Exemple 5
Identité de l'ego et alcoolisme chez l'adolescent

Talashek, Marie L. Parental (1987) Alcoholism and Adolescent Ego Identity. *Journal of Community Health Nursing*, 4(4), (1987), pp. 211-222.

L'*identité du moi* est le sentiment d'être distinct des autres et d'être le même au fil du temps; le sentiment que le petit que j'étais avant, c'était bien le moi d'aujourd'hui et celui que je serai après. L'identité du moi évolue chez l'adolescent/e à mesure qu'il/elle répond lui-même ou elle-même aux questions : D'où viens-je? Qui suis-je? Que deviendrai-je?

L'hypothèse de travail est que l'adolescent ayant un parent alcoolique est à risque de développer une identité du moi inadéquate.

Afin de démontrer cette hypothèse, les chercheurs font passer le *Ego Identity Scale*, un questionnaire d'identité du moi, à deux groupes d'adolescents : un groupe de 44 sujets ayant un parent alcoolique (le groupe expérimental) et un groupe de 44 sujets n'ayant pas de parent alcoolique (le groupe témoin). L'auteure parle d'un appariement fait sur la base de l'âge, du sexe, et du niveau socio-économique. Mais cette affirmation est contredite par un tableau qui montre à tout le moins que l'appariement n'a pas été fait sur la base du sexe.

Le questionnaire produit un score mesurant la force de l'identité de l'ego.

Question 1 *Facteurs ayant un effet sur l'identité du moi*
En un premier temps, on essaie de déterminer quels sont les facteurs qui affectent l'identité de l'ego. Les facteurs considérés sont l'âge, le sexe, le niveau socio-économique, la séparation parentale, et la fréquence des congés maladie. Pour chacun de ces facteurs, déterminer, à partir des données présentées, si le facteur en question contribue à l'identité de l'ego. Vous utiliserez soit une analyse de variance, soit un test d'égalité de moyennes. Dans le cas d'une analyse de variance, explorer la possibilité de grouper des catégories lorsque les données le justifient.

Bien que les données soient appariées, vous supposerez que les observations sont toutes indépendantes (voir la discussion à la fin de la section 6.6).

Au bas de chaque tableau, on présente l'écart-type de l'échantillon entier (indépendamment de la strate), la statistique F et la valeur p (vp).

1.1 *Effet de l'âge*

Strate (h)	n_h	Moyenne ($\bar{y}_h$)	Écart-type (S_h)
13 ans	12	44,16	11,15
14 ans	24	45,83	6,82
15 ans	24	45,21	6,95
16 ans	16	50,81	7,69
17 ans	10	51,40	8,95
18 ans	2	60,00	2,83

$\hat{\sigma}$ = 7,97856; F = 2,48359151; vp = 0,05006332

a) Montrer que S_1 (= 11,15, dernière colonne) est significativement différent de S_2 (6,82).
F = 2,6728894; vp = 0,0225388

[Cette conclusion signifie que les hypothèses du modèle ne sont pas toutes vérifiées.   On ignorera cette conclusion dans la suite, puisqu'elle ne risque pas trop de fausser les résultats. En pratique, on jettera un coup d'œil sur les données pour voir s'il n'y a pas eu erreur dans les données (erreur de calcul ou de transcription, ou inclusion d'un individu n'appartenant pas à la population visée).]

b) Effectuer une analyse de variance afin de déterminer si le score moyen d'identité d'ego dépend de l'âge. Exclure les deux individus âgés de 18 ans (les données étant trop peu nombreuses pour

permettre une analyse fiable. On concède donc que l'analyse ne prétend pas porter sur une population comprenant des adolescents de plus de 17 ans).
$n = 86$; $\bar{y} = 46{,}998$; $\hat{\sigma} = 8{,}2524$; $F = 2{,}4836$; $vp = 0{,}0501$

c) On pourrait tenter de simplifier le modèle en regroupant des catégories contiguës. Par exemple, on pourrait adopter, comme partie du modèle, l'hypothèse que les trois premiers groupes sont de même moyenne.

i) Avant de combiner les trois premiers groupes, tester l'hypothèse que les trois moyennes sont égales en effectuant une analyse de variance sur les seules données des trois premiers groupes, ignorant les autres groupes.
$n = 60$; $\bar{y} = 45{,}248$; $\hat{\sigma} = 7{,}8899$; $F = 0{,}1797$; $vp = 0{,}8360$

ii) En i), en utilisant l'estimation $\hat{\sigma} = 7{,}8899$ basée sur les trois premiers groupes, on se prive de la contribution que pourraient apporter les deux derniers groupes à l'estimation de σ. On pourrait considérer que l'estimation $\hat{\sigma}$ basée sur l'ensemble de toutes les données est crédible et donc préférer celle-là à celle calculée en i). Vérifier que dans ce cas, la statistique F est de loi de Fisher à 2 et 81 degrés de liberté sous H_o; et qu'on obtient $F = 0{,}1757$ et $vp = 0{,}839$.

[*Remarque* : le résultat ici n'est pas ce auquel on s'attendrait, soit une valeur p inférieure à celle obtenue en i) étant donné le fait que l'estimation de σ est plus précise, ce qui se manifeste par un nombre de degrés de liberté plus important (81 plutôt que 57). Mais le hasard a fait que $\hat{\sigma}$ est ici plus grand, ce qui réduit la valeur F.

d) Explorer le modèle qui considérerait les âges 13, 14 et 15 comme appartenant à un même groupe.

i) Vérifier les données de la première ligne du tableau suivant :

Strate (h)	n_h	Moyenne ($\bar{y}_h$)	Écart-type (S_h)
13, 14,15	60	45,25	7,78
16	16	50,81	7,69
17	10	51,40	8,95

ii) Reprendre l'analyse de variance avec ces nouvelles données
$n = 86$; $\bar{y} = 47{,}000$; $\hat{\sigma} = 7{,}899$; $F = 4{,}885$; $vp = 0{,}0099$

e) Afin de simplifier le modèle encore plus, vérifier s'il est raisonnable de combiner les deux derniers groupes.
$n = 26$; $\bar{y} = 51{,}0369$; $\hat{\sigma} = 7{,}7464$; $t = 0{,}1788$; $vp = 0{,}860$

f) On aura vu qu'en effet il n'est pas déraisonnable de combiner les groupes d'âge 16 et 17. Examiner donc le modèle qui retient deux groupes d'âge : 13 à 15 ans et 16 à 17 ans. Tester l'hypothèse qu'il n'y a pas de différence entre les deux groupes.l

Âge	n_h	Moyenne ($\bar{y}_h$)	Écart-type (S_h)
13 à 15	60	45,25	7,78
16 à 17	26	51,04	8,02

$n = 86$; $\bar{y} = 47{,}00$; $\hat{\sigma} = 7{,}8522$; $t = 3{,}14$; $vp = 0{,}0023$

g) Résumer l'ensemble de vos conclusions.

h) Les données présentées dans le tableau ne font pas cas du groupe d'appartenance des sujets. Commenter sur les effets de cette omission (voir l'exercice 6.41).

1.2 *Effet du sexe*

Strate (h)	n_h	Moyenne ($\bar{y}_h$)	Écart-type (S_h)
Filles	60	46,67	8,67
Garçons	28	48,64	7,72

$\hat{\sigma} = 8{,}3833458$; $t = 1{,}0267439S$; $pv = 0{,}3074203$

1.3 *Effet du niveau socio-économique*

Strate (h)	n_h	Moyenne ($\bar{y}_h$)	Écart-type (S_h)
Professionnel	10	52,40	7,90
Cadre	22	46,23	7,25
Ouvrier spécialisé	42	47,21	7,96
Ouvrier non spécialisé	14	45,57	10,89

$\hat{\sigma}$ = 8,3130888; F = 1,5797878; vp = 0,2003342

1.4 *Effet de la séparation parentale*

Strate (h)	n_h	Moyenne ($\bar{y}_h$)	Écart-type (S_h)
Séparés	31	46,19	6,72
Non séparés	57	47,89	9,17

$\hat{\sigma}$ = 8,3969335; t = 0,9072044; vp = 0,366833722

1.5 *Relations avec les absences maladie*

Strate (h)	n_h	Moyenne ($\bar{y}_h$)	Écart-type (S_h)
2 semaines ou plus	22	48,00	9,67
Moins de 2 semaines	66	47,06	7,99

$\hat{\sigma}$ = 8,4311806; t = 0,4528782; vp = 0,6517769.

1.6 *Effet de l'usage de l'alcool de l'adolescent*

Strate (h)	n_h	Moyenne ($\bar{y}_h$)	Écart-type (S_h)
Abstinence	18	46,89	9,10
Buveur modérés	36	49,03	7,51
Buveurs problématiques	33	45,61	8,90

$\hat{\sigma}$ = 8,392523; F = 1,455345; vp = 0,239142.

1.7 *Effet d'un parent alcoolique*

Parent alcoolique	n_h	Moyenne ($\bar{y}_h$)	Écart-type (S_h)
Oui	44	45,52	9,40
Non	44	49,06	6,90

$\hat{\sigma}$ = 8,245.0169; t = 2,01376; pv = 0,04716068

Question 2

Les deux numéros suivants concernent deux conséquences possibles sur l'adolescent d'un parent alcoolique : une consommation abusive d'alcool et des absences fréquentes à l'école.

2.1 *Consommation abusive d'alcool*

L'auteure affirme avoir effectué un test de McNemar à partir du tableau suivant (légèrement modifié pour les besoins de cet exercice):

Adolescent buveur?	Parent alcoolique	Parent non alcoolique	Total
Problématique	16	14	30
Modéré	28	30	58
Total	44	44	88

Elle a calculé la valeur $|Z| = 2{,}16$ et la valeur p, $pv = 0{,}03$ (0,015 si le test est unilatéral); elle concluait donc à un effet modérément significatif.

a) Du fait que l'effectif total est 88, on conclut qu'en réalité le test effectué n'est pas un test de McNemar. Expliquer.

On constate que les données présentées au tableau ci-dessus sont incomplètes : elles ne suffiraient pas pour effectuer un test de McNemar.

b) Le test de McNemar est normalement effectué partir des valeurs X et Y inscrites dans le format de tableau suivant. Indiquer dans les marges les noms des quantités qui seraient nécessaires pour pouvoir compléter un test de McNemar.

		Titre		
		Valeur 1	Valeur 2	Total
Titre	Valeur 1		X	
	Valeur 2	Y		
	Total			

c) Soient

p_1 : la probabilité qu'un adolescent présente un comportement problématique s'il a au moins un parent alcoolique.

p_2 : la probabilité qu'un adolescent présente un comportement problématique si aucun de ses parents n'est alcoolique.

L'objectif de l'auteure était de tester l'hypothèse $H_o : p_1 = p_2$. Pour les fins de cet exercice, supposez que les observations faites sur les deux membres d'une paire sont indépendantes. Effectuer un test classique d'égalité de proportions. Pouvez-vous rejeter H_o?

d) Le test de McNemar est essentiellement équivalent à un test d'égalité de moyennes pour données appariées, où les observations $[X_i\,;\,Y_i]$ sont des variables indicatrices définies sur la i^e paire de sujets ($i = 1, 2, \ldots, 44$) par

$X_i = 1$ si le membre du groupe expérimental est un buveur problématique, $X_i = 0$ sinon;
$Y_i = 1$ si le membre du groupe témoin est un buveur problématique, $Y_i = 0$ sinon.

Vérifier que selon les données présentées, $\sum_{i=1}^{44} X_i = 16; \sum_{i=1}^{44} Y_i = 14$.

Ce qui est nécessaire pour compléter le test, c'est la valeur de $\sum_{i=1}^{44} X_i Y_i$, qui n'est pas donnée dans l'article.

Supposons que $\sum_{i=1}^{44} X_i Y_i = 2$. Utiliser cette information pour déterminer S_x ; S_y et S_{xy} puis effectuer le test d'égalité de moyennes (avec données appariées).

e) Utiliser la supposition que $\sum_{i=1}^{44} X_i Y_i = 2$, compléter le tableau en b) et effectuer un test de McNemar.

f) Le test en c) suppose une covariance nulle entre les X_i et les Y_i, ce qu'on peut juger irréaliste étant donné l'appariement. Une objection contre la procédure en c) est qu'en supposant l'indépendance le test perd de sa puissance du fait que $\text{Cov}(X_i\,;\,Y_i)$ est vraisemblablement positive (et par conséquent S_{xy} devrait l'être aussi). Cet argument est-il défendable?

g) Si l'argument en e) est valide, la supposition (non justifiée) que $\sum_{i=1}^{44} X_i Y_i = 2$ n'est pas raisonnable : elle donne une covariance échantillonnale négative alors qu'on s'attend à ce qu'elle soit positive. Pour faire une supposition plus réaliste, posons $\sum_{i=1}^{44} X_i Y_i = 10$. Montrer qu'alors $S_{xy} = 0{,}11416$.

2.2 *Absences fréquentes*

Absences fréquentes?	Parent non alcoolique	Parent alcoolique	Total
Oui	8	14	22
Non	36	30	66
Total	44	44	88

La même problématique qu'au numéro 2.1 se pose ici. Analyser ces données à la manière du numéro précédent. Il sera nécessaire de donner arbitrairement certaines valeurs à la somme inconnue $\sum_{i=1}^{44} X_i Y_i$.