejet de Ho	Cas 3	Cas 4

TAB. 6 - *Cas dans la logique des tests statistiques d'hypothèses*

Au niveau de confiance 1- α		Cas 1	Cas 2	Cas 3	Cas 4
	Décision…				
Analyse Statistique Implicative	Quasi-implication retenue	Conf_1	Conf_2	Conf_3	Conf_4
	Quasi-implication non retenue	Conf_5	Conf_6	Conf_7	Conf_8

TAB. 7 - *Configurations issues du croisement de la logique des tests et de la logique ASI*

Le tableau suivant fournis les tableaux de contingences correspondant à chacune des huit configurations possibles quant à la prise de décision pour un risque de 1ère espèce d'une valeur α = 0,05 et pour un niveau de confiance 1- α en la quasi-implication supérieur à la valeur minimale requise dans la théorie de l'ASI, à savoir que l'intensité d'implication soit supérieure à 0,50.

Nous pouvons constater dans les configurations 6 5, 6, 7 et 8 que les valeurs de l'intensité d'implication sont inférieures à 0,50. Dans celles-ci la présomption de quasi-implication ne peut absolument pas être retenue. L'approche Mac Nemar conduit cependant à conclure que dans la configuration 7, l'hypothèse de symétrie de changement d'états doit être rejetée à un seuil de risque bien inférieur à 0,05 et même inférieur à 0,005. Cette dissymétrie aurait pu alors être interprétée comme une tendance à une implication de réussite entre a et b , malgré le non rejet de l'indépendance entre les deux variables a et b.

χ^2 =	9,65	χ^2_{MC} =	7,81	
a-b	b=1	b=0		$\varphi_P(a,b)$
a=1	36	10	46	0,9628
a=0	27	29	56	$\varphi_{BIN}(a,b)$
	63	39	102	0,9742

TAB. 8 - *Configuration 1*

χ^2 =	34,56	χ^2_{MC} =	3,2	
a-b	b=1	b=0		$\varphi_P(a,b)$
a=1	56	6	62	0,9996
a=0	14	26	40	$\varphi_{BIN}(a,b)$
	70	32	102	0,9998

TAB.. 9 - *Configuration 2*

χ^2 =	1,75	χ^2_{MC} =	21,3	
a-b	b=1	b=0		$\varphi_P(a,b)$
a=1	38	8	46	**0,7519**
a=0	40	16	56	$\varphi_{BIN}(a,b)$
	78	24	102	0,7667

TAB.. 10 - *Configuration 3*

χ^2 =	2,96	χ^2_{MC} =	2,81	
a-b	b=1	b=0		$\varphi_P(a,b)$
a=1	30	16	46	**0,797**
a=0	27	29	56	$\varphi_{BIN}(a,b)$
	57	45	102	0,826

TAB. 11- *Configuration 4*

χ^2 =	4,005	χ^2_{MC} =	5,22	
a-b	b=1	b=0		$\varphi_P(a,b)$
a=1	24	22	46	0,101
a=0	40	16	56	$\varphi_{BIN}(a,b)$
	64	38	102	0,081

TAB.. 12 - *Configuration 5*

χ^2 =	4,22	χ^2_{MC} =	0,068	
a-b	b=1	b=0		$\varphi_P(a,b)$
a=1	34	28	46	**0,1318**
a=0	30	10	56	$\varphi_{BIN}(a,b)$
	64	38	102	0,1026

TAB. 13 - *Configuration 6*

χ^2 =	0,259	χ^2_{MC} =	8,01	
a-b	b=1	b=0		$\varphi_P(a,b)$
a=1	29	17	46	**0,320**
a=0	38	18	56	$\varphi_{BIN}(a,b)$
	67	35	102	0,310

TAB.. 14 - *Configuration 7*

χ^2 =	0,019	χ^2_{MC} =	1,23	
a-b	b=1	b=0		$\varphi_P(a,b)$
a=1	24	22	46	**0,413**
a=0	30	26	56	$\varphi_{BIN}(a,b)$
	54	48	102	0,409

TAB. 15 - *Configuration 8*

Face à ces configurations, il nous semble qu'un paradoxe surgisse puisque le même tableau de contingence est susceptible d'être interprété de manière contradictoire. Une façon de lever ce paradoxe est de considérer la logique sous-jacente à chacune des trois approches : approche ASI, approche χ^2 Mac Nemar, approche χ^2 d'indépendance.

3 Conclusion

Comme nous avons pu le voir au travers des propos tenus tout au long de ce qui précède, ceux-ci s'appuient sur un point de vue soutenu par I.-C. Lerman (1992) appliqué à l'étude d'une certaine relation de dépendance orientée entre des variables descriptives. Ce point de vue oppose la logique des tests statistiques, comme celui dit du χ^2 d'indépendance ou encore celui du χ^2 de Mac Nemar, à celle des méthodes classificatoires de la manière suivante : pour les premiers, dit I.-C. Lerman, « relativement à l'existence d'un lien, on a FAUX, jusqu'à preuve du contraire » par le rejet de l'hypothèse nulle ; pour les secondes, « pour l'optique des données, on a VRAI, jusqu'à preuve du contraire », c'est-à-dire vrai selon une certaine échelle de probabilité du lien. Pour terminer nous pouvons rappeler que le test de Mac Nemar se généralise à des variables catégorielles qui ont plus de deux modalités. Si k est ce nombre, le test s'appuie sur un tableau de contingence de dimension k x k. (Pupion et Pupion, 1998 p.94).

Références

Lerman I.-C (1992) Conception et analyse de la forme limite d'une famille de coefficients statistiques d'association entre variables relationnelles. *Mathématiques, Informatique et Sciences Humaines* (118)

Pupion G. et P.-C. Pupion (1998) *Tests non paramétriques. Avec applications à l'économie et à la gestion.* Paris : Economica

Régnier, J.C. (1980) *Élaboration d'un livret auto-correctif. Étude préliminaire : questionnaire sur l'équation du second degré. Projet de livret autocorrectif.* Mémoire de DEA de Didactique de mathématiques Université Nancy 1- ULP Strasbourg. Directeur du mémoire : Georges Glaeser. Irem de Nancy. 172 p.

Régnier J.C. (1983) *Étude didactique d'un test autocorrectif en trigonométrie*. Thèse de doctorat en mathématiques (mention : didactique des mathématiques) ULP Strasbbourg Directeur de thèse : Georges Glaeser. Irem de Strasbourg Tome 1 : 307 p. Tome 2 : 50 p. suivies des Annexes.

Summary

In this chapter we try to compare three approaches to study the likely links between two binary variables, which are: The SIA, the test of Mac Nemar and the test of independence based on the measurement of $\chi 2$. For this purpose, we make a return on data from our past work in the field of mathematics education