L'année psychologique

L'utilisation des contrastes dans l'analyse des données : Comment tester les hypothèses spécifiques dans la recherche en psychologie ?

Markus Brauer, Gary McClelland

Abstract

Summary : The use of contrasts in data analysis : How to test specific hypotheses in psychological research

Omnibus tests with multiple degrees of freedom test relatively vague questions whereas most hypotheses that we derive from our theoretical models make quite precise predictions. In order to respond to this level of precision, we suggest to test specific contrasts rather than to perform omnibus tests. Two conditions have to be satisfied before one can claim that a given contrast is a parsimonious description of the observed means : 1/ the contrast itself has to explain a significant part of the variance, and 2/ if one statistically controls for the effects of this contrast, the residual variance has to be non-significant. Using specific examples, this article presents the analyses that allow researchers to test these two conditions in different experimental designs.

Key words : contrast, omnibus test, residual variance, degree of freedom.

Résumé

Résumé

Les tests omnibus à plusieurs degrés de libertés ne fournissent que des réponses vagues alors que la plupart des hypothèses que nous dérivons de nos modèles théoriques font des prédictions relativement précises. Pour répondre à ce niveau de précision, nous suggérons de tester des contrastes spécifiques plutôt que d'effectuer des tests omnibus. Deux conditions se doivent d'être satisfaites avant que l'on puisse affirmer qu'un contraste donné est une description parcimonieuse des moyennes observées : 1/ le contraste lui-même doit expliquer une partie significative de la variance, et 2/ si l'on contrôle statistiquement les effets de ce contraste, la variance intergroupe résiduelle doit être non significative. À l'aide d'exemples concrets, l'article présente les analyses permettant de tester ces deux conditions avec différents plans expérimentaux.

Mots clés : contraste, test omnibus, variance résiduelle, degré de liberté.

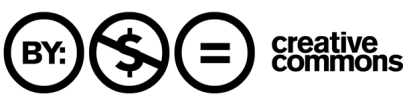
Citer ce document / Cite this document :

Brauer Markus, McClelland Gary. L'utilisation des contrastes dans l'analyse des données : Comment tester les hypothèses spécifiques dans la recherche en psychologie ?. In: L'année psychologique. 2005 vol. 105, n°2. pp. 273-305.

doi : 10.3406/psy.2005.29696

http://www.persee.fr/doc/psy_0003-5033_2005_num_105_2_29696

Document généré le 17/10/2015


BY:
creative
commons

L'Année psychologique, 2005, *105*, 273-305

NOTES MÉTHODOLOGIQUES

*CNRS et Université de Clermont-Ferrand, France**
*University of Colorado at Boulder, États-Unis***

L'UTILISATION DES CONTRASTES DANS L'ANALYSE DES DONNÉES : COMMENT TESTER DES HYPOTHÈSES SPÉCIFIQUES DANS LA RECHERCHE EN PSYCHOLOGIE ?

Markus BRAUER*[1], Gary MCCLELLAND**

Persée creative commons

SUMMARY : The use of contrasts in data analysis : How to test specific hypotheses in psychological research

Omnibus tests with multiple degrees of freedom test relatively vague questions whereas most hypotheses that we derive from our theoretical models make quite precise predictions. In order to respond to this level of precision, we suggest to test specific contrasts rather than to perform omnibus tests. Two conditions have to be satisfied before one can claim that a given contrast is a parsimonious description of the observed means : 1 / the contrast itself has to explain a significant part of the variance, and 2 / if one statistically controls for the effects of this contrast, the residual variance has to be non-significant. Using specific examples, this article presents the analyses that allow researchers to test these two conditions in different experimental designs.

Key words : contrast, omnibus test, residual variance, degree of freedom.

1. Correspondance : Markus Brauer, Laboratoire de Psychologie sociale et cognitive, Université Blaise-Pascal, 34, avenue Cournot, 63037 Clermont-Ferrand Cedex, E-mail : brauer@srupsy.univ-bpclermont.fr.

L'UTILISATION DES CONTRASTES DANS L'ANALYSE DES DONNÉES : COMMENT TESTER DES HYPOTHÈSES SPÉCIFIQUES DANS LA RECHERCHE EN PSYCHOLOGIE ?

La plupart des travaux de recherche en psychologie adoptent une logique déductive et sont guidés par des modèles théoriques. Le but d'une étude expérimentale est alors de tester une hypothèse spécifique. La majorité de nos modèles théoriques permet de faire des prédictions précises sur la taille relative des moyennes. Or souvent les analyses statistiques effectuées ne reflètent pas ce niveau de précision. Il n'est pas rare de voir des articles où l'auteur a fait une prédiction précise dans l'introduction (par exemple, « les scores dans la condition tristesse seront plus faibles que ceux dans la condition joie, et les scores de la condition contrôle seront au milieu ») mais où une autre prédiction qui, elle, est beaucoup plus vague, est testée dans la partie résultats (retraduit en langage parlé, la prédiction testée pourrait être décrite comme « il y a des différences entre les différentes conditions expérimentales »).

Avec des variables indépendantes à plus de deux modalités, la meilleure façon de tester une prédiction précise est d'utiliser des contrastes. Or l'examen des revues scientifiques en psychologie révèle qu'il y a une certaine inconsistance quant à l'utilisation des contrastes dans l'analyse des données. Dans certains cas, des tests omnibus à plusieurs degrés de liberté sont effectués alors qu'un contraste spécifique aurait été plus approprié. Dans d'autres cas, un contraste significatif est considéré comme la validation d'une hypothèse théorique alors que les moyennes sont clairement en contradiction avec cette interprétation. L'objectif de cet article est de revoir le rôle des contrastes dans l'analyse de données et de spécifier les tests statistiques nous permettant de tester des prédictions théoriques.

Quand ils ont affaire à une variable indépendante à plus de deux modalités, de nombreux chercheurs ont l'habitude d'effectuer un test omnibus avec plusieurs degrés de libertés. Si ce test est significatif et si les moyennes sont dans la direction prédite, les chercheurs déclarent que les résultats prouvent l'hypothèse théorique. Certains chercheurs vont plus loin et effectuent, après

le test omnibus, des comparaisons *post-hoc* pour contraster certaines conditions expérimentales à d'autres (parfois en omettant les ajustements de type Scheffé, malheureusement). Alors que des comparaisons *post-hoc* sont plus appropriées qu'un simple test omnibus, nous suggérons néanmoins que cette pratique est insuffisante pour tester des hypothèses précises. De plus, si un chercheur effectue de nombreuses comparaisons *post-hoc* (comme c'est le cas quand il compare toutes les conditions expérimentales deux par deux), l'ajustement de Scheffé peut l'amener à ne pas rejeter l'hypothèse nulle alors qu'en réalité, les résultats sont parfaitement compatibles avec l'hypothèse théorique.

Dans cet article, nous allons revoir l'utilisation des contrastes dans l'analyse de données. Comme nous allons le démontrer ci-dessous, un contraste spécifique est souvent plus approprié qu'un test omnibus à plusieurs degrés de libertés. Mais un contraste donné doit satisfaire certaines conditions avant que l'on puisse considérer que l'hypothèse précise correspondant à ce contraste soit validée.

LE CAS D'UNE VARIABLE INDÉPENDANTE À TROIS MODALITÉS

En accord avec d'autres (Abelson, 1995 ; Judd, 2000 ; Judd et McClelland, 1989 ; Rosenthal et Rosnow, 1985 ; Rosenthal, Rosnow, et Rubin, 2000), nous suggérons de tester les hypothèses théoriques à l'aide de contrastes spécifiques. Avec cette approche, le test omnibus à plusieurs degrés de liberté va devenir superflu car il ne permet pas de tirer des conclusions sur l'hypothèse théorique. Pour commencer avec un exemple simple, imaginons un neuropsychologue faisant l'hypothèse selon laquelle un certain neurotransmetteur influencerait les capacités de mémorisation. Quinze singes sont attribués à trois conditions expérimentales : condition « placebo » (on leur injecte une solution saline), condition « neurotransmetteur élevé » (on leur injecte le neurotransmetteur en question), et condition « neurotransmetteur faible » (on leur injecte une substance dont on sait qu'elle empêche la production naturelle du neurotransmetteur en question). La variable dépendante est le score que les singes obtiennent à un test de mémorisation. Les résultats de cette étude fictive sont montrés dans la figure 1. Par

rapport à la condition « placebo » (M = 41,0, s = 12,0, N = 5), les capacités de mémorisation sont relativement bonnes dans la condition « neurotransmetteur élevé » (M = 51,0, s = 14,0, N = 5), et relativement mauvaises dans la condition « neurotransmetteur faible » (M = 33,0, s = 13,0, N = 5)[1].

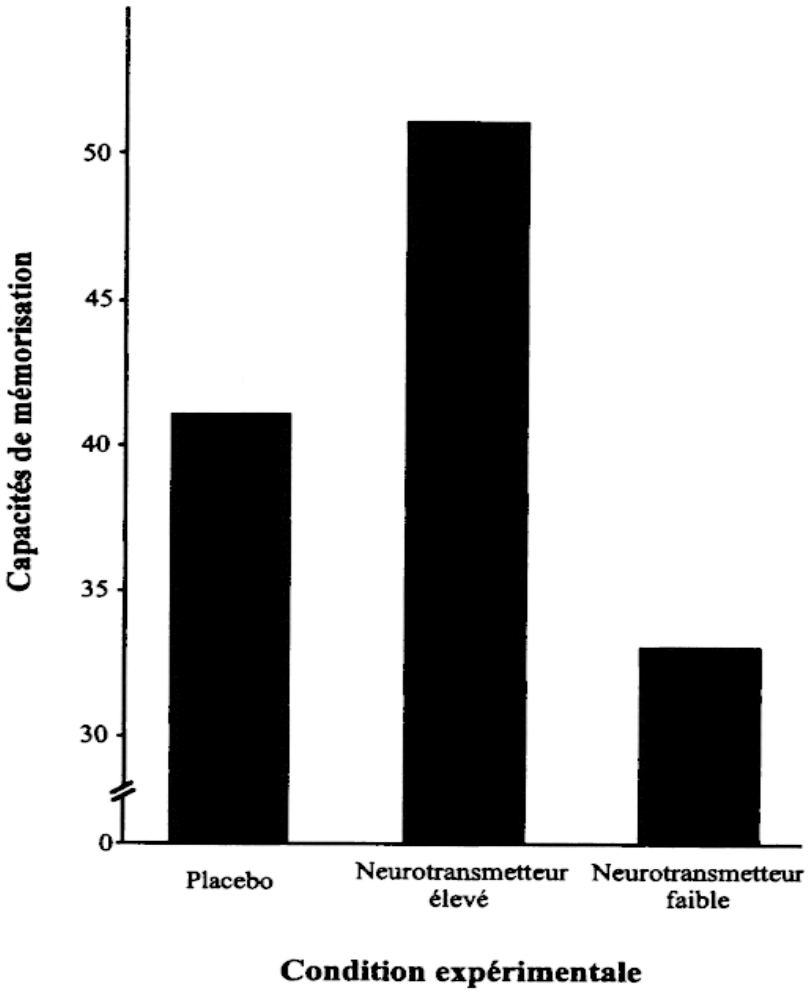


Fig. 1. — Les résultats de l'étude fictive sur l'effet d'un neurotransmetteur particulier sur les capacités de mémorisation

The results of the hypothetical study on the effects of a particular neurotransmitter on memorization

1. Voici les données brutes : 20, 24, 29, 40, 52 dans la condition « neurotransmetteur faible » ; 26, 34, 43, 44, 58 dans la condition « placebo » ; et 33, 44, 50, 58, 70 dans la condition « neurotransmetteur élevé ». Pour simplifier la présentation et pour permettre aux lecteurs de refaire les analyses avec leur propre logiciel d'analyse de données, nous utilisons un très petit nombre de participants dans les exemples de cet article.

Les analyses inférentielles révèlent que le test omnibus examinant l'effet principal de la condition expérimentale n'est pas significatif, $F(2,12) = 2,40$, $p = .13$. Le neuropsychologue s'arrête probablement ici et tire la conclusion que les résultats ne confirment pas l'hypothèse théorique. Quelques jours plus tard, des collègues conseillent au neuropsychologue d'effectuer des comparaisons *post-hoc* (en réalisant l'ajustement de Scheffé). Mais, à sa grande déception, la différence entre « placebo » et « neurotransmetteur élevé » n'est pas significative ($p = .64$), tout comme la différence entre « neurotransmetteur faible » et « placebo » ($p = .50$). Même la différence entre les deux conditions extrêmes, « neurotransmetteur élevé » et « neurotransmetteur faible », n'est pas significative ($p = .13$)[1]. À nouveau, le neuropsychologue est amené à douter de sa prédiction théorique concernant l'effet du neurotransmetteur en question sur les capacités de mémorisation.

Nous suggérerions au neuropsychologue d'effectuer d'autres analyses. En effet, son hypothèse opérationnelle est relativement exacte : il prédit que les capacités de mémorisation des singes varient de manière systématique avec la quantité du neurotransmetteur en question dans le corps. En d'autres mots, il prédit qu'en comparaison à la condition contrôle, les singes ont de meilleures performances de mémorisation si l'on leur a injecté le neurotransmetteur et de moins bonnes performances si l'on leur a injecté une substance dont on sait qu'elle empêche la production naturelle du neurotransmetteur. À cette prédiction correspond un contraste spécifique que l'on appellera C1 (placebo = 0 ; neurotransmetteur élevé = 1 ; neurotransmetteur faible = – 1). C'est ce contraste que le neuropsychologue devrait utiliser pour tester son hypothèse.

Rappelons que la logique sous-jacente à toute analyse de données est la suivante : DONNÉES = MODÈLE + ERREUR (ou encore MESSAGE = SIGNAL + BRUIT ; Abelson et Prentice, 1997). Il convient donc de démontrer deux choses, d'un côté que le MODÈLE (l'hypothèse) explique une partie significative des DONNÉES (de la variance), d'un autre côté que l'ERREUR (la variance résiduelle) est tellement petite qu'elle n'explique

1. Le neuropsychologue aurait pu s'épargner le travail supplémentaire. Par définition, si le test omnibus n'est pas significatif, aucune comparaison *post-hoc* avec l'ajustement de Scheffé ne peut être significative.

qu'une partie non significative des DONNÉES (de la variance)[1]. Plus précisément, deux conditions doivent être satisfaites avant que l'on puisse affirmer qu'un contraste spécifique fournit une description satisfaisante des moyennes : 1 / le contraste lui-même doit être significatif, et 2 / l'effet résiduel – la partie de la variance qui reste une fois que l'on a enlevé la partie de la variance expliquée par le premier contraste – doit être non significatif. Pour pouvoir tester ces deux conditions, il est nécessaire de créer $m - 1$ contrastes orthogonaux centrés, m étant le nombre de conditions expérimentales. L'annexe contient des explications détaillées sur la définition et l'utilisation des contrastes orthogonaux centrés. Rappelons ici simplement que plusieurs contrastes sont centrés et orthogonaux si pour chacun des contrastes, la somme de toutes les valeurs est égale à zéro (*i.e.*, chacun des contrastes est centré autour de zéro) et si pour chaque paire de deux contrastes possible, la somme des produits des valeurs pour chaque condition est égale à zéro (*i.e.*, les contrastes ne sont pas corrélés entre eux).

Pour tester l'hypothèse du neuropsychologue, il convient donc de créer $m - 1$ contrastes centrés orthogonaux. Le premier de ces contrastes correspond toujours à l'hypothèse théorique, alors que les « autres » contrastes, nécessairement orthogonaux entre eux et au premier contraste, sont souvent dépourvus de sens. Leur seul rôle est de tester l'effet résiduel après avoir enlevé la variance expliquée par le premier contraste. Dans le cas présent, le neuropsychologue devra créer deux contrastes. Le premier est le contraste C1 mentionné plus haut. Pour le deuxième contraste, supposons que le neuropsychologue choisit le contraste C2 (placebo = – 2 ; neurotransmetteur élevé = 1 ; neurotransmetteur faible = 1)[2].

La prochaine étape consiste à effectuer une analyse de régression multiple où la variable dépendante est régressée sur l'ensemble des $m - 1$ contrastes. La décomposition de la variance de

1. Ce qui constitue « une partie significative de la variance » dépend en partie du nombre de participants. C'est la limitation de toute inférence basée sur le test de l'hypothèse nulle. Il est toujours conseillé de considérer non seulement la significativité d'un contraste mais aussi la proportion de la variance expliquée par ce contraste.

2. D'autres contrastes orthogonaux possibles auraient été – 1/2, – 1/2, + 1 ; ou + 387, + 387, – 774 ; ou encore – 3π, – 3π, + 6π ; mais tous ces contrastes donnent les mêmes valeurs F, valeurs p et tailles d'effet.

cette analyse de régression est montré dans le tableau I[1]. On constate deux résultats intéressants : 1 / le premier contraste C1 est statistiquement différent de zéro, $F(1,12) = 4,77, p < .05$, et 2 / le deuxième contraste C2 n'est pas significatif, $F(1,12) = .02$, ns. Cette deuxième analyse montre que si l'on contrôle statistiquement l'effet du premier contraste, il ne reste plus de variance qu'un autre contraste pourrait expliquer. Un examen des sommes des carrés révèle que le premier contraste explique (810,00/813,33 =) 99,6 % de la variance intergroupe alors que le deuxième contraste n'en explique que (3,33/813,33 =) 0,4 %[2]. Somme toute, les deux conditions spécifiées plus haut sont satisfaites, et les résultats semblent confirmer l'hypothèse du chercheur : il y a un effet du neurotransmetteur sur les capacités de mémorisation.

TABLEAU I. — *La décomposition de la variance de l'étude fictive sur l'effet d'un neurotransmetteur particulier sur les capacités de mémorisation*

The source table for the hypothetical study on the effects of a particular neurotransmitter on memorization

Source de variation	*b*	Somme des carrés (*SC*)	*ddl*	*F*	*p* <
Modèle		813,33	2	2,40	.14
C1 (0, + 1, – 1)	9,00	810,00	1	4,77	.05
C2 (– 2, + 1, + 1)	0,33	3,33	1	0,02	.90
Erreur		2 036,00	12		
Total		2 849,33	14		

1. On aurait pu obtenir les mêmes résultats en passant par le module « ANOVA » du logiciel de traitement de données et en spécifiant les contrastes appropriés. Rappelons au passage que l'analyse de variance est un cas particulier de l'analyse de régression et que pour cette raison cela ne fait pas de différence si l'on passe par le module « régression » ou par le module « ANOVA » dans le logiciel de traitement de données.

2. Toute comparaison entre conditions expérimentales ne peut expliquer au mieux que la variance intergroupe. La variance à l'intérieur de chaque groupe expérimental sera toujours de la variance résiduelle.

Cet exemple démontre que les tests omnibus à plusieurs degrés de liberté sont d'une utilité limitée quand il s'agit de tester des hypothèses spécifiques. L'hypothèse nulle étant soumise à l'épreuve dans le test omnibus effectué par le neuropsychologue est la suivante : $\mu_1 = \mu_2 = \mu_3$, μ étant la moyenne de la population dans chacune des trois conditions expérimentales. Cette hypothèse nulle peut être falsifiée de plusieurs manières (par exemple parce que $\mu_1 = \mu_2 \neq \mu_3$, ou $\mu_1 \neq \mu_2 \neq \mu_3$, ou encore $\mu_1 = \mu_3 \neq \mu_2$). Les tests omnibus examinent donc des prédictions relativement grossières telles que « il y aura des différences entre les conditions expérimentales ». Ce manque de précision va de pair avec une puissance statistique faible car la variance expliquée par la variable indépendante à plusieurs modalités est divisée par $m - 1$ avant de constituer le numérateur dans le calcul du F. Ainsi, un test omnibus peut être non significatif alors que le contraste spécifique correspondant à l'hypothèse théorique peut satisfaire les deux conditions spécifiées plus haut. Cette configuration est d'autant plus probable qu'il y a un nombre important de groupes expérimentaux dans l'expérience[1].

Dans certains livres de statistiques on peut lire qu'on ne peut examiner les contrastes spécifiques que si le test omnibus est significatif. Ce n'est pas exact. Le test omnibus et le contraste spécifique décrivant l'hypothèse théorique testent deux questions différentes. Il n'y a pas de raison pour laquelle on aurait le droit de tester la question pertinente uniquement après avoir vérifié si l'on peut répondre par l'affirmative à une autre question qui, elle, n'est pas pertinente. De manière similaire, certains experts exigent toujours que l'auteur d'un manuscrit rapporte d'abord le test omnibus à plusieurs degrés de liberté et ensuite les contrastes spécifiques. Étant donné que le test omnibus n'est que marginalement pertinent pour une hypothèse théorique précise, nous proposons d'en finir avec cette pratique.

1. Notons qu'un contraste spécifique ne peut jamais « rattraper » un plan expérimental inapproprié. Mieux vaut réfléchir sur les contrastes avant la mise en place de l'expérience et choisir le plan expérimental en fonction (voir McClelland, 1997).

UNE VARIABLE INDÉPENDANTE AVEC PLUS DE TROIS MODALITÉS

Prenons l'exemple d'une psychologue expérimentaliste faisant l'hypothèse selon laquelle différentes opérations cognitives sont effectuées dans des modules cognitifs qui sont relativement spécifiques et indépendants les uns des autres. Pour tester son hypothèse elle demande à ses participants d'effectuer une tâche de détection de mouvement sur ordinateur (voir Christie et Klein, 1995). Elle crée quatre conditions expérimentales : 1 / condition « contrôle » où seuls les stimuli de la tâche de détection de mouvement apparaissent sur l'écran ; 2 / condition « chat immobile » où le dessin d'un chat apparaît en tant que distracteur sur l'écran en même temps que les stimuli ; 3 / condition « chat lent » où le chat se déplace lentement sur l'écran ; et 4 / condition « chat rapide » où le chat se déplace rapidement sur l'écran. La variable dépendante est un score de performance reflétant la capacité des participants à détecter correctement un mouvement. Selon l'hypothèse de la psychologue expérimentaliste, il y a détérioration de la performance uniquement si le distracteur utilise les ressources du même module que celui dont on se sert pour effectuer la tâche (ici : le module « mouvement »). Par conséquent, la performance dans la tâche de détection de mouvement devrait être relativement bonne dans les conditions « contrôle » et « chat immobile » et relativement mauvaise dans les conditions « chat lent » et « chat rapide ». Les résultats de cette étude fictive sont montrés dans la figure 2. Les moyennes sont comme suit : $M = 75{,}0$, $s = 36{,}0$, $N = 10$ dans la condition « contrôle », $M = 52{,}0$, $s = 34{,}0$, $N = 10$ dans la condition « chat immobile », $M = 45{,}0$, $s = 32{,}0$, $N = 10$ dans la condition « chat lent », et $M = 25{,}0$, $s = 32{,}0$, $N = 10$ dans la condition « chat rapide »[1].

Le tableau II montre différentes décompositions de la variance de cette étude fictive. Le test omnibus à trois degrés de liberté est statistiquement significatif, $F(3,36) = 3{,}78$, $p < .02$. Mais ce test indique simplement qu'au moins une des quatre

1. Voici les données brutes : 16, 38, 49, 65, 68, 83, 85, 92, 119, 135 dans la condition « contrôle » ; 0, 7, 32, 44, 49, 55, 60, 72, 97, 104, dans la condition « chat immobile » ; 0, 6, 19, 35, 39, 50, 55, 66, 79, 101, dans la condition « chat lent » ; 0, 2, 2, 5, 9, 14, 29, 33, 55, 101, dans la condition « chat rapide ».

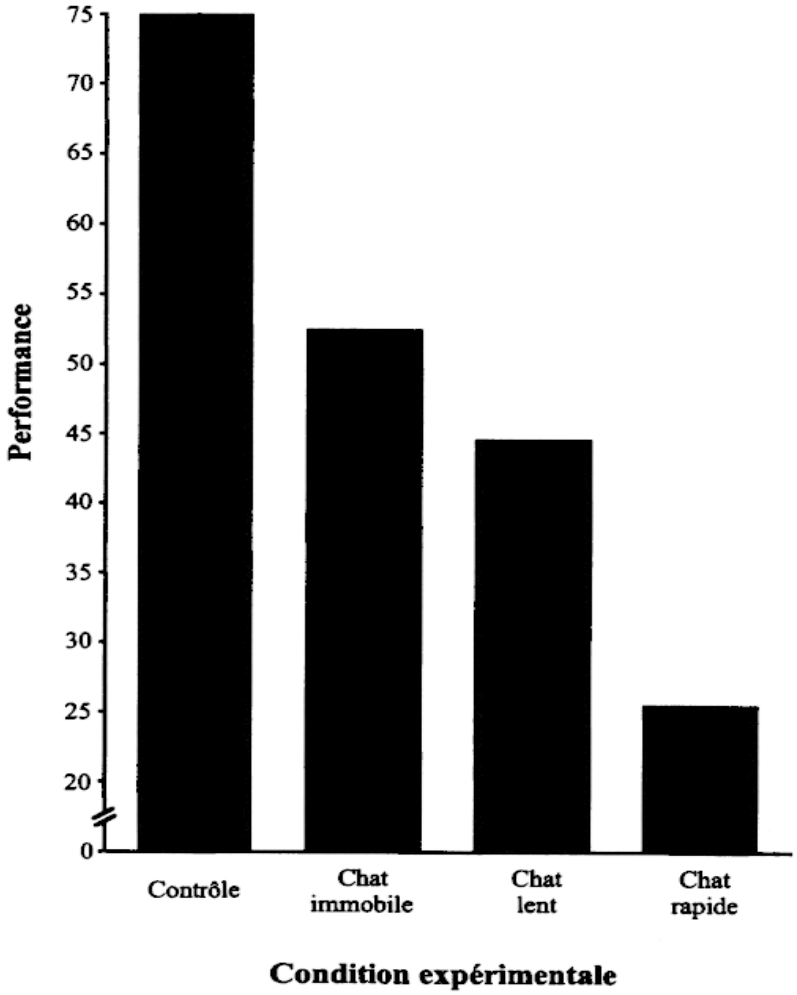


Fig. 2. — Les résultats de l'étude fictive sur la spécificité des modules cognitifs

The results of the hypothetical study on the specificity of cognitive modules

moyennes est différente des autres. Consciente du fait que les tests omnibus ne fournissent que de réponses vagues, la psychologue crée un contraste qui correspond à son hypothèse principale : C1 (+ 1, + 1, – 1, – 1, pour les conditions « contrôle », « chat immobile », « chat lent », et « chat rapide », respectivement). Elle crée aussi deux autres contrastes qui lui permettent d'examiner deux questions d'importance secondaire : C2 (+ 1, – 1, 0, 0) lui permet de voir si la présence d'un stimulus distracteur immobile est suffisant pour produire une détérioration la performance, et C3 (0, 0, + 1, – 1) qui répond à la question à savoir si la vitesse de mouvement du chat fait une diffé

TABLEAU II. — *La décomposition de la variance de l'étude fictive sur la spécificité des modules cognitifs*

Different source tables of the hypothetical study on the specifiticy of cognitive modules

Source de variation	*b*	Somme des carrés *(SC)*	*ddl*	*F*	*p* <
MODÈLE DE DÉPART					
Modèle		12 767,50	3	3,78	.02
C1 (+ 1, + 1, – 1, – 1)	14,25	8 122,50	1	7,22	.02
C2 (+ 1, – 1, 0, 0)	11,50	2 645,00	1	2,35	.14
C3 (0, 0, + 1, – 1)	10,00	2 000,00	1	1,78	.20
Erreur		40 500,00	36		
Total		53 267,50	39		
MODÈLE RESPÉCIFIÉ					
Modèle		12 767,50	3	3,78	.02
C4 (+ 3, + 1, – 1, – 3)	7,85	12 324,50	1	10,96	.002
C5 (+ 1, – 1, – 1, + 1)	0,75	22,50	1	0,02	.89
C6 (+ 1, – 3, + 3, – 1)	1,45	420,50	1	0,37	.55
Erreur		40 500,00	36		
Total		53 267,50	39		
MODÈLE DE DÉPART AVEC D'AUTRES CONTRASTES ORTHOGONAUX					
Modèle		12 767,50	3	3,78	.02
C1 (+ 1, + 1, – 1, – 1)	14,25	8 122,50	1	7,22	.02
$C2_b$ (+ 1, – 1, + 1, – 1)	10,75	4 622,50	1	4,11	.05
$C3_b$ (+ 1, – 1, – 1, + 1)	0,75	22,50	1	0,02	.89
Erreur		40 500,00	36		
Total		53 267,50	39		

rence. Ensuite, elle régresse sa variable dépendante sur les trois contrastes.

Les analyses révèlent que le contraste correspondant à l'hypothèse théorique (C1) est statistiquement significatif, $F(1,36) = 7,22$, $p < .02$ (voir la partie supérieure du tableau II). En revanche, ni C2 ni C3 ne sont significatifs, $F(1,36) = 2,35$, ns,

et $F(1,36) = 1{,}78$, ns. À partir de ces résultats, la psychologue expérimentaliste affirme que son hypothèse est confirmée. Il y a donc bien détérioration de la performance si le stimulus distracteur est en mouvement (C1). En revanche, il n'y a pas d'indication dans les données que la présence ou l'absence d'un stimulus distracteur immobile fait une différence (C2), ou que la vitesse de déplacement du stimulus distracteur joue un rôle (C3).

Se peut-il que les conclusions de la psychologue expérimentaliste soient erronées parce qu'elle n'a pas effectué les tests appropriés ? Rappelons-nous les différentes étapes que la psychologue aurait dû traverser. Son hypothèse théorique était précise et pouvait être décrite par le contraste C1 (+ 1, + 1, – 1, – 1, pour les conditions « contrôle », « chat immobile », « chat lent » et « chat rapide », respectivement). Pour pouvoir examiner ce contraste, la psychologue avait besoin de $m - 1$ contrastes orthogonaux en tout. Comme il y avait quatre conditions expérimentales, il lui fallait, outre le contraste C1, encore deux « autres » contrastes. Comme nous l'avons déjà dit, ces « autres » contrastes sont souvent dépourvus de sens, leur seul rôle étant d'utiliser tous les degrés de libertés. La psychologue avait donc le choix entre plusieurs ensembles d' « autres » contrastes, *e.g.* l'ensemble $C2_a$ (+ 1, – 1, 0, 0) et $C3_a$ (0, 0, + 1, – 1), l'ensemble $C2_b$ (+ 1, – 1, + 1, – 1) et $C3_b$ (+ 1, – 1, – 1, + 1), ou encore l'ensemble $C2_c$ (– 3, + 3, – 1, + 1) et $C3_c$ (+ 1, – 1, – 3, + 3). Dans chacun des ensembles, les deux contrastes sont orthogonaux entre eux et au premier contraste. La psychologue a choisi le premier ensemble et, puisque le choix n'a aucune importance, nous n'avons rien à reprocher à la façon dont elle a procédé jusqu'ici.

En revanche, avant de pouvoir affirmer que son hypothèse était confirmée, la psychologue aurait dû tester les deux conditions que nous avons spécifiées plus haut. De toute évidence, la première condition est satisfaite car le contraste d'intérêt (C1) est statistiquement significatif, $F(1,36) = 7{,}22$, $p < .02$. Constatons au passage que ce contraste capte (8 122,50/12 767,50 =) 64 % de la variance intergroupe. Alors que les calculs pour tester la seconde condition seront expliqués dans les prochains paragraphes, anticipons déjà le résultat en révélant que l'erreur résiduelle est significative, $F(1,36) = 4{,}13$, $p < .05$. Par conséquent, cette seconde condition n'est pas satis-

faite car si le premier contraste était une description parcimonieuse et satisfaisante des moyennes observées, l'erreur résiduelle devrait être non significative. Cette erreur résiduelle constitue ((2 000 + 2 645)/12 767,50 =) 36 % de la variance intergroupe. En somme, on peut conclure que l'hypothèse de la psychologue est infirmée car une des deux conditions pour tester un contraste spécifique n'est pas satisfaite.

Pour tester la seconde condition (l'erreur résiduelle), il convient de calculer d'abord la somme des carrés expliquée par les « autres » contrastes en tant qu'ensemble, et ensuite calculer le F selon la formule $F = (SC_{\text{« autres » contrastes}}) / (SC_{\text{erreur}}/ddl_{\text{erreur}})$. Avec des N's égaux, la somme des carrés expliquée par les « autres » contrastes en tant qu'ensemble s'obtient en additionnant les sommes des carrés individuelles, donc : $SC_{\text{« autres » contrastes}} = SC_{C2} + SC_{C3} = 2\,645{,}00 + 2\,000{,}00 = 4\,645{,}00$ (voir le tableau II). Le F est alors $F = (4\,645{,}00) / (40\,500{,}00/36) = 4{,}13$. Avec des N's égaux, un autre moyen d'obtenir la valeur F est d'additionner simplement les valeurs F associées aux deux « autres » contrastes, donc ici $F = F_{C2} + F_{C3} = 2{,}35 + 1{,}78 = 4{,}13$.

Avec des N's inégaux, il n'est pas possible d'additionner les sommes des carrés des « autres » contrastes car dans un tel cas la somme des carrés associée à plusieurs contrastes en tant qu'ensemble n'est pas égale à la somme des sommes des carrés. Pour la même raison, on ne peut pas additionner les valeurs F des « autres » contrastes pour obtenir le F total. Dans un tel cas, il convient de tester plusieurs contrastes non pas individuellement mais en tant qu'ensemble. Pour cela, on peut utiliser l'option « test » dans la plupart des logiciels de traitement de données. En SPSS, par exemple, la commande :

```
REGRESSION
    /STATISTICS COEFF OUTS R ANOVA F
    /DEPENDENT perform
    /METHOD=TEST (C1) (C2, C3)
```

permet d'obtenir la valeur F du test omnibus à plusieurs degrés de liberté testant les contrastes C2 et C3 en tant qu'ensemble. En SAS, on obtient le même test avec la commande :

```
proc reg ;
    model perform = c1 c2 c3 ;
    test c2=0, c3=0 ;
```

Le logiciel de traitement de données affichera une valeur F basée sur plusieurs degrés de libertés (ici : $F = 2{,}06$). Pour obtenir le F du test à un degré de liberté mentionné ci-dessus, il faut alors multiplier la valeur F affichée par le nombre d' « autres » contrastes ($m - 2$). Dans l'exemple présent, la psychologue expérimentaliste devra donc multiplier la valeur F affichée par 2, donc $F(1{,}36) = 4{,}13$.

Indépendamment de la façon dont la valeur F a été obtenue, cette valeur devra être comparée à la valeur critique de F pour $N - m$ degrés de libertés à $\alpha = .05$. Dans ce cas précis, la valeur critique pour (40 – 4 =) 36 degrés de liberté est de 4,11 (voir une table de F). La valeur F que nous avons calculée excède cette valeur, donc $F(1{,}36) = 4{,}13$, $p < .05$. En somme, la valeur F testant les « autres » contrastes en tant qu'ensemble est significative à $\alpha = .05$, ce qui signifie que la seconde condition n'est pas satisfaite. On est alors obligé de conclure que le premier contraste ne fournit qu'une description suboptimale des moyennes observées, et que l'hypothèse de la psychologue expérimentaliste est infirmée.

En vue de ces résultats, la psychologue expérimentaliste pourrait être amenée à reconsidérer son hypothèse de départ. Certes, nous procédons souvent de manière déductive en testant si oui ou non une hypothèse théorique est exacte, mais rien ne nous empêche de formuler une nouvelle hypothèse théorique à partir de résultats observés (méthode inductive, voir McClelland, Judd et Brauer, 2001, pour plus de détails sur la construction de modèles). Bien sûr, il convient de mettre la nouvelle hypothèse théorique à l'épreuve dans des études ultérieures qui ont été conçues spécifiquement pour tester cette nouvelle hypothèse (et non pas pour tester une autre hypothèse comme c'est le cas avec l'étude dont on est en train d'interpréter les résultats).

Au vu des résultats montrés dans la figure 2, et du matériel expérimental utilisé par la psychologue expérimentaliste, on pourrait avancer l'hypothèse alternative selon laquelle la diminution de la performance dépend de la force de distraction des stimuli distracteurs. Un chat se déplaçant rapidement est plus distrayant qu'un chat se déplaçant lentement qui, à son tour, est plus distrayant qu'un chat immobile, lui-même étant plus distrayant que l'absence d'un stimulus distracteur. Le contraste correspondant à cette nouvelle hypothèse est C4 (+ 3, + 1, – 1, – 3, pour les conditions « contrôle », « chat immo-

bile », « chat lent », et « chat rapide », respectivement). Parmi les contrastes orthogonaux possibles, nous avons choisi C5 (+ 1, – 1, – 1, + 1) et C6 (+ 1, – 3, + 3, – 1). Sans que cela ait une influence sur l'interprétation des résultats, remarquons que les contrastes C4, C5 et C6 correspondent respectivement à la tendance linéaire, la tendance quadratique, et la tendance cubique des contrastes polynomiaux. La décomposition de la variance en fonction de ces contrastes est montrée dans la partie moyenne du tableau II.

La première condition est satisfaite, car le contraste d'intérêt est statistiquement significatif, $F(1,36) = 10,95$, $p < .002$. Ce contraste explique (12 324,50/12 767,50 =) 97 % de la variance intergroupe. La seconde condition est aussi satisfaite, car la variance résiduelle est non significative, $F(1,36) = .02 + .37 = .39$, ns (3 % de la variance intergroupe). Ces résultats montrent que la nouvelle hypothèse faisant référence à la force de distraction des stimuli distracteurs fait des prédictions particulièrement exactes. On n'a donc pas besoin de supposer une quelconque spécificité des modules cognitifs pour pouvoir interpréter les résultats observés, et cela, ne l'oublions pas, malgré le fait que le premier contraste de la psychologue expérimentaliste (C1) était statistiquement significatif.

Le lecteur intéressé est invité à consulter Niedenthal, Brauer, Robin et Innes-Ker (2002) pour un autre exemple où des contrastes spécifiques ont été utilisés pour tester les effets d'une variable indépendante à quatre modalités. Une présentation détaillée des études effectuées par ces auteurs dépasserait le cadre de cet article mais cela pourrait être intéressant pour certains lecteurs de savoir que ces auteurs ont testé des prédictions plus complexes (par exemple C1 : – .25, + .75, + .75, – 1,25) que celle utilisée dans l'exemple de l'étude fictive présentée ici.

CONSIDÉRATIONS TECHNIQUES SUR LE TEST DE LA VARIANCE RÉSIDUELLE

Le lecteur averti aura remarqué que le test proposé ci-dessus pour vérifier la seconde condition n'est pas tout à fait le même que celui suggéré par Abelson et Prentice (1997). Évoquons quelques considérations techniques sur la différence entre les deux tests. Ces considérations sont pertinentes avant tout pour

les lecteurs avec un intérêt avancé en statistiques. Nous suggérons aux lecteurs moins spécialisés n'ayant pas cet intérêt de passer immédiatement à la prochaine section de l'article (voir « L'interaction entre deux variables interparticipant »).

Abelson et Prentice proposent de tester la variance intergroupe résiduelle (la seconde condition) avec un test omnibus avec $m - 2$ degrés de liberté. Plus précisément, ils suggèrent de diviser la somme des carrés expliquée par les « autres » contrastes en tant qu'ensemble par les degrés de liberté utilisés par ces contrastes. Appliqué à l'hypothèse initiale de la psychologue dans l'étude fictive sur la spécificité des modules cognitifs, cela donne $F = (SC_{\text{« autres » contrastes}}/ddl_{\text{« autres » contrastes}}) / (SC_{\text{erreur}}/ddl_{\text{erreur}})$ = (4 645,00/2) / (40 500,00/36) = 2,06, donc $F(2,36) = 2,06$, ns. Avec des N's égaux, un autre moyen d'obtenir la valeur F est de faire la moyenne des valeurs F associés aux « autres » contrastes, donc ici $F = (F_{C2} + F_{C3}) / 2 = (2,35 + 1,78) / 2 = 2,06$.

Or le test proposé par Abelson et Prentice (1997) peut induire de fausses conclusions. Étant donné l'imprécision et le manque de puissance statistique des tests omnibus, le test sur la seconde condition proposé par ces auteurs peut être non significatif même s'il reste encore de la variance résiduelle à expliquer. Pour illustrer ce point, imaginons que la psychologue ait choisit un autre ensemble d' « autres » contrastes orthogonaux, $C2_b$ (+ 1, – 1, + 1, – 1) et $C3_b$ (+ 1, – 1, – 1, + 1), en plus de son contraste d'intérêt C1 (+ 1, + 1, – 1, – 1). Le panneau inférieur du tableau II montre la décomposition de la variance si l'on utilise les contrastes C1, $C2_b$ et $C3_b$. La somme des carrés et le F associés au contraste C1 n'ont évidemment pas changé. En revanche, cette analyse révèle que le contraste $C2_b$ capte une partie significative de la variance résiduelle, $F(1,36) = 4,11$, $p = .05$. En d'autres mots, les données ne correspondent pas à l'hypothèse de la psychologue expérimentaliste. Après avoir enlevé la variance expliquée par le premier contraste (celui correspondant à l'hypothèse), il reste encore suffisamment de variance intergroupe résiduelle pour qu'un deuxième contraste puisse être statistiquement significatif. Or le test proposé par Abelson et Prentice révèle un résultat non significatif, $F(2,36) = 2,06$, ns, et nous aurait amenés à tirer la conclusion fausse selon laquelle la seconde condition serait satisfaite.

Nous proposons de tester la totalité de la variance résiduelle avec un test à un degré de liberté. Comme le lecteur s'en est

peut-être déjà rendu compte, notre test est très similaire à celui d'Abelson et Prentice (1997). Le F se calcule de la même manière, la seule différence étant que les $ddl_{\text{« autres » contrastes}}$ sont fixés à 1. La formule devient alors $F = (SC_{\text{« autres » contrastes}}/1) / (SC_{\text{erreur}}/ddl_{\text{erreur}}) = (4\,645{,}00/1) / (40\,500{,}00/36) = 4{,}13$, donc $F(1{,}36) = 4{,}13$, $p < .05$. Le F de notre test est le F d'Abelson et Prentice multiplié par le nombre d' « autres » contrastes. Avec deux « autres » contrastes, le F de notre test sera simplement le double du F du test d'Abelson et Prentice. Avec trois « autres » contrastes (cinq conditions expérimentales), notre F correspondra à trois fois celui d'Abelson et Prentice, et ainsi de suite.

La force de notre test est qu'il est le test le plus conservateur que l'on puisse imaginer : si le résultat de ce test n'est pas significatif, on peut être certain qu'il n'y a aucun « autre » contraste orthogonal – aussi optimal qu'il soit – qui peut être significatif. Cette garantie n'existe pas avec le test d'Abelson et Prentice. Dans l'étude fictive, leur test n'était pas significatif, mais il existait un « autre » contraste ($C2_b$) qui expliquait une partie significative de la variance. Si C1 et $C2_b$ sont tous les deux significatifs, qu'est-ce qui nous permet de dire que c'est C1, et non pas $C2_b$, qui décrit les phénomènes psychologiques ayant déterminé les moyennes observées ? Le test que nous proposons règle ce problème car un résultat non significatif nous garantit que parmi les nombreux « autres » contrastes possibles, il n'y en a aucun qui peut être significatif.

En résumé, nous proposons d'utiliser la méthode d'Abelson et Prentice (1997) qui consiste à tester aussi bien le contraste d'intérêt (condition 1) que la variance résiduelle (condition 2), mais nous conseillons de vérifier la seconde condition avec un test plus conservateur que celui proposé par ces auteurs. Comme cette seconde condition est satisfaite si on ne peut pas rejeter l'hypothèse nulle, il convient d'utiliser le test le plus conservateur possible.

L'INTERACTION ENTRE DEUX VARIABLES INTERPARTICIPANT

Dans la majorité des plans expérimentaux, plusieurs variables indépendantes sont croisées entre elles pour examiner leur effet interactif. Si un des ces variables a plus de deux modalités, il convient d'utiliser des contrastes. Imaginons un psychologue

du travail qui veut tester l'hypothèse que la promesse d'une récompense incite les gens à penser que la tâche à effectuer demande de l'effort (*e.g.*, Kemmelmeier, Bless, Schwarz et Bohner, 2004). Il présente à ses participants une tâche et leur promet (ou non) une récompense s'ils réussissent à effectuer la tâche correctement. Cette récompense varie entre les conditions expérimentales : 0, 1, 2, 3 ou 4 bouteilles d'une boisson non alcoolisée de leur choix (variable « Récompense »). Le psychologue dit à la moitié des participants que la tâche est facile, à l'autre moitié qu'elle est difficile (variable « Difficulté »). L'expérience a donc un plan 5 × 2. Le psychologue fait deux hypothèses : 1 / Plus la récompense est élevée plus les participants penseront qu'ils devront dépenser de l'effort pour réussir la tâche, mais l'accroissement de l'effort nécessaire perçu entre 2 niveaux de récompense diminue avec chaque bouteille supplémentaire. 2 / Cet effet est plus prononcé dans la condition « difficile » que dans la condition « facile ». Les résultats de cette étude fictive sont montrés dans la figure 3[1].

L'hypothèse du psychologue peut être représentée par le contraste C1 (– 4, – 1, 1, 2, 2, pour les conditions 0, 1, 2, 3 et 4 bouteilles respectivement). Comme la variable « récompense » a 5 modalités, il convient de créer ($m - 2 =$) 3 « autres » contrastes orthogonaux. Parmi les nombreux « autres » contrastes que le psychologue a à sa disposition, imaginons qu'il choisisse les contrastes C2 (0, 0, 0, – 1, 1), C3 (2, – 3, – 3, 2, 2) et C4 (2, – 6, + 6, – 1, – 1). Si le psychologue souhaite passer par le module « ANOVA » de son logiciel de traitement, il suffit d'effectuer une analyse de variance avec deux variables inter-participant et de spécifier les contrastes C1 à C4 mentionnés ci-dessus. S'il souhaite passer par le module « Régression », il lui faut effectuer trois opérations : Il doit créer les contrastes C1 à C4 dans la base de données en donnant aux conditions leur valeur dans chaque contraste (en recodant quatre fois la variable « récompense » ou en utilisant des expressions conditionnelles). Il doit recoder la variable « difficulté » en – 1 et 1

1. Voici les données brutes. Condition facile : 9, 10, 38 (pas de récompense) ; 3, 13, 23 (1 bouteille) ; 8, 27,4 (2 bouteilles) ; 20, 31, 42 (3 bouteilles) ; 4, 8, 24 (4 bouteilles). Condition difficile : 10, 15, 29 (pas de récompense) ; 29, 37, 63 (1 bouteille) ; 25, 44, 51 (2 bouteilles) ; 28, 49, 61 (3 bouteilles) ; 51, 63, 69 (4 bouteilles).

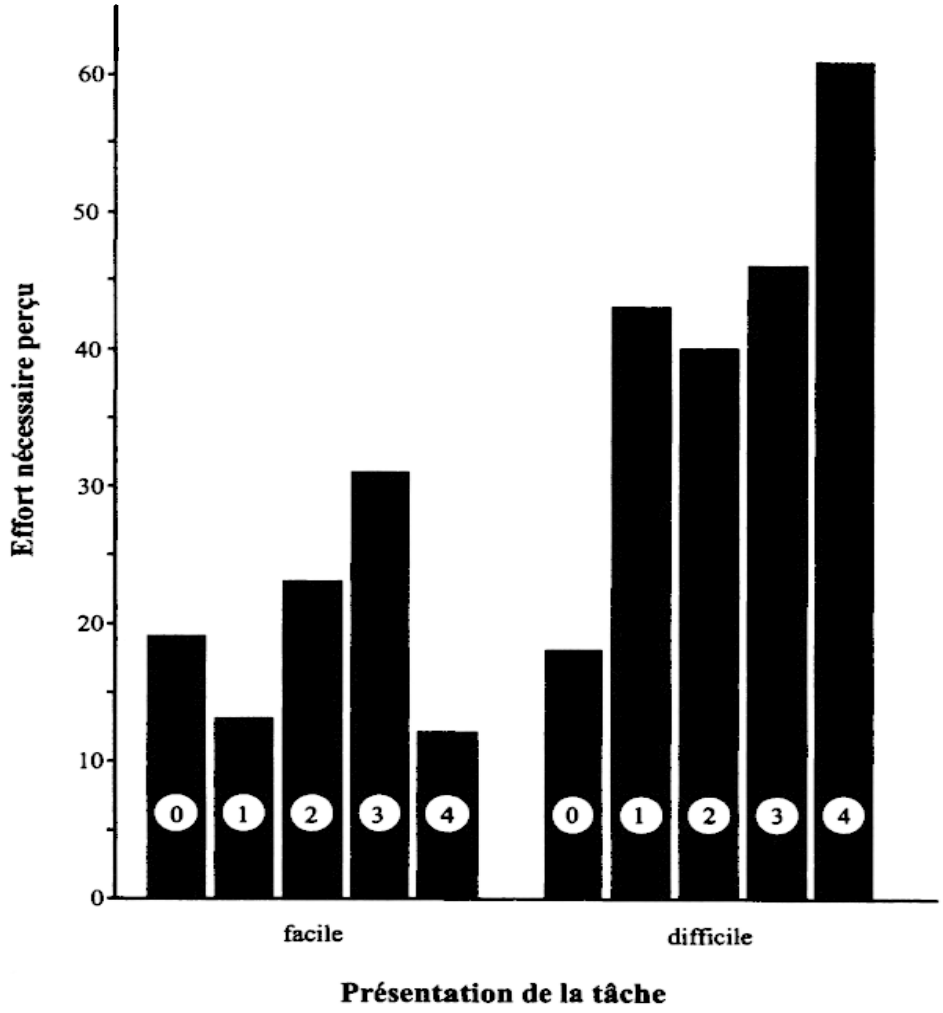


Fig. 3. — Les résultats de l'étude fictive sur la perception d'effort nécessaire pour effectuer correctement une tâche

The results of the hypothetical study on the perception of effort necessary to correctly accomplish a task

pour qu'elle soit centrée autour de zéro. Et il doit créer quatre variables d'interaction en multipliant les contrastes C1 à C4 avec la variable « difficulté » recodée. Ensuite il effectue une analyse de régression linéaire avec effort en tant que variable dépendante et avec neuf variables indépendantes (les quatre contrastes, la variable « difficulté » recodée, et les quatre variables d'interaction). L'analyse de variance et l'analyse de régression produisent évidemment les mêmes résultats. La décomposition de la variance est montrée dans le tableau III.

TABLEAU III. — *La décomposition de la variance de l'étude fictive sur la perception d'effort nécessaire pour effectuer correctement une tâche*

The source table for the hypothetical study on the perception of effort necessary to correctly accomplish a task

Source de variation	*b*	Somme des carrés *(SC)*	*ddl*	*F*	*p* <
Récompense		1 507,20	4	2,16	.12
C1 (– 4, – 1, 1, 2, 2)	3,06	1 458,58	1	8,37	.009
C2 (0, 0, 0, – 1, 1)	– 1,00	12,02	1	0,07	.80
C3 (2, – 3, – 3, 2, 2)	0,28	14,46	1	0,08	.78
C4 (2, – 6, + 6, – 1, – 1)	– 0,22	22,30	1	0,13	.73
Difficulté	10,20	3 630,00	1	20,84	.001
Récompense × Difficulté		2 094,00	4	3,01	.05
C1 × Difficulté	2,67	817,00	1	4,69	.05
C2 × Difficulté	8,00	866,99	1	4,98	.04
C3 × Difficulté	– 0,52	11,32	1	0,07	.81
C4 × Difficulté	– 0,99	398,74	1	2,29	.15
Erreur		3 484,00	20		
Total		2 858,80	29		

Pour chacune de ses hypothèses, le psychologue de travail devra tester les deux conditions spécifiées par Abelson et Prentice (1997). Le contraste C1 correspond à la première hypothèse. Il s'avère que ce contraste est statistiquement significatif, $F(1,20) = 8,37$, $p < .009$. Ce contraste capte 1 458,58/1 507,20 = 97 % de la variance intergroupe que la variable « récompense » peut expliquer. L'effet des « autres » contrastes, en tant qu'ensemble, n'est pas significatif car $F(1,20) = .07 + .08 + .13 = .28$, ns. Ici et dans les exemples ultérieurs, nous utilisons « notre » test (celui à *1* degré de liberté) et non pas celui initialement proposé par Abelson et Prentice (à $m - 2$ degrés de libertés).

L'interaction entre le contraste C1 et la variable « difficulté » correspond à la deuxième hypothèse du psychologue.

Cette interaction est statistiquement significative, $F(1,20) = 4,69$, $p < .05$, tout comme l'interaction entre les « autres » contrastes et la variable « difficulté », $F(1,20) = 4,98 + .07 + 2,29 = 7,34$, $p < .02$. L'interaction du contraste d'intérêt explique 817,00/2 094,00 = 39 % de la variance intergroupe que l'interaction entre les variables « récompense » et « difficulté » peut expliquer. Il s'avère donc que pour la première hypothèse les deux conditions d'Abelson et Prentice sont satisfaites alors que pour la deuxième hypothèse ce n'est pas le cas (ici, la première condition est satisfaite mais pas la seconde). Le psychologue de travail conclura alors que sa première hypothèse est confirmée mais que sa deuxième hypothèse ne l'est pas.

Notons au passage, que pour chacune des deux hypothèses, le test omnibus à quatre degrés de liberté fournit le résultat contraire. Le test omnibus de la variable « récompense » n'est pas significatif, $F(4,20) = 2,16$, ns, alors que les deux conditions pour le contraste C1 sont satisfaites. Le test omnibus de l'interaction entre « récompense » et « difficulté » est significatif, $F(4,20) = 3,01$, $p < .05$, alors que le test précis de l'hypothèse de modulation (l'interaction entre C1 et « difficulté ») nous indique que les données ne confirment pas cette hypothèse. Cet exemple illustre à nouveau le fait que les tests omnibus ne sont pas pertinents pour tester des hypothèses théoriques précises. Il n'y a pas de lien entre le résultat d'un test omnibus et le résultat de l'analyse appropriée. Notons également, que le test initialement proposé par Abelson et Prentice fournit une réponse insatisfaisante quant à l'interaction entre le contraste C1 et la variable « difficulté ». Comme nous l'avons vu ci-dessus, la seconde condition n'est pas satisfaite quand on teste les « autres » contrastes avec un test (conservateur) à un degré de liberté, $F(1,20) = 4,98 + .07 + 2,29 = 7,34$, $p < .02$. Abelson et Prentice suggèrent de tester ces « autres » contrastes avec un test à $m - 2$ degrés de liberté, donc $F(3,20) = (4,98 + .07 + 2,29) / 3 = 2,45$, *ns*. Leur test aurait donc mené à la conclusion que les deux conditions sont satisfaites pour la deuxième hypothèse du psychologue du travail. Une inspection des moyennes montrées dans la figure 3, le fait que l'interaction entre C1 et « difficulté » n'explique que 39 % de la variance explicable par l'interaction, et le fait que l'interaction entre un autre contraste (C2) et « difficulté » n'est pas seulement

significative mais explique aussi plus de variance que l'interaction entre C1 et « difficulté » montrent à quel point une telle conclusion serait aberrante.

Deux commentaires s'imposent à ce stade. Premièrement, l'exemple précédent peut facilement être adapté à un plan expérimental contenant une variable indépendante continue. Dans l'étude fictive décrite ci-dessus, la variable « Difficulté » était une variable catégorielle à deux modalités (facile *vs* difficile). Mais on pourrait imaginer une étude où le psychologue ne dit rien sur la difficulté de la tâche mais demande à ses participants, après la présentation de la tâche, d'évaluer la difficulté perçue sur une échelle à 9 points. Dans ce cas, on se retrouverait avec une variable « Difficulté » continue. L'article de Brauer (2002) contient une description détaillée des analyses avec des variables indépendantes continues et catégorielles. Nous présenterons un bref résumé des analyses suggérées pour l'exemple présent : Après la création de 4 variables correspondant aux 4 contrastes C1 à C4, il convient de transformer la variable continue en forme centrée, de calculer 4 variables d'interaction en multipliant chacun des 4 contrastes avec la variable continue en forme centrée, et d'effectuer une analyse de régression. Dans cette analyse, les 4 contrastes, la variable continue en forme centrée et les 4 variables d'interaction sont les variables indépendantes et l'effort perçu est la variable dépendante. « L'ouput de cette analyse de régression ressemblera au tableau III, et l'interprétation des valeurs F sera la même. »

Deuxièmement, plus on dévie des contrastes « standard » (contrastes polynomiaux, contrastes Helmert, voir annexe), plus il devient nécessaire de justifier ses contrastes théoriquement. Considérons à nouveau l'exemple précédent où le psychologue du travail a choisi le contraste C1 (– 4, – 1, 1, 2, 2) pour tester ses hypothèses. En choisissant ce contraste, il fait implicitement la supposition que l'augmentation de l'effort perçu entre 2 niveaux de récompense diminue de manière constante (la différence entre – 4 et – 1 est 3, celle entre – 1 et 1 est 2, celle entre 1 et 2 est 1 et celle entre 2 et 2 est 0). Or on pourrait supposer que l'hypothèse du psychologue du travail est mieux représentée par un contraste où l'augmentation diminue de manière exponentielle. Dans ce cas, il faudrait utiliser le contraste $C1_b$ (– 4,9, – 0,9, 1,1, 2,1, 2,6). Notons aussi qu'il est difficile de prédire à partir de quand une bouteille de jus de fruit supplémentaire n'a plus d'effet. Et si

l'effet plafonnait à deux bouteilles ? Dans ce cas, le contraste $C1_c$ (– 2,2, – 0,2, 0,8, 0,8, 0,8) serait approprié. Et si l'effet d'une bouteille supplémentaire était beaucoup plus important quand on n'a rien que quand on a déjà au moins une bouteille ? Dans ce cas, il convient d'utiliser le contraste $C1_d$ (– 8, 0, 2, 3, 3). Comme ces exemples le montrent, plusieurs contrastes peuvent correspondre à une hypothèse donnée. Cela ne veut pas dire qu'il faut se limiter aux contrastes standard. Au contraire, un contraste non standard peut contribuer à l'originalité d'un travail de recherche. Ces exemples ont été donnés pour montrer à quel point il est nécessaire de justifier théoriquement l'utilisation d'un contraste donné, d'autant plus qu'il s'agit d'un contraste non standard. La qualité d'un article scientifique serait moindre si le lecteur gagnait l'impression que le contraste d'intérêt (C1) a été choisi non pas en fonction des considérations théoriques mais parce qu'il satisfait les deux conditions d'Abelson et Prentice.

UNE VARIABLE INDÉPENDANTE INTRAPARTICIPANT À PLUSIEURS MODALITÉS

L'analyse d'une variable intraparticipant suit la même logique, c'est-à-dire qu'il convient de tester aussi bien le MODÈLE que l'ERREUR. Mais étant donné qu'il y a certaines particularités à considérer avec une variable intraparticipant, nous allons illustrer la façon de procéder avec deux autres exemples, un sans et un avec interaction. Imaginons qu'une psychologue clinicienne ait développé une nouvelle mesure implicite de l'anxiété. Dans un premier temps, elle souhaite tester si elle réussit à répliquer un résultat typique qu'on observe dans le traitement des individus phobiques : l'anxiété est faible au début, augmente au moment où le phobique rentre en contact avec le stimulus sur lequel porte la phobie et diminue une fois que ce stimulus n'est plus visible (Bandura, Adams, Hardy, et Howells, 1980). La psychologue clinicienne fait venir cinq participants phobiques des serpents au laboratoire et place un serpent devant eux sur une table pendant cinq minutes. Elle mesure le niveau d'anxiété des participants à quatre moments différents : moment 1 (après avoir annoncé qu'elle va leur montrer un serpent), moment 2 (immédiatement après avoir posé le serpent sur la table), moment 3 (trois minutes après avoir posé le serpent sur la table), moment 4 (immédiatement après avoir remis le serpent dans sa

caisse). La psychologue prédit une relation curvilinéaire entre le moment de mesure et l'anxiété. Les moyennes semblent correspondre aux attentes de la psychologue clinicienne : $M = 3{,}00$, $s = 1{,}58$ (moment 1), $M = 8{,}00$, $s = 3{,}16$ (moment 2), $M = 8{,}00$, $s = 2{,}12$ (moment 3) et $M = 4{,}00$, $s = 2{,}55$ (moment 4)[1].

TABLEAU IV. — *La décomposition de la variance de l'étude fictive sur la nouvelle mesure implicite d'anxiété*

The source table for the hypothetical study on the new implicit measure of anxiety

Source de variation	*b*	Somme des carrés *(SC)*	*ddl*	*F*	*p* <
Moment		103,75	3	6,92	.006
Tendance linéaire	1,0	2,25	1	0,98	.38
Tendance quadratique	− 6,0	101,25	1	50,63	.002
Tendance cubique	− 2,0	0,25	1	0,02	.89
Erreur		60,00	12		
Tendance linéaire		9,20	4		
Tendance quadratique		8,00	4		
Tendance cubique		42,80	4		
Total		2 858,80	29		

Pour analyser ses données, la psychologue clinicienne effectue une analyse de variance avec une variable intraparticipant (mesure répétée) à 4 modalités tout en spécifiant des contrastes polynomiaux. Le tableau IV contient une description de la décomposition de la variance. Le contraste d'intérêt est la tendance quadratique. Ce contraste est statistiquement significatif, $F(1,4) = 50{,}63$, $p < .002$. L'effet des « autres » contrastes, en tant qu'ensemble, est non significatif, $F(1,8) = (2{,}25 + 0{,}25) / ((9{,}20 + 42{,}80) / (4 + 4)) = 2{,}50/6{,}50 = .38$, *ns*. Le contraste d'intérêt explique 101,25/103,75 = 98 %, les « autres » contrastes

1. Voici les données brutes : participant n° 1 : 1, 4, 8, 3 ; participant n° 2 : 3, 8, 11, 2 ; participant n° 3 : 4, 12, 8, 8 ; participant n° 4 : 2, 10, 5, 5 ; participant n° 5 : 5, 6, 8, 2.

expliquent (2,25 + 0,25)/103,75 = 2 % de la variance. Les deux conditions d'Abelson et Prentice (1997) étant satisfaites, la psychologue clinicienne tire la conclusion (correcte) que son hypothèse est confirmée et qu'elle a réussi à répliquer la courbe classique avec sa nouvelle mesure d'anxiété.

Dans un deuxième temps, la psychologue clinicienne souhaite utiliser sa nouvelle mesure d'anxiété pour évaluer l'efficacité de deux types de thérapies, une thérapie psychanalytique et une thérapie cognitivo-comportementaliste. Dix individus, tous avec une phobie envers les serpents, sont attribués aléatoirement à un des deux types de thérapie. À la fin de la thérapie, la psychologue clinicienne fait venir les patients au laboratoire et leur fait faire les mêmes tâches que les participants de l'expérience précédente. Elle se retrouve avec un plan expérimental de type 4 (moment de la mesure : moment 1 *vs* moment 2 *vs* moment 3 *vs* moment 4, intraparticipant) × 2 (type de thérapie : psychanalytique *vs* cognitivo-comportementaliste, interparticipant). Elle prédit : 1 / qu'il y aurait une relation curvilinéaire entre le temps de mesure et l'anxiété, et 2 / que cette relation curvilinéaire serait moins prononcée chez les individus ayant suivi une thérapie cognitivo-comportementaliste que chez les individus ayant suivi une thérapie psychanalytique. Les résultats de cette étude fictive sont montrés dans la figure 4[1].

Pour examiner ses données, la psychologue clinicienne effectue une ANOVA 4 × 2 avec mesures répétées, tout en spécifiant des contrastes polynomiaux pour la variable intraparticipant « moment de mesure ». La décomposition de la variance se trouve dans le tableau V. La première hypothèse fait référence à une relation curvilinéaire entre le moment de mesure et l'anxiété. La tendance quadratique est significativement supérieure à zéro, $F(1,8) = 23,23$, $p < .001$, et explique (90,00/92,50 =) 97 % de la variance explicable par la seule variable « moment de mesure ». La première condition est donc largement satisfaite. La variance résiduelle, en revanche, n'est pas significative, car $F(1,16) = (0,50 + 2,00)$ /

1. Voici les données brutes : participants n^{os} 1-5 sont identiques à l'expérience précédente ; participant n^{o} 6 : 1, 5, 4, 4 ; participant n^{o} 7 : 6, 3, 2, 1 ; participant n^{o} 8 : 4, 5, 2, 5 ; participant n^{o} 9 : 1, 4, 6, 1 ; participant n^{o} 10 : 3, 8, 6, 4 ; participants n^{os} 1-5 ont suivi une thérapie psychanalytique, participants n^{os} 6-10 une thérapie cognitivo-comportementaliste.

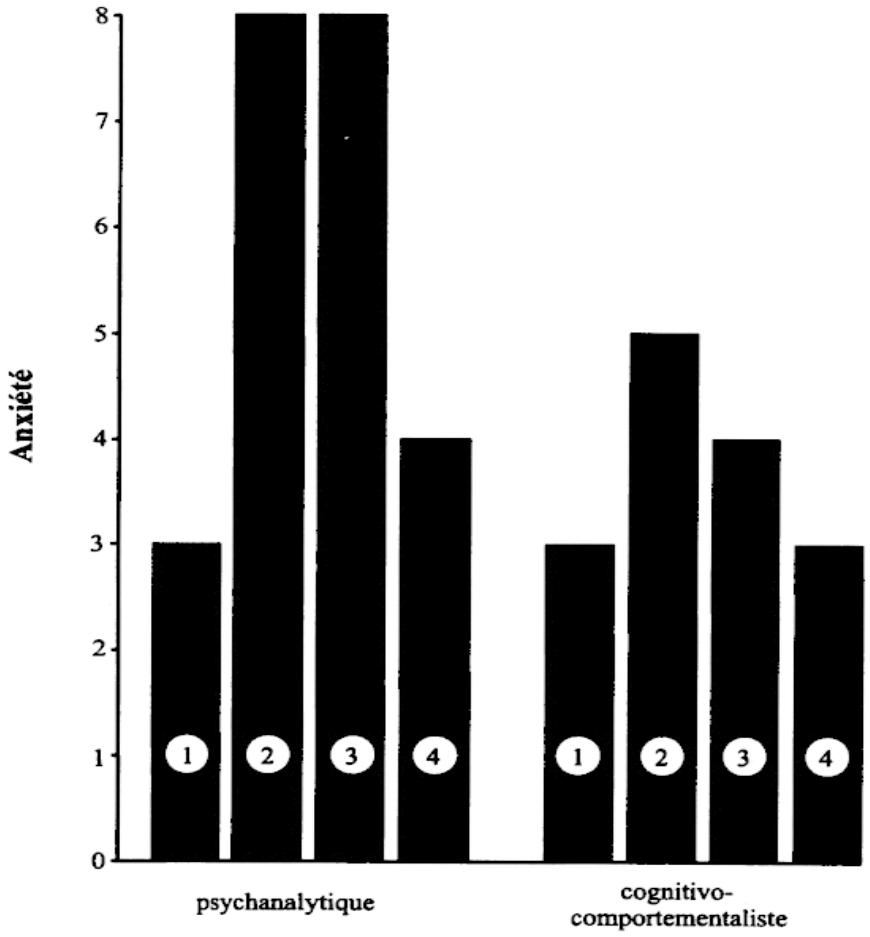


Fig. 4. — Les résultats de l'étude fictive sur l'efficacité de différentes thérapies

The results of the hypothetical study on the efficiency of differents types of therapies

((25,20 + 51,80) / (8 + 8)) = 2,50/4,81 = .52, *ns*, et n'explique que ((0,50 + 2,00)/92,50 =) 3 % de la variance. Donc, la seconde condition est également satisfaite, et on peut en conclure que la première hypothèse de la psychologue clinicienne est confirmée : elle a réussi à répliquer l'effet classique que l'on observe chez les individus phobiques et qu'elle a déjà observé dans l'expérience précédente.

La deuxième hypothèse prédit que la relation curvilinéaire entre le moment de mesure et l'anxiété dépend du type de thérapie. L'interaction entre type de thérapie et la tendance quadratique est statistiquement significative, $F(1,8) = 5,81$,

TABLEAU V. — *La décomposition de la variance de l'étude fictive sur l'efficacité de différentes thérapies*

The source table for the hypothetical study on the efficiency of differents types of therapies

Source de variation	*b*	Somme des carrés *(SC)*	*ddl*	*F*	*p* <
Test des effets interparticipant					
Modèle (thérapie)	1,0	40,00	1	6,67	.04
Erreur interparticipant		48,00	8		
Total interparticipant		88,00	9		
Test des effets intraparticipant					
Moment		92,50	3	6,85	.002
Tendance linéaire	1,0	0,50	1	0,16	.71
Tendance quadratique	− 6,0	90,00	1	23,23	.001
Tendance cubique	− 2,0	2,00	1	0,31	.60
Moment × Thérapie		25,00	3	1,85	.17
Tendance linéaire	2,0	2,00	1	0,63	.45
Tendance quadratique	− 3,0	22,50	1	5,81	.05
Tendance cubique	1,0	0,50	1	0,08	.79
Erreur		108,00	24		
Tendance linéaire		25,20	8		
Tendance quadratique		31,00	8		
Tendance cubique		51,80	8		
Total intraparticipant		225,50	30		
Total inter et intra		313,50	39		

$p < .05$. Elle explique (22,50/25,00 =) 90 % de la variance explicable par l'interaction des deux facteurs. L'examen de la figure 4 montre que cette interaction est dans « le bon sens ». En revanche, l'interaction entre le type de thérapie et les deux « autres » contrastes n'est pas significative, $F(1,16) = (2,00 + 0,50) / ((25,20 + 51,80) / (8 + 8) = 2,50/4,81 = .52$, *ns* (10 % de la variance). Ainsi, les deux conditions sont satisfaites, et on peut dire que les résultats confirment l'hypothèse de modulation

de la psychologue clinicienne : l'effet classique est moins prononcé chez les individus ayant suivi une thérapie cognitivo-comportementaliste que chez les individus ayant suivi une thérapie psychanalytique. Notons également que le test omnibus de l'interaction n'est pas significatif, $F(3,24) = 1,85$, ns. La psychologue clinicienne n'aurait pas tiré les bonnes conclusions s'il s'était fondé sur ce test omnibus (voir Abelson et Prentice, 1997, p. 324-325, pour un exemple avec une variable intraparticipant à six modalités et une variable interparticipant à deux modalités).

CONCLUSION

Nous proposons d'adopter une certaine méfiance envers les tests non spécifiques à plusieurs degrés de liberté. Certes, ces tests omnibus ont leur place quand le chercheur n'a pas d'hypothèses précises et veut explorer des données qu'il a récoltées « juste pour voir ». Mais à part ces situations bien spécifiques, les tests omnibus devraient être utilisés avec la plus grande précaution. Avec deux ou trois degrés de liberté, ils sont souvent non significatifs alors qu'il y a des résultats intéressants. Avec beaucoup de degrés de libertés, les tests omnibus sont presque toujours significatifs (surtout si une des variables est intraparticipant) perdant ainsi toute capacité de confirmation d'hypothèse. Même si c'est écrit dans certains livres d'introduction à l'analyse des données, et même si certains experts l'exigent, un test omnibus à plusieurs degrés de liberté significatif n'est pas une précondition nécessaire pour pouvoir examiner des contrastes spécifiques.

Comme les exemples dans cet article l'ont montré, le test omnibus et le contraste spécifique peuvent parfois mener à des conclusions contradictoires. Comme le formulaient les fameux statisticiens Snedecor et Cochran en 1937 : « D'ailleurs, si plusieurs comparaisons sont faites, une ou deux de ces comparaisons peuvent être significatives même si le test F global n'est pas significatif » (Snedecor et Cochran, 1937-1980, p. 225). Seul le contraste spécifique permet de confirmer une hypothèse théorique précise. Étant donné le rôle central de ce contraste, il se doit d'être testé rigoureusement. Il doit expliquer une partie significative de la variance (condition 1). Et une fois que l'on

contrôle statistiquement les effets de ce premier contraste, la variance résiduelle doit ne pas être significative (condition 2). Si ces deux conditions sont satisfaites, le chercheur peut affirmer avec confiance que son contraste donne une description parcimonieuse et satisfaisante des moyennes observées.

RÉSUMÉ

Les tests omnibus à plusieurs degrés de libertés ne fournissent que des réponses vagues alors que la plupart des hypothèses que nous dérivons de nos modèles théoriques font des prédictions relativement précises. Pour répondre à ce niveau de précision, nous suggérons de tester des contrastes spécifiques plutôt que d'effectuer des tests omnibus. Deux conditions se doivent d'être satisfaites avant que l'on puisse affirmer qu'un contraste donné est une description parcimonieuse des moyennes observées : 1 / le contraste lui-même doit expliquer une partie significative de la variance, et 2 / si l'on contrôle statistiquement les effets de ce contraste, la variance intergroupe résiduelle doit être non significative. À l'aide d'exemples concrets, l'article présente les analyses permettant de tester ces deux conditions avec différents plans expérimentaux.

Mots clés : contraste, test omnibus, variance résiduelle, degré de liberté.

RÉFÉRENCES

Abelson R. P. — (1995) *Statistics as Principled Argument*, Hillsdale (NJ), Lawrence Erlbaum Associates.

Abelson R. P., Prentice D. A. — (1997) Contrast tests of interaction hypotheses, *Psychological Methods, 2*, 315-329.

Bandura A., Adams N. E., Hardy A. B., Howells G. N. — (1969) Tests of the generality of self-efficacy theory, *Cognitive Therapy and Research, 4*, 39-66.

Christie J., Klein R. — (1995) Familiarity and attention : Does what we know affect what we notice, *Memory and Cognition, 23*, 547-550.

Judd C. M. — (2000) Everyday data analysis in social psychology : Comparisons of linear models, *in* H. T. Reis et C. M. Judd (Edit.), *Handbook of Research Methods in Social and Personality Psychology*, Cambridge (UK), Cambridge University Press, 370-392.

Judd C. M., McClelland G. H. — (1989) *Data Analysis : A Model-Comparison Approach*, San Diego (CA), Harcourt Brace Jovanovich.

Kemmelmeier M., Bless H., Schwarz N., Bohner G. — (2004) What research participants learn from rewards : A conversational logic analysis of rewarding reasoning performance, *Cahiers de psychologie cognitive / Current Psychology of Cognition.*

McClelland G. H. — (1997) Optimal design in psychological research, *Psychological Methods, 2*, 3-19.

McClelland G. H., Judd C. M., Brauer M. — (2001) *Unfettering Contrastes from ANOVA's Restrictions : Comment on Abelson and Prentice (1997)*, manuscrit non publié, University of Colorado at Boulder.

Niedenthal P. M., Brauer M., Robin L., Innes-Ker A. H. — (2002) Adult attachment and the perception of facial expression, *Journal of Personality and Social Psychology, 82,* 419-433.

Rosenthal R., Rosnow R. L. — (1985) *Contrast analysis : Focused Comparisons in the Analysis of Variance,* Cambridge (UK), Cambridge University Press.

Rosenthal R., Rosnow R. L., Rubin D. B. — (2000) *Contrasts and Effect Sizes in Behavioral Research : A Correlational Approach,* Cambridge (UK), Cambridge University Press.

Snedecor G. W., Cochran W. G. — (1937-1980) *Statistical Methods,* Menasha (WI), Collegiate Press.

Note de l'Auteur

Nous voudrions remercier Joel Swendsen, François Ric, Nadine Chaurand et Élise Perrier pour leurs commentaires sur des versions antérieures de cet article.

Toute correspondance concernant cet article peut être envoyée à Markus Brauer, Laboratoire de Psychologie sociale et cognitive, Université Blaise-Pascal, 34, avenue Carnot, 63037 Clermont-Ferrand Cedex, France. L'adresse internet est <brauer@srvpsy.univ-bpclermont.fr>.

ANNEXE

Dans les analyses présentées dans cet article, nous suggérons d'utiliser des contrastes orthogonaux centrés (angl. *contrast codes*). Rappelons les deux conditions qui doivent être satisfaites avant que l'on puisse affirmer que plusieurs contrastes sont orthogonaux et centrés (voir Judd et McClelland, 1989) : 1 / Pour chacun des contrastes, la somme de toutes les valeurs doit être égale à zéro. 2 / Pour chaque paire de deux contrastes possible, la somme des produits des valeurs pour chaque groupe expérimental doit être égale à zéro. Ces conditions peuvent être exprimés de manière suivante :

(1) $\Sigma\lambda_k = 0$;

(2) $\Sigma\lambda_{1k}\,\lambda_{2k} = 0$;

où λ est la valeur du contraste pour chacun des k groupes expérimentaux et les chiffres 1 et 2 font référence au premier et au deuxième contraste d'une paire donnée.

Considérons un exemple. Un chercheur fait une expérience avec 4 groupes expérimentaux. Il prédit que les participants dans les groupes [1] et [2] auront des scores élevés, ceux dans le groupe [3] des scores faibles, et ceux dans le groupe [4] auront des scores intermédiaires sur la variable dépendante. Le contraste C1 (+ 3, + 3, – 5, – 1) correspond à cette prédiction (voir colonne A du tableau ci-dessous). Comme il y a 4 groupes, le chercheur doit créer en tout (m – 1 =) 3 contrastes centrés et orthogonaux.

	A	B	C	D	E	F	G	H	I	J	K	L
Groupe	C1	C2	C1 * C2	$C3_a$	$C3_b$	C1 * $C3_b$	$C3_c$	C1 * $C3_c$	C2 * $C3_c$	$C3_d$	C1 * $C3_d$	C2 * $C3_d$
[1]	+ 3	+ 1	+ 3	− 2	− 3	− 9	− 4	− 12	− 4	− 1	− 3	− 1
[2]	+ 3	+ 1	+ 3	+ 3	+ 1	+ 3	+ 5	+ 15	+ 5	+ 1	+ 3	+ 1
[3]	− 5	+ 2	− 10	+ 1	− 1	+ 5	+ 1	− 5	+ 2	0	0	0
[4]	− 1	− 4	+ 4	− 1	+ 3	− 3	− 2	+ 2	+ 8	0	0	0
Somme	0	0	0	+ 1	0	− 4	0	0	+ 11	0	0	0

Le chercheur trouve rapidement un contraste C2 (+ 1, + 1, + 2, − 4) qui satisfait les deux conditions spécifiées ci-dessus. C2 est « centré » (*i.e.*, la somme des quatre valeurs est égale à zéro, voir colonne B) et orthogonal à C1 (*i.e.*, la somme des produits des valeurs pour chaque groupe expérimental est égale à zéro, voir colonne C). Ensuite il génère le contraste $C3_a$ (− 2, + 3, + 1, − 1). Or ce contraste ne convient pas car il n'est pas centré autour de zéro (*i.e.*, la somme des quatre valeurs n'est pas égale à zéro, voir colonne D). Un autre contraste $C3_b$ (− 3, + 1, − 1, + 3) satisfait la première condition (*i.e.*, il est centré, voir colonne E) mais pas la seconde condition (*i.e.*, il n'est pas orthogonal à C1, voir colonne F). Le contraste $C3_c$ paraît plus prometteur, car il est centré autour de zéro (voir colonne G) et il est orthogonal à C1 (voir colonne H). En revanche, ce contraste ne convient pas non plus car il n'est pas orthogonal à C2 (voir colonne I). Seul le contraste $C3_d$ satisfait toutes les conditions nécessaires : il est centré autour de zéro, il est orthogonal à C1 et il est orthogonal à C2 (voir colonnes J, K et L). C'est donc l'ensemble C1, C2 et $C3_d$ qui convient pour l'expérience du chercheur. C1 est le contraste d'intérêt et C2 et $C3_d$ sont les « autres » contrastes.

Les contrastes standard

Contrastes polynomiaux (3 groupes)

Groupe	Tendance linéaire	Tendance quadratique
[1]	− 1	− 1
[2]	0	+ 2
[3]	+ 1	− 1

Contrastes polynomiaux (4 groupes)

Groupe	Tendance linéaire	Tendance quadratique	Tendance cubique
[1]	− 3	+ 1	− 1
[2]	− 1	− 1	+ 3
[3]	+ 1	− 1	− 3
[4]	+ 3	+ 1	+ 1

Contrastes polynomiaux (5 groupes)

Groupe	Tendance linéaire	Tendance quadratique	Tendance cubique	Tendance quartique
[1]	− 2	+ 2	− 1	+ 1
[2]	− 1	− 1	+ 2	− 4
[3]	0	− 2	0	6
[4]	+ 1	− 1	− 2	− 4
[5]	+ 2	+ 2	+ 1	+ 1

Contrastes Helmert (3 groupes)

Groupe	[1] *vs* [2]-[3]	[2] *vs* [3]
[1]	− 2	0
[2]	+ 1	− 1
[3]	+ 1	+ 1

Contrastes Helmert (4 groupes)

Groupe	[1] *vs* [2]-[4]	[2] *vs* [3]-[4]	[3] *vs* [4]
[1]	− 3	0	0
[2]	+ 1	− 2	0
[3]	+ 1	+ 1	− 1
[4]	+ 1	+ 1	+ 1

Contrastes Helmert (5 groupes)

Groupe	[1] *vs* [2]-[5]	[2] *vs* [3]-[5]	[3] *vs* [4]-[5]	[4] *vs* [5]
[1]	− 4	0	0	0
[2]	+ 1	− 3	0	0
[3]	+ 1	+ 1	− 2	0
[4]	+ 1	+ 1	+ 1	− 1
[5]	+ 1	+ 1	+ 1	+ 1