

SIMON LAFLAMME

Université Laurentienne, département de sociologie

RUN-MIN ZHOU

Université Laurentienne, département de psychologie

Méthodes statistiques en sciences humaines

Avec des illustrations tirées du logiciel *SPSS*

Cahier d'exercices

Cognitio

Prise
de parole

Méthodes statistiques en sciences humaines

Avec des illustrations tirées du logiciel *SPSS*

Cahier d'exercices

SIMON LAFLAMME
RUN-MIN ZHOU

COLLECTION COGNITIO
Éditions Prise de parole
Sudbury, 2016

Nous tenons à remercier l'International Business Machines Corporation (IBM) de nous avoir permis de reproduire des captures d'écran de résultats générés grâce au logiciel *SPSS*. Nous reconnaissons le © *SPSS*, une compagnie IBM, sur l'ensemble des résultats ainsi reproduits.

Conception de la première de couverture : Stéphane Cormier d'après la couverture du manuel produite par Olivier Lasser

Prise
de parole

Éditions Prise de parole
C.P. 550, Sudbury (Ontario)
Canada P3E 4R2
www.prisedeparole.ca

Nous reconnaissons l'aide financière du gouvernement du Canada par l'entremise du Fonds du livre du Canada (FLC) et du programme Développement des communautés de langue officielle de Patrimoine canadien. La maison d'édition remercie également la Ville du Grand Sudbury de son appui financier. Cet ouvrage a bénéficié d'une contribution de l'Université Laurentienne.

Canada

 Greater Grand
Sudbury

 Université Laurentienne
Laurentian University

Tous droits de traduction, de reproduction
et d'adaptation réservés pour tous pays.

Copyright © Ottawa, 2016

ISBN 978-2-89423-961-2 (PDF)

1 - INTRODUCTION

Ce cahier d'exercices accompagne nos *Méthodes statistiques en sciences humaines (avec des illustrations tirées du logiciel SPSS)*. Il a été conçu dans l'esprit qui a animé la rédaction du manuel. Cela signifie qu'il met l'accent davantage sur la méthode que sur la statistique. Il a pour principe moteur qu'en sciences humaines la statistique est rarement une fin en soi, qu'elle est l'un des outils avec lesquels on donne un sens à des données. Les exercices qu'il propose ne sont donc pas essentiellement des invitations à calculer ceci ou cela : une moyenne, ou un chi-carré, ou un test t, ou un Kruskal-Wallis. La première partie, celle des « notions générales », comporte un peu d'arithmétique. Mais l'essentiel des travaux qu'il propose consiste en des interprétations de résultats générés par un logiciel. Un des traits distinctifs de notre manuel est de centrer l'enseignement de la statistique sur son lien avec la méthodologie : il a donc été conçu surtout à l'usage des chercheurs virtuels ou actuels des sciences humaines. Or, à trop encadrer l'enseignement de la statistique, on perd de vue ses fonctions méthodologiques. En outre, méthodologiquement, on peut faire énormément de choses avec la statistique qui ne sont pas réellement réalisables si l'on fait trop de place au calcul : on ne peut pas travailler sur de grands échantillons, on ne peut pas faire de log-linéaires avec plusieurs variables, ni d'analyse de variance à plus d'un facteur, par exemple. Ce sont les logiciels qui ont donné accès à ces outils analytiques.

Chaque chapitre du cahier d'exercices correspond à un chapitre du manuel et comprend une ou plusieurs tâches. Les solutions détaillées se trouvent dans un cahier séparé. Hormis les questions auxquelles on répondra par vrai ou faux, et d'autres devoirs techniques simples à

solutionner, les réponses sont autre part forcément incomplètes. Dans tous ces autres cas, soit pour la plupart des exercices, on se référera à une matrice de données clairement identifiée. De manière pratique, on recourra à cette matrice pour générer quelques résultats, puis on interprétera ces résultats. Nous avons chaque fois livré les sorties que produit le logiciel afin que l'utilisateur sache s'il est sur la bonne voie. Nous avons aussi offert une interprétation en condensé, toujours pour que l'utilisateur puisse vérifier si ses observations, dans les grandes lignes, sont justifiées. Cependant, il y aura à dire sur les résultats que livre le logiciel beaucoup plus que ce que nous avons écrit, aussi bien sur les plans statistique et méthodologique qu'en ce qui a trait à la théorie. Il appartiendra au professeur de faire parler les résultats techniques s'il le désire. Les propos que nous offrons sur les produits statistiques ne sont qu'indicatifs, sommaires. Ils ne sont pas exhaustifs ; ils ne pourraient pas l'être.

Nous avons fourni une ou deux matrices pour toutes les analyses statistiques qui relèvent de la méthodologie, à des fins d'entraînement. Mais il y a, à nos yeux, deux meilleurs exercices. L'un consiste à entrer dans une matrice qui comprend plusieurs variables et à découvrir, en fonction de ce qu'on a sous les yeux, les analyses qui peuvent être effectuées et les résultats que ces analyses offrent. L'autre est de fabriquer des matrices en fonction des résultats qu'on veut obtenir ; il n'y a pas de meilleure preuve de sa maîtrise de la méthode statistique que la capacité à inventer des données qui permettent d'aboutir à des résultats souhaités (bien entendu, la fin est ici didactique ; elle n'est pas scientifiquement admissible).

Toutes les matrices du cahier d'exercices, de même que celles qui ont été utilisées dans le manuel à des fins d'illustration, sont mises à la disposition des enseignants sur le site web de l'éditeur à l'adresse suivante : <http://www.prisedeparole.ca/titres-livre/?id=532>.

EXERCICES

2 - LES VARIABLES

Indiquez votre opinion sur chacun des énoncés suivants en cochant « vrai » ou « faux ».

- 1) L'analyse statistique est concevable en dehors de la notion de variable.
 Vrai
 Faux

- 2) Les notions de variables nominales, variables d'espèce, variables à caractère spécifique et variables qualitatives sont équivalentes.
 Vrai
 Faux

- 3) Les variables nominales sont hiérarchiques.
 Vrai
 Faux

- 4) La variable dont les modalités sont « pour » et « contre » est nominale.
 Vrai
 Faux

- 5) La variable dont les modalités sont « réussite » et « échec » est dichotomique.
 Vrai
 Faux

- 6) Si l'on a, au point de départ, quatre niveaux pour une variable ordinale, il est possible de créer une variable à huit niveaux.
 Vrai
 Faux

- 7) Si l'on a, au point de départ, huit niveaux pour une variable ordinale, il est possible de les ramener à quatre niveaux.
- Vrai
 Faux
- 8) Il est parfois justifié de faire en sorte que l'étendue des classes d'une échelle ordinale ne soit pas la même pour chacune des classes.
- Vrai
 Faux
- 9) Les variables cardinales se distinguent des variables ordinales et des variables nominales en grande partie par les opérations arithmétiques qu'elles rendent possibles.
- Vrai
 Faux
- 10) La valeur de zéro (0) sur une échelle de Likert qui va de 0, soit « pas du tout d'accord », à 5, soit « tout à fait d'accord », est arbitraire.
- Vrai
 Faux
- 11) On peut dire d'une échelle de Likert qui va de 0, soit « pas du tout d'accord », à 5, soit « tout à fait d'accord », qu'elle est cardinale et de proportion.
- Vrai
 Faux
- 12) On peut dire d'une échelle de température en degrés Celsius qu'elle est cardinale et de proportion puisqu'elle a une valeur de zéro (0).
- Vrai
 Faux
- 13) Le budget familial en dollars constitue une échelle cardinale dont on peut dire qu'elle est de rapport parce que sa valeur de zéro (0) est absolue.
- Vrai
 Faux

- 14) Une échelle de Likert dont le nombre de valeurs est pair est fondamentalement différente d'une autre dont le nombre de valeurs est impair.
- Vrai
 - Faux
- 15) Quand une échelle de Likert est destinée à des enfants, elle doit comporter un grand nombre de valeurs afin de leur permettre de faire des nuances.
- Vrai
 - Faux
- 16) Les échelles de Likert peuvent être traitées comme ordinales.
- Vrai
 - Faux
- 17) Les échelles de Likert peuvent être traitées comme cardinales.
- Vrai
 - Faux
- 18) Au plan statistique, il importe peu, sur une échelle de Likert, que la valeur la plus faible soit celle du désaccord ou celle de l'accord.
- Vrai
 - Faux
- 19) Au plan sémantique, il importe peu, sur une échelle de Likert, que la valeur la plus faible soit celle du désaccord ou celle de l'accord.
- Vrai
 - Faux
- 20) Une variable originellement cardinale peut être transformée en une variable nominale.
- Vrai
 - Faux
- 21) Une variable originellement nominale peut être transformée en une variable ordinale.
- Vrai
 - Faux

- 22) Une variable originellement ordinale peut être transformée en une variable nominale.
- Vrai
 - Faux
- 23) La variable indépendante est celle dont il est convenu qu'elle influe sur une autre.
- Vrai
 - Faux
- 24) Quand une variable agit comme dépendante, c'est qu'on considère qu'elle est déterminée par une autre.
- Vrai
 - Faux
- 25) « Variable dépendante » est synonyme de « variable exogène ».
- Vrai
 - Faux

EXERCICES

3 - STATISTIQUES DESCRIPTIVES UNIVARIÉES

Effectuez les exercices suivants.

1) Dans la distribution suivante

Mode de transport pour se rendre au travail	f
Automobile, camion ou fourgonnette en tant que conducteur	328
Automobile, camion ou fourgonnette en tant que passager	63
Transport en commun	145
À bicyclette	22
À pied	54
Autres moyens	19
Total	631

le mode est 328.

- Vrai
- Faux

2) Calculez le rapport de variation de cette distribution sur les modes de transport.

3) Transformez en distribution de fréquence relative cette distribution de fréquence absolue.

4) On a la distribution suivante :

Âge de l'enfant	f
De 0 à 3 mois	29
De 4 à 11 mois	44
De 1 à 3 ans	67
De 4 à 7 ans	121
De 8 à 10 ans	71
Total	332

Trouvez-en la médiane.

5) Dans cette distribution de l'âge, le mode est « de 4 à 7 ans ».

- Vrai
- Faux

6) On a la distribution suivante :

Résultat à un examen sur 20	f
3	1
4	4
5	3
6	4
7	9
8	11
9	8
10	14
11	22
12	32
13	13
14	9
15	4
16	2
17	1
Total	137

Trouvez-en la médiane.

7) Dans cette distribution sur les résultats d'examen, le mode est de 12.

- Vrai
- Faux

8) Dans cette distribution sur les résultats d'examen, l'étendue ou la plage est de 136.

- Vrai
- Faux

9) On a le nombre de personnes décédées suite à un accident de la route au cours de 10 années.

Année	1	2	3	4	5	6	7	8	9	10
Décès	49	47	64	53	55	72	46	54	62	52

Calculez la moyenne du nombre de décès par année.

10) Calculez l'écart-type de cette distribution sur les accidents de la route.

11) Dans cette distribution, la médiane est de 53,5.

- Vrai
- Faux

12) Dans cette distribution, l'étendue ou la plage est de 26.

- Vrai
- Faux

13) On a la distribution suivante du nombre d'enfants par famille pour 43 familles :

Nombre d'enfants	f
0	11
1	14
2	12
3	9
4	6
5	3
6	1
7	1
Total	57

Trouvez-en la moyenne.

14) Trouvez l'écart-type de cette distribution sur le nombre d'enfants par famille.

15) Dans cette distribution sur le nombre d'enfants par famille, l'étendue ou la plage est de 13.

- Vrai
 Faux

16) Calculez le coefficient de variation de cette distribution sur le nombre d'enfants par famille.

17) Un chercheur doit établir le niveau de sexisme d'une population. Il construit une échelle de 1 à 10 où 1 indique un faible niveau de sexisme et 10, un niveau de sexisme élevé. Il saisit ses résultats dans la matrice du logiciel *SPSS*. Pour analyser ces données, repérez le fichier *SPSS* de données intitulé « Statistiques descriptives univariées ». À l'aide du logiciel *SPSS*, obtenez le mode, la médiane, la moyenne, l'écart-type et la plage. Interprétez vos résultats.

EXERCICES

4 - L'INFÉRENCE

Indiquez votre opinion sur chacun des énoncés suivants en cochant « vrai » ou « faux ».

- 1) Il importe de recourir à la statistique inférentielle même quand les calculs sont établis à partir de la population.
 Vrai
 Faux

- 2) Souvent, pour des raisons économiques, il est préférable de mener une enquête à partir d'un échantillon plutôt que de le faire sur la population elle-même.
 Vrai
 Faux

- 3) Dans bien des circonstances, il est tout simplement impossible d'effectuer une collecte de données sur l'entièreté d'une population.
 Vrai
 Faux

- 4) La notion de « population », en statistique, renvoie nécessairement à la notion de « personnes ».
 Vrai
 Faux

- 5) La notion de « population », en statistique, renvoie nécessairement à la notion d'« individu ».
 Vrai
 Faux

- 6) Dans la constitution d'un échantillon, la manière dont les individus sont sélectionnés importe peu ; un échantillon est d'autant meilleur qu'il comprend une grande proportion de la population.
- Vrai
- Faux
- 7) La loi des grands nombres a peu à voir avec le mode de sélection des individus qui constituent l'échantillon.
- Vrai
- Faux
- 8) La loi des grands nombres veut que plus un échantillon est grand, plus il soit en mesure de représenter une population.
- Vrai
- Faux
- 9) La loi des grands nombres ne connaît pas de palier quant au principe de l'adéquation entre l'échantillon et la population en fonction de l'augmentation de la taille de l'échantillon.
- Vrai
- Faux
- 10) La sélection des individus au hasard est l'une des meilleures manières de faire en sorte que les résultats qui reposent sur un échantillon ne soient pas attribuables au hasard.
- Vrai
- Faux
- 11) L'échantillon accidentel est probabiliste.
- Vrai
- Faux
- 12) L'échantillon aléatoire simple est probabiliste.
- Vrai
- Faux

- 13) L'échantillon de volontaires est probabiliste.
- Vrai
 - Faux
- 14) L'échantillon systématique est probabiliste.
- Vrai
 - Faux
- 15) L'échantillon par quota est probabiliste.
- Vrai
 - Faux
- 16) L'échantillon par grappe est probabiliste.
- Vrai
 - Faux
- 17) Il n'est permis de recourir à la statistique inférentielle que si l'échantillon est probabiliste.
- Vrai
 - Faux
- 18) La distribution normale centrée réduite a une moyenne de 1 et un écart-type de 1.
- Vrai
 - Faux
- 19) La courbe normale centrée réduite peut servir à standardiser des résultats.
- Vrai
 - Faux
- 20) La courbe normale centrée réduite peut servir à calculer des probabilités.
- Vrai
 - Faux
- 21) La courbe normale centrée réduite peut servir à faire de l'inférence.
- Vrai
 - Faux

22) La courbe normale centrée réduite est symétrique de chaque côté de la moyenne.

- Vrai
 Faux

23) Le théorème central limite suggère qu'une distribution échantillonnale sera anormalement distribuée si elle est aléatoirement tirée d'une population anormalement distribuée dès lors que l'échantillon sera composé d'au moins trente individus.

- Vrai
 Faux

24) Dans la formule pour trouver un intervalle de confiance sur une moyenne

$$\mu = \bar{X} \pm z \frac{\sigma}{\sqrt{n}}$$

on peut remplacer σ par s si $n \geq 30$.

- Vrai
 Faux

25) Dans la formule pour trouver un intervalle de confiance sur une proportion

$$P = p \pm z \sqrt{\frac{PQ}{n}}$$

on ne peut remplacer P et Q par p et q même si $n \geq 30$.

- Vrai
 Faux

26) Conclure qu'une chose est vraie quand elle est fausse, c'est commettre une erreur de première espèce.

- Vrai
 Faux

27) Conclure qu'une chose est vraie quand elle est fausse, c'est commettre une erreur alpha.

- Vrai
 Faux

- 28) Dans les tests d'hypothèse, la décision se prend sur l'hypothèse rivale plutôt que sur l'hypothèse nulle.
- Vrai
 Faux
- 29) L'intervalle de confiance et le test d'hypothèses ont pour fondement un calcul de probabilité.
- Vrai
 Faux
- 30) Si l'hypothèse nulle est de forme $\mu = A$ et que l'hypothèse rivale est $\mu > A$, alors le test d'hypothèse est bilatéral.
- Vrai
 Faux
- 31) Si l'hypothèse nulle est de forme $\mu = A$ et que l'hypothèse rivale est $\mu < A$, alors le test d'hypothèse est bilatéral.
- Vrai
 Faux
- 32) Si l'hypothèse nulle est de forme $\mu = A$ et que l'hypothèse rivale est $\mu \neq A$, alors le test d'hypothèse est bilatéral.
- Vrai
 Faux

Effectuez les exercices suivants

- 33) Soit z une variable aléatoire normale centrée réduite, déterminez la probabilité de $p(0 \leq z \leq 0,8)$.
- 34) Soit z une variable aléatoire normale centrée réduite, déterminez la probabilité de $p(0 \leq z \leq 2,6)$.

- 35) Soit z une variable aléatoire normale centrée réduite, déterminez la probabilité de $p(1,61 \leq z \leq 1,96)$.
- 36) Soit z une variable aléatoire normale centrée réduite, déterminez la probabilité de $p(z \leq -1,92)$.
- 37) Soit z une variable aléatoire normale centrée réduite, déterminez la probabilité de $p(-1,42 \leq z \leq 0,99)$.
- 38) Soit z une variable aléatoire normale centrée réduite, déterminez la probabilité de $p(z \geq 2,56)$.
- 39) Soit z une variable aléatoire normale centrée réduite, déterminez la probabilité de $p(z \leq -1,49)$.
- 40) Soit z une variable aléatoire normale centrée réduite, déterminez la probabilité de $p(z \geq -1,0)$.

- 41) Soit z une variable aléatoire normale centrée réduite, déterminez la probabilité de p ($-2 \leq z \leq -1,0$).
- 42) Soit z une variable aléatoire normale centrée réduite, déterminez la probabilité de p ($-0,5 \leq z \leq 0$).
- 43) Quel est le score z pour le revenu d'une personne qui gagne 38 000 \$ par année dans un environnement où la moyenne est de 44 000 \$ et l'écart-type, de 8 000 \$?
- 44) À un test de raisonnement, une personne obtient 29 quand la moyenne est de 27 et l'écart-type de 6. À un test de mémoire, cette personne obtient 69 quand la moyenne est de 64 et l'écart-type de 26. Ses performances sont-elles meilleures en raisonnement ou en mémoire?
- 45) Une démographe dénombre les individus dans 22 localités d'une région périphérique. Elle doit transformer la population de chacune de ces localités en valeur de z . À l'aide du logiciel *SPSS*, effectuez cette opération pour elle à partir de la matrice « Transformation en z – démographie ». Cela fait, trouvez la moyenne et l'écart-type de cette nouvelle distribution.

- 46) On a un échantillon d'enfants âgés de 3 à 6 ans qui fréquentent des écoles de langue française en Ontario, une province à majorité anglophone. Un de leurs parents a estimé le nombre d'heures pendant lesquelles leur enfant est exposé à l'anglais hebdomadairement. Avec le logiciel *SPSS*, trouvez un intervalle de confiance à 95 % de la moyenne de cette distribution. Les données se trouvent dans le fichier « IC moyenne – exposition au français ».
- 47) Un chercheur dispose d'un échantillon de travailleurs sociaux. Ces travailleurs sociaux ont répondu à un questionnaire. L'une des questions invite les travailleurs sociaux à signaler, en répondant par « oui » ou par « non », s'ils ont manqué des journées de travail au cours des trois années qui ont précédé l'enquête à cause de leurs états psychiques : stress ; angoisse ; sentiments d'impuissance, de frustration, de découragement, de déception, d'agitation, de manque d'empathie... Avec le logiciel *SPSS*, trouvez un intervalle de confiance pour la proportion de ceux qui ont manqué des journées de travail. Les données se trouvent dans le fichier « IC proportion – journées de travail manquées ».
- 48) On a soumis un échantillon d'enfants à un test de vocabulaire. Pour ce test, la norme est de 100. Déterminez si le score moyen de ces enfants de l'échantillon diffère de la norme. Les données se trouvent dans le fichier « Test hypothèse – un échantillon – test de vocabulaire ».
- 49) Dans une enquête par questionnaire, un échantillon de 1 620 jeunes ont indiqué leur moyenne scolaire sur une échelle de 0 à 100. Déterminez si cette distribution correspond à celle d'une courbe normale. Les données se trouvent dans le fichier *SPSS* « KS – moyenne scolaire ».

EXERCICES

5 - LE CHI-CARRÉ

Effectuez les exercices suivants.

- 1) Pour interpréter les tableaux de contingence, lorsque les groupes qui sont comparés ne sont pas de la même taille et que cette taille le justifie, il est préférable de transformer les fréquences absolues en fréquences relatives.
 Vrai
 Faux

- 2) Pour interpréter les tableaux de contingence, on calcule les pourcentages dans le sens de la variable dépendante.
 Vrai
 Faux

- 3) On ne peut construire des tableaux de contingence que si les groupes qui sont comparés sont de taille semblable.
 Vrai
 Faux

- 4) Le test du chi-carré est établi à partir des chiffres absolus.
 Vrai
 Faux

- 5) Dans le test du chi-carré, les valeurs observées de chacune des cases sont comparées à celles que le hasard aurait générées.
 Vrai
 Faux

- 6) Dans son rapport au chi-carré, le degré de liberté renvoie aux dimensions du tableau de contingence.
- Vrai
 Faux
- 7) Plus le degré de liberté est grand, alors plus doit être faible la valeur de chi-carré pour que le test soit significatif.
- Vrai
 Faux
- 8) Plus la valeur de chi-carré est grande, alors plus la probabilité de commettre une erreur de première espèce est grande elle aussi.
- Vrai
 Faux
- 9) Le test du chi-carré, quand il se rapporte à un tableau de contingence, peut avoir pour hypothèse nulle l'indépendance des variables.
- Vrai
 Faux
- 10) Le test du chi-carré, quand il se rapporte à un tableau de contingence, peut avoir pour hypothèse nulle la similitude de la distribution de fréquences des modalités de la variable indépendante dans leur rapport aux modalités de la variable dépendante.
- Vrai
 Faux
- 11) Dans un tableau de contingence, quand les valeurs théoriques sont inférieures à cinq, alors il faut interpréter le résultat d'un test positif avec réserve.
- Vrai
 Faux
- 12) Une valeur théorique de zéro dans une case d'un tableau de contingence doit automatiquement éveiller un doute quant à la légitimité du calcul d'un chi-carré.
- Vrai
 Faux

- 13) Une valeur observée de zéro dans une case d'un tableau de contingence doit automatiquement éveiller un doute quant à la légitimité du calcul d'un chi-carré.
- Vrai
 Faux
- 14) Il est préférable de ne pas recourir au correctif de Yates quand on veut inférer les variations dans un tableau de contingence de 2 x 2 et quand il est justifiable d'employer le chi-carré.
- Vrai
 Faux
- 15) Pour les tableaux de contingence de 2 x 2, on substituera le test exact de Fisher à celui du chi-carré corrigé quand le nombre d'individus qui sont pris en considération est inférieur à 20.
- Vrai
 Faux
- 16) Une chercheuse s'intéresse à l'abandon des études chez des jeunes qui se sont inscrits dans des programmes de niveau universitaire. Elle soupçonne que la persévérance est en partie attribuable à ce que les parents aient eux-mêmes parachevé de telles études, deux parents étant plus favorables à cette persistance qu'un seul. Elle entend vérifier cette hypothèse en recueillant des informations auprès d'un échantillon aléatoire de personnes qui sont en âge d'avoir terminé des études de premier cycle mais qui n'ont pas atteint l'âge de 30 ans, chacune d'elles ayant été inscrite dans un programme d'études universitaires. Avec le logiciel *SPSS*, vérifiez si elle confirme son hypothèse. Les données se trouvent dans le fichier « Chi-carré – deuxième génération ».
- 17) Un chercheur souhaite vérifier l'hypothèse selon laquelle les individus qui travaillent à plein temps seront proportionnellement plus nombreux à faire des dons en argent à des personnes que ceux qui travaillent à temps partiel. Une enquête lui a permis d'obtenir le statut de ces individus relativement à l'emploi de même que la réponse à la question qui consiste à déterminer si, oui ou non, ils font des dons en argent à des personnes. Avec le logiciel *SPSS*, vérifiez s'il confirme son hypothèse. Les données se trouvent dans le fichier « Chi-carré – don en argent et travail ».

- 18) Une chercheuse a obtenu un échantillon de 19 jeunes contrevenants qui ont été tenus coupables de voies de fait graves. Elle souhaite déterminer si les peines de détention pour ces crimes sont aussi communes pour les filles que pour les garçons. Avec le logiciel *SPSS*, vérifiez s'il est permis de conclure que pour ce type de crime la détention est aussi probable pour les filles que pour les garçons. Les données se trouvent dans le fichier « Fisher – délinquance juvénile ».
- 19) Le test de McNemar permet de croiser deux variables dichotomiques – l'une indépendante, dont les modalités peuvent être considérées comme étant des échantillons indépendants, l'autre dépendante.
- Vrai
- Faux
- 20) Un gouvernement s'apprête à adopter une loi qui entend assouplir les règles qui encadrent la présence sur le territoire de la nation de travailleurs saisonniers d'origine étrangère. Un organisme s'oppose fortement à l'orientation que prend le gouvernement. Il prépare une campagne de sensibilisation dont le but est de mobiliser l'opinion publique contre la loi qui est prévue. Dans sa campagne, il souhaite diffuser à grande échelle une courte vidéo de six minutes. Avant de le faire, il teste l'incidence de cette vidéo sur un échantillon de 100 individus. À cette fin, il demande à chacune des personnes échantillonnées si elle est *a priori* plutôt pour ou plutôt contre la loi ; il fait visionner la vidéo à chacune d'elles, après quoi il lui repose la question. Vérifiez avec le logiciel *SPSS* s'il lui est permis de croire que la vidéo sert bien ses intérêts. Les données se trouvent dans le fichier « McNemar – opinion sur la loi ».
- 21) Un traitement suppose, pour être efficace, une prise de médicament à intervalle régulier et sans interruption. La plupart des sujets, quand ils sont laissés à eux-mêmes, ne parviennent pas à suivre correctement les directives. La compagnie pharmaceutique crée une application téléphonique pour venir en aide aux malades. Elle vérifie auprès de 44 d'entre eux qui disposent d'un téléphone intelligent si cette application est bénéfique. Découvrez à l'aide du logiciel *SPSS* si cette application se révèle bénéfique. Les données se trouvent dans le fichier « McNemar – application téléphonique ».

EXERCICES

6 - L'ANALYSE D'UNE SEULE VARIABLE

Effectuez les exercices suivants.

- 1) Un psychopédagogue est amené à se demander si les troubles d'apprentissage parmi les élèves d'une population immigrante concernent davantage les filles que les garçons. Pour vérifier son hypothèse, il tire un échantillon aléatoire dans lequel la variable sexe est prise en considération. En vous servant du logiciel *SPSS*, déterminez s'il confirme son hypothèse. Les données se trouvent dans le fichier « Test binomial – trouble d'apprentissage ».
- 2) Une sociologue du travail a établi que dans une population de travailleurs sociaux on dénombrait, à un moment donné, 14 % qui présentaient des symptômes de surmenage. Ses analyses l'ont conduite à avancer que la situation des infirmiers est très semblable et que, par conséquent, on devrait en trouver 14 % également qui présentent ces symptômes. Pour vérifier son hypothèse, elle obtient les données d'un échantillon de personnes qui exercent la profession d'infirmier. Avec le logiciel *SPSS*, vérifiez si elle peut soutenir son hypothèse. Les données se trouvent dans le fichier « Test binomial – surmenage ».
- 3) Un récréologue a des raisons de croire que la population de touristes qui se déplace vers les pays chauds pendant la saison hivernale se divise en quatre groupes dans des proportions à peu près égales : les individus qui privilégient les activités sportives, ceux qui préfèrent les activités culturelles, ceux qui penchent vers la farniente et ceux qui apprécient quelque combinaison de ces trois possibilités. Il tire un échantillon de touristes parmi les clients d'une agence de voyage et les interroge en ce sens. Découvrez, à l'aide du logiciel *SPSS*, s'il est en mesure de vérifier son hypothèse. Les données se trouvent dans le fichier « Ajustement chi-carré – touristes ».

- 4) Une politologue doute qu'un échantillon ait réellement été tiré d'une population donnée. Elle sait que dans la population, on dénombre 6 % de personnes qui n'ont pas terminé le cours secondaire, 39 % qui ont atteint le diplôme d'études secondaires sans poursuivre leurs études, 32 % qui ont obtenu un diplôme d'études collégiales et 23 % qui détiennent un diplôme d'études universitaires. Pour vérifier son hypothèse, elle compare cette distribution théorique à celle de l'échantillon suspect. À l'aide du logiciel *SPSS*, découvrez si son doute est justifié. Les données se trouvent dans le fichier « Ajustement chi-carré – instruction ».

EXERCICES

7 - LE LOG-LINÉAIRE

Effectuez l'exercice suivant.

Vous êtes embauché-e par un organisme culturel pour déterminer si la préférence pour les comédies cinématographiques est liée à l'attrait pour la musique classique et pour découvrir si cette relation est affectée par l'âge. Pour effectuer cette analyse, vous disposez d'un échantillon aléatoire. Le rapport au cinéma et à la musique classique se révèle dans des variables dichotomiques dont les modalités sont « non » et « oui ». Le rapport à l'âge est lui aussi binaire et ses modalités sont « adolescent » et « adulte ». Les données se trouvent dans le fichier « Log-linéaire – cinéma musique âge ». Avec l'aide du logiciel *SPSS*,

- i. utilisez l'analyse log-linéaire hiérarchique par élimination descendante pour trouver un modèle adapté aux données ;
- ii. effectuez une analyse log-linéaire généralisée pour interpréter ce modèle.

EXERCICES

8 - LES MESURES D'ASSOCIATION POUR LES TABLEAUX DE CONTINGENCE

Effectuez les exercices suivants.

- 1) Dans les écoles de natation de plusieurs municipalités, il y a un groupe d'élite. Pour faire partie de ce groupe, il faut se soumettre à des épreuves de qualification. Un kinésologue se demande si le fait de s'être préalablement entraîné seul ou en équipe a une incidence sur le résultat des épreuves. Pour le vérifier, il se dote d'un échantillon d'élèves ayant participé aux dernières épreuves et se renseigne sur la manière dont ils se sont entraînés, de même que sur leur performance lors des qualifications. À l'aide du logiciel *SPSS*, découvrez si les modalités de l'entraînement influent sur le résultat des qualifications. Les données se trouvent dans le fichier « Phi – qualifications ».
- 2) Des personnes ont une maladie incurable qui les fait souffrir. La médecine ne peut pas vraiment intervenir pour calmer la douleur sans fortement réduire la qualité de vie des malades. Une clinique met à la disposition de ces malades un service psychologique dont l'objectif est de permettre de s'adonner aux activités de la vie ordinaire tout en contrôlant la douleur. Un gestionnaire doit vérifier si ce service est bénéfique. Il construit un échantillon d'individus chez qui la maladie en est approximativement au même niveau. Il s'informe sur le fait qu'ils occupent ou non un emploi et les répartit selon qu'ils fréquentent ou non la clinique psychologique. Découvrez, avec le logiciel *SPSS*, si le recours à la clinique favorise le maintien d'un emploi. Les données se trouvent dans le fichier « Phi – clinique psychologique ».
- 3) Des leaders d'une ethnie qui habite dans une région éloignée estiment que leurs membres sont discriminés en matière d'emploi. Ils soupçonnent notamment que les postes à temps plein sont attribués en grand nombre à des migrants et que les citoyens des autres ethnies qui résident dans la même localité sont moins communément au chômage que leurs propres membres. Pour découvrir s'il y a effectivement une corrélation entre la situation d'emploi et le statut de la personne, ils obtiennent les registres officiels de l'État et répartissent les individus dans un tableau de contingence. À l'aide du logiciel *SPSS*, vérifiez s'il y a

corrélation et, le cas échéant, si elle confirme les soupçons des leaders. Les données se trouvent dans le fichier « V de Cramér – discrimination ».

- 4) Une spécialiste en sociologie de l'éducation s'interroge sur les aspirations éducationnelles des élèves en voie de terminer leurs études secondaires dans une région excentrée. Elle soupçonne que, dans cette région, les filles sont plus interpellées que les garçons par les études universitaires et que les garçons tendent plus que les filles à donner à leurs études une orientation nettement professionnelle. Pour le vérifier, elle obtient un échantillon aléatoire des finissants du secondaire et les interroge sur leurs perspectives d'études, qu'elle classe en quatre catégories. À l'aide du logiciel *SPSS*, découvrez s'il y a quelque corrélation entre le sexe et le type d'études envisagées, et déterminez si l'hypothèse de la sociologue se confirme. Les données se trouvent dans le fichier « V de Cramér – aspiration éducationnelle ».
- 5) Un psycho-anthropologue soutient que la perception de l'agressivité dans le visage d'une personne dépend de la culture. Pour le vérifier, il stratifie son échantillon en trois modalités : culture 1, culture 2 et culture 3. La culture 1 est théoriquement celle qu'il considère comme étant la plus proche d'une certaine expression de l'agressivité ; la culture 3 en est, en principe, la plus éloignée. Sur un écran, cette expression apparaît parmi onze autres faciès. Chacun des participants doit balayer l'écran des yeux et repérer le visage agressif. Trois résultats sont possibles : identifie le visage au premier balayage ; identifie le visage au second balayage ; n'identifie pas le visage. L'intention du psycho-anthropologue est d'établir si le résultat est déterminé par la culture et, le cas échéant, dans quelle proportion. Découvrez si le logiciel *SPSS* confirme son hypothèse et, le cas échéant, jusqu'à quel point il le fait. Les données se trouvent dans le fichier « Lambda – culture et agressivité ».
- 6) Un vote de grève doit avoir lieu au sein d'une grande entreprise. Plusieurs craignent que le vote soit trop déterminé en fonction de la situation des travailleurs, selon qu'ils sont permanents, en probation ou que leur contrat soit à durée limitée, les premiers tendant à être majoritairement en faveur de cette grève, les deuxièmes présentant un peu plus d'hésitation, les autres exprimant une certaine résistance. Le syndicat embauche une spécialiste pour découvrir jusqu'à quel point la position des travailleurs peut prédire leur vote. En vous servant du logiciel *SPSS*, découvrez la réponse que peut fournir la spécialiste après s'être pourvue d'un échantillon des travailleurs. Les données se trouvent dans le fichier « Lambda – vote de grève ».

- 7) Un sociolinguiste a obtenu des enregistrements de 64 locuteurs. Il souhaite les classer en deux catégories pour mener des analyses comparées selon que la langue orale serait dite régionale ou standard. Il établit une grille de classification et demande à deux juges d'appliquer séparément cette grille et de catégoriser les enregistrements. À l'aide du logiciel *SPSS*, déterminez le degré d'entente entre les deux juges. Les données se trouvent dans le fichier « Kappa – sociolinguiste ».
- 8) Une diététicienne a mené des entretiens auprès de 140 individus dans le but de découvrir leur mode alimentaire. Elle souhaite classer ces individus dans trois ensembles. Dans le premier apparaîtront les personnes qui ne s'imposent pas de restriction alimentaire ; dans le deuxième, celles qui s'interdisent des aliments ; dans le troisième, celles qui alternent entre l'exclusion et la non-exclusion d'aliments. Pour que sa catégorisation soit adéquate, la diététicienne établit des critères et, pour s'assurer que ces critères sont efficaces, elle demande à deux juges de les mettre en application, chacun de son côté. Vérifiez si leurs deux classements concordent en vous servant du logiciel *SPSS*. Les données se trouvent dans le fichier « Kappa – modes alimentaires ».
- 9) Un spécialiste en santé communautaire en vient à déduire que meilleure est l'estimation de la santé physique, meilleure est aussi l'estimation de la santé émotionnelle. Il obtient un échantillon de 1 493 jeunes du secondaire. Il leur donne à lire deux énoncés : « Comparativement à d'autres personnes de mon âge, ma santé physique est... » et « Comparativement à d'autres personnes de mon âge, ma santé émotionnelle est... ». À côté de chacun de ces énoncés, il y a une échelle qui va de 1, pour « mauvaise », à 6, pour « excellente ». Les jeunes doivent encercler une valeur sur chacune de ces deux échelles. À l'aide du logiciel *SPSS*, vérifiez s'il y a une corrélation positive entre ces deux formes d'auto-estimation de la santé. Les données se trouvent dans le fichier « Gamma – auto-estimation de la santé ».
- 10) Une psychologue du travail soulève l'hypothèse que plus les travailleurs sociaux estiment que l'organisme qui les embauche connaît des réductions de personnel, plus ils tendront à éprouver du stress. Pour le vérifier, elle recueille des informations auprès d'un échantillon de ces travailleurs et leur propose deux énoncés. L'un d'eux est le suivant : « Au cours des dernières années, l'organisme pour lequel je travaille a connu des réductions d'effectifs » ; y est associée une échelle qui va de 1, « pas du tout d'accord », à 6, « tout à fait d'accord ». L'autre est formulé en ces termes : « À cause des conditions dans lesquelles je travaille, j'éprouve du stress » ; l'échelle va aussi de 1 à 6, mais sémantiquement elle oscille entre

« jamais » et « très souvent ». À l'aide du logiciel *SPSS*, découvrez si corrélation positive il y a. Les données se trouvent dans le fichier « Gamma – stress ».

- 11) Un spécialiste en santé publique veut calculer, dans un échantillon de 14 adultes âgés entre 30 et 50 ans, la corrélation entre le revenu annuel en dollars et l'auto-estimation de la qualité de la santé. Cette estimation est donnée sur une échelle qui va de 1, pour « très mauvaise », à 10, pour « excellente ». À l'aide du logiciel *SPSS*, découvrez la réponse qu'il obtient. Les données se trouvent dans le fichier « Tau de Kendall – revenu et santé ».
- 12) Deux juges ont 17 performances athlétiques à classer par ordre de qualité allant de la meilleure (dont le rang est 1) à la moins appréciable (dont le rang est 17). À l'aide du logiciel *SPSS*, déterminez si ces classements sont corrélés. Les données se trouvent dans le fichier « Tau de Kendall – performances athlétiques ».
- 13) Une chercheuse sait que le chi-carré est significatif pour un tableau de contingence dans lequel la variable indépendante a les modalités « minoritaire » et « majoritaire » et la variable dépendante a trois modalités, soit « aime ma communauté », « aime plus ou moins ma communauté » et « n'aime pas ma communauté ». La chercheuse croit qu'au-delà de la variation du sentiment à l'égard de la communauté en fonction du statut du citoyen, il y a quelque linéarité entre les deux variables. À l'aide du logiciel *SPSS*, vérifiez s'il faut donner crédit à cette hypothèse. Les données se trouvent dans le fichier « Association linéaire – affection pour la communauté ».
- 14) Un orthophoniste dispose de données sur des élèves. Ces données lui indiquent si un élève a été identifié par un orthophoniste comme ayant un trouble du langage et si cet élève est de sexe masculin ou féminin. Il croit que l'ensemble des jeunes chez qui un trouble du langage a été signalé comportera plus de garçons que de filles, et il s'attend même à ce que cette relation entre la présence d'un trouble et le sexe comporte quelque linéarité. À l'aide du logiciel *SPSS*, déterminez s'il a raison. Les données se trouvent dans le fichier « Association linéaire – trouble du langage ».
- 15) Une sociologue se demande s'il y a un lien entre le groupe ethnique auquel s'identifie un enfant qui fréquente l'école secondaire et son niveau d'aspiration professionnelle. Pour le vérifier, elle demande aux individus qui composent son échantillon : « Parmi ces groupes, auquel t'identifies-tu le plus? » Et elle propose trois possibilités de réponse : Premières Nations ; francophones ; anglophones (ceux-ci constituant un groupe majoritaire). Pour

connaître l'aspiration professionnelle, la sociologue demande : « Cinq ans après la fin de tes études, quel genre d'emploi auras-tu? ». Les métiers qu'envisagent d'exercer les élèves sont ensuite ramenés à cinq niveaux qui vont de 1, « inférieur », à 5, « supérieur ». Les niveaux sont déterminés en prenant en considération le revenu moyen des professions, le niveau d'instruction médian et le prestige qui leur est normalement accordé. On a ainsi une variable nominale et une variable cardinale. Déterminez si la sociologue peut répondre positivement à sa question et, si oui, quelle est la force du lien entre ces deux variables. Les données se trouvent dans le fichier « Eta – aspiration professionnelle ».

- 16) On a tiré d'une école un échantillon de 22 élèves des classes de première année du secondaire : 11 filles et 11 garçons. On leur a demandé d'estimer leur revenu annuel cinq ans après la fin de leurs études secondaires. À l'aide du logiciel *SPSS*, déterminez s'il y a une corrélation entre le sexe des élèves et le revenu auquel ils aspirent. Les données se trouvent dans le fichier « Eta – revenu prisé ».
- 17) Un chercheur veut savoir s'il y a un lien entre le sexe (femme ou homme) et l'attitude (positive ou négative) envers une loi du gouvernement tout en contrôlant le niveau d'instruction (secondaire ; universitaire de premier cycle ; universitaire des études supérieures). Pour découvrir quel est le résultat de cette analyse, effectuez les tests de Cochran et de Mantel-Haenszel avec l'aide du logiciel *SPSS*. Les données se trouvent dans le fichier « Cochran et Mantel-Haenszel – sexe loi instruction ».

EXERCICES

9 - TEST T

Effectuez les exercices suivants.

- 1) On veut vérifier l'hypothèse selon laquelle le mode de mémorisation peut influencer la performance de la mémoire. 40 sujets ont été assignés de façon aléatoire à 2 conditions de mémorisation : A et B. La mémoire est mesurée par un test dont les scores varient de 1 à 80. À l'aide du logiciel *SPSS*, et en recourant au test t, vérifiez cette hypothèse. Les données se trouvent dans le fichier « Test t - mémoire méthode ».

- 2) On veut tester l'hypothèse selon laquelle la méthode de mémorisation A favorise la mémoire. Un échantillon aléatoire de 20 sujets effectuent une tâche de mémoire dans deux conditions : la condition expérimentale, où la méthode A est utilisée, et la condition de contrôle, où aucune intervention n'a lieu. La séquence pour les deux conditions est contrebalancée : la moitié des sujets est d'abord soumise à la condition expérimentale et, ensuite, à la condition contrôle ; l'ordre est inversé pour l'autre moitié des sujets. La mémoire est mesurée par un test dont les scores varient de 1 à 90. Testez cette hypothèse à l'aide du logiciel *SPSS*. Les données se trouvent dans le fichier « Test t appariés - mémoire méthode ».

EXERCICES

10 - L'ANALYSE DE VARIANCE (ANOVA)

Effectuez les exercices suivants.

- 1) On veut savoir si deux méthodes de mémorisation, A et B, peuvent influencer la performance de mémoire. On divise de façon aléatoire 60 sujets dans trois groupes : le groupe A utilise la méthode A, le groupe B utilise la méthode B et le troisième est le groupe contrôle, dans lequel aucune méthode n'est imposée. Leur performance de mémoire est définie par les scores au test de mémoire, qui varient théoriquement de 1 à 60. À l'aide du logiciel *SPSS*, déterminez si la méthode a quelque incidence. Les données se trouvent dans le fichier « ANOVA 1 facteur – mémoire méthode ».

- 2) On veut savoir si le niveau d'activation physiologique pendant l'écoute de la musique classique est différent chez les femmes et les hommes selon leur préférence musicale. On choisit un échantillon de sujets. On note leur préférence musicale (rock, hip-hop ou classique) et on mesure leur activation physiologique pendant qu'ils écoutent de la musique classique (sur une échelle de 10 à 100, plus le score étant élevé, plus grande étant l'activation). À l'aide du logiciel *SPSS*, découvrez quels résultats permet de trouver cette analyse. Vous effectuez donc une analyse de variance 2 x 3 (sexe par préférence) pour l'activation psychologique ; vous présentez et interprétez les effets principaux et vous portez attention à l'effet d'interaction si tel est le cas. Les données se trouvent dans le fichier « ANOVA 2 facteurs – activation psychologique ».

- 3) Une sociolinguiste a demandé à des élèves de première année à l'université de rédiger un texte, en anglais ou en français, selon leur langue de préférence. Les personnes dont la langue de préférence est le français sont normalement en situation minoritaire, les autres appartiennent à la majorité ou sont interpellées par elle. Dans les textes, la sociolinguiste identifie les erreurs selon une grille dans laquelle doivent être repérées les fautes sémantiques, morphologiques, grammaticales, syntaxiques, orthographiques et interlinguistiques. Son hypothèse veut que le nombre d'erreurs soit plus élevé dans les textes écrits dans la langue minoritaire que dans ceux qui sont rédigés dans la langue de la majorité. Un test t lui donne raison sur ce point, la moyenne des erreurs dans les textes en français étant de 51,94 ($s = 23,04$) et, dans les textes en anglais, de 24,83 ($s = 19,17$)

($t_{(198,75)} = 9,19$ $p < 0,001$). Cependant, elle soupçonne qu'en retirant de l'analyse les erreurs interlinguistiques, qui devraient être plus communes dans le groupe minoritaire, cette inégalité de moyennes pourrait disparaître. Elle le vérifie en utilisant ces erreurs spécifiques comme covariante. Déterminez, à l'aide du logiciel *SPSS*, si cette seconde hypothèse se vérifie. Les données se trouvent dans le fichier « ANOVA avec covariante – erreurs linguistiques ».

- 4) Un chercheur veut savoir si différentes façons de solliciter l'opinion des gens peuvent influencer les résultats : les sollicitations par téléphone, par entretien et par questionnaire. Pour le vérifier, il recueille les réponses à 20 questions sur le programme d'immigration. Il obtient ainsi des scores qui varient de 0 à 20, un score élevé indiquant une opinion positive. Le chercheur divise 48 participants aléatoirement en 3 groupes de 16 pour utiliser une différente méthode dans chaque groupe. Il dispose aussi d'informations sur le nombre d'années de scolarité des participants. À l'aide du logiciel *SPSS*, effectuez une analyse de variance pour découvrir s'il y a une différence significative entre les trois méthodes ; effectuez ensuite une analyse avec une covariante pour découvrir si l'observation est la même après avoir tenu compte de la scolarité des participants. Les données se trouvent dans le fichier « ANOVA avec covariante – sollicitation de réponses ».
- 5) Dans une classe de 40, chaque présentation orale des étudiants est évaluée par le professeur, par son assistant et par les autres étudiants. L'appréciation des autres étudiants est ramenée à une seule valeur. L'évaluation se fait sur une échelle de 1 à 10. Effectuez, à l'aide du logiciel *SPSS*, une analyse de la variance à mesures répétées pour découvrir s'il y a des différences entre les évaluations selon qu'il s'agit de celle du professeur, de l'assistant ou des étudiants. Les données se trouvent dans le fichier « ANOVA à mesures répétées – présentation orale ».
- 6) Dans une classe de 40, chaque présentation orale des étudiants est évaluée par le professeur, par son assistant et par les autres étudiants. L'appréciation des autres étudiants est ramenée à une seule valeur. L'évaluation se fait sur une échelle de 1 à 10. Effectuez, à l'aide du logiciel *SPSS*, une analyse de la variance à mesures répétées pour découvrir s'il y a des différences entre les évaluations selon qu'il s'agit de celle du professeur, de l'assistant ou des étudiants eux-mêmes et vérifiez si cette différence, le cas échéant, dépend du sexe de la personne qui fait la présentation. Les données se trouvent dans le fichier « ANOVA mixte – présentation orale ».

EXERCICES

11 - L'ANALYSE DE VARIANCE MULTIVARIÉE (MANOVA)

Effectuez les exercices suivants.

- 1) Dans une recherche, on veut savoir si le genre de musique qu'on écoute a une influence sur la performance de la mémoire. 60 sujets sont divisés de façon aléatoire en trois groupes. Pendant le testage de la mémoire, les sujets du premier groupe écoutent de la musique classique, ceux du deuxième groupe écoutent de la musique rock et ceux du troisième n'écoutent aucune musique. La mémoire est mesurée à partir de trois tests dont les mesures sont cardinales : 1) la mémoire de rappel libre, dont les scores se situent entre 0 et 20 ; 2) la mémoire de reconnaissance, dont les scores varient de 0 à 40 ; 3) le temps de réaction, qui mesure la vitesse moyenne de la mémoire en secondes. Effectuez, à l'aide du logiciel *SPSS*, une analyse de la variance multivariée pour déterminer cette incidence de la musique. Présentez et interprétez l'effet multivarié et les effets univariés. Les données se trouvent dans le fichier « MANOVA 1 facteur – mémoire musique ».

- 2) Suite à cette analyse, on veut savoir si l'influence de la musique sur la mémoire sera différente chez les jeunes et les adultes. L'échantillon est augmenté et la recherche comporte maintenant deux variables indépendantes : le groupe d'âge, soit 39 jeunes âgés de 13 à 16 ans, et 42 adultes âgés de 25 à 30 ans ; le type de musique, dont les trois conditions sont « avec musique classique », « avec musique rock » et « sans musique ». À l'aide du logiciel *SPSS*, effectuez une analyse de variance multivariée, présentez et interprétez les effets principaux multivariés et univariés, les interactions multivariées et univariées, et les effets simples multivariés et univariés. Les données se trouvent dans le fichier « MANOVA 2 facteurs – mémoire musique âge ».

EXERCICES

12 - LES TESTS W DE WILCOXON ET U DE MANN-WHITNEY

Effectuez les exercices suivants.

- 1) Une sociologue de l'éducation en vient à concevoir que, dans une région donnée, les aspirations éducationnelles des filles sont plus élevées que celles des garçons. Elle dispose d'un échantillon dans lequel le sexe des élèves du secondaire est connu, de même que le niveau d'instruction auquel aspirent ces élèves. Les modalités de cette scolarité envisagée sont les suivantes :
 - 1 : quelques années de l'école secondaire
 - 2 : diplôme d'études secondaires
 - 3 : diplôme d'études collégiales
 - 4 : diplôme d'études universitaires de premier cycle
 - 5 : diplôme d'études universitaires de niveau supérieur

À l'aide du logiciel *SPSS*, vérifiez si la sociologue a raison. Les données se trouvent dans le fichier « Wilcoxon et Mann-Whitney – aspirations scolaires ».

- 2) Des marathoniens proviennent de deux écoles différentes, A et B. On se demande si l'une d'elles est plus performante. On tire un échantillon aléatoire de 10 coureurs dans chacune de ces écoles. Tous ont participé à trois marathons identiques au cours des six derniers mois. On calcule leur rang moyen pour ces trois marathons. À l'aide du logiciel *SPSS*, découvrez s'il est permis de déclarer quelque ascendance d'une école sur l'autre. Les données se trouvent dans le fichier « Wilcoxon et Mann-Whitney – marathon ».

EXERCICES

13 - LE TEST KRUSKAL-WALLIS

Effectuez les exercices suivants.

- 1) On a un échantillon d'individus issus de quatre secteurs d'emploi : gestion, enseignement, métier ouvrier, soins médicaux. Toutes ces personnes travaillent à plein temps depuis au moins deux ans. On leur demande d'estimer sur une échelle ordinale leur temps normal de travail hebdomadaire. Les modalités en sont les suivantes :
 - 1 : moins de 30 heures
 - 2 : 30 à 39 heures
 - 3 : 40 à 49 heures
 - 4 : 50 heures et plus

À l'aide du logiciel *SPSS*, déterminez s'il y a quelque différence inférable entre ces quatre secteurs. Les données se trouvent dans le fichier « Kruskal-Wallis – temps de travail ».

- 2) On a des ménages de trois personnes, de quatre personnes et de cinq personnes ou plus. Dans chacun d'eux, les enfants sont d'âge scolaire. Dans chacun d'eux, il y a un père et une mère. La question consiste à déterminer si la taille du ménage fait varier le temps que la mère passe devant un écran (télévision, ordinateur, tablette, téléphone intelligent y compris) hebdomadairement à des fins autres que liées au travail. Vérifiez-le à l'aide du logiciel *SPSS*. Les données se trouvent dans le fichier « Kruskal-Wallis – temps devant un écran ».

EXERCICES

14 - LE TEST DE LA MÉDIANE

Effectuez les exercices suivants.

- 1) Dans un questionnaire, on trouve une définition de ce qu'est un artiste professionnel ; la définition est la suivante : une personne qui vit de son art (par distinction, par exemple, à un artiste semi-professionnel rémunéré par son métier artistique mais exerçant aussi un autre métier). Le questionnaire est distribué dans une municipalité. La distribution a lieu à l'intérieur de trois populations : des artistes (professionnels, semi-professionnels et amateurs), des grands consommateurs d'art, de petits consommateurs d'art. Le questionnaire demande à tous d'estimer le nombre d'artistes professionnels qui vivent dans leur municipalité. En effectuant, à l'aide du logiciel *SPSS*, un test sur la médiane, déterminez si l'estimation varie selon le type de participant. Les données se trouvent dans le fichier « Test de la médiane – estimation d'artistes professionnels ».

- 2) On a des étudiants de deuxième année universitaire. On en tire trois échantillons selon 1) qu'aucun de leurs parents n'a fait d'études universitaires, 2) qu'un de leurs parents en a fait, 3) que les deux parents ont fait de telles études. Une chercheuse veut vérifier si le rapport à la situation de leurs parents est déterminant de la moyenne qu'ils ont obtenue lors de leurs études en première année. Effectuez, à l'aide du logiciel *SPSS*, un test de la médiane pour le découvrir. Les données se trouvent dans le fichier « Test de la médiane – moyenne de la première année universitaire ».

EXERCICES

15 - LE TEST DU SIGNE

Effectuez les exercices suivants.

- 1) Le gouvernement vient d'adopter un programme social. On veut savoir si un débat télévisé influence l'opinion des gens sur le programme. On a un petit échantillon de 15 personnes. On évalue leur opinion sur une échelle à 3 valeurs (1 = opinion négative ; 2 = indécis ; 3 = opinion positive) avant et après le débat télévisé. À l'aide du logiciel *SPSS*, en effectuant un test du signe, déterminez si l'on peut avancer que le débat a eu une incidence favorable. Les données se trouvent dans le fichier « Test du signe – débat ».

- 2) On a un échantillon d'élèves qui en sont à la première année de leurs études secondaires. On leur a demandé d'estimer leur compétence dans leur langue maternelle, qui est aussi celle de l'école. Il leur est demandé de juger de leur compétence par rapport à leurs camarades de classe et par rapport à leurs professeurs. Une valeur de 1 signifie « inférieure », de 2, « équivalente », de 3, « supérieure ». À l'aide du logiciel *SPSS*, et en recourant au test du signe, déterminez s'il y a une différence entre ces deux estimations. Les données se trouvent dans le fichier « Test du signe – compétence linguistique ».

EXERCICES

16 - LE TEST DES RANGS SIGNÉS DE WILCOXON

Effectuez les exercices suivants.

- 1) On a un échantillon d'élèves qui sont tous en classe terminale de leurs études secondaires. On leur demande d'évaluer leur propre compétence dans leur langue maternelle, le français, qui est aussi la langue de l'école qu'ils fréquentent. Pour les fins de cette estimation, ils ont une échelle qui va de 1, « mauvaise », à 6, « excellente ». Il leur est demandé d'effectuer cette estimation pour ce qui est de lire et de parler. À l'aide du logiciel *SPSS*, et en recourant au test des rangs signés de Wilcoxon, déterminez s'il y a une différence entre ces deux registres d'auto-estimation. Les données se trouvent dans le fichier « Test des rangs signés – aptitude à lire et parler ».

- 2) Un intervenant social offre des séances de sensibilisation au racisme. Il veut découvrir si les séances qu'il propose ont quelque efficacité. Il se donne un échantillon de 18 personnes. Juste avant que ne débute une séance, celles-ci sont invitées à estimer leur niveau de racisme sur une échelle de 1 à 10, 1 signifiant « pas du tout raciste » et 10 signifiant « très raciste ». Elles referont le même exercice immédiatement après la séance. À l'aide du logiciel *SPSS*, et en employant le test des rangs signés de Wilcoxon, vérifiez s'il y a une différence pour cette estimation entre les deux moments. Les données se trouvent dans le fichier « Test des rangs signés – racisme ».

EXERCICES

17 - LE TEST DE FRIEDMAN

Effectuez les exercices suivants.

- 1) On a un échantillon d'élèves qui sont tous en classe terminale du secondaire. On leur demande d'évaluer leur propre compétence dans leur langue maternelle, le français, qui est aussi la langue de l'école qu'ils fréquentent. Pour les fins de cette estimation, ils ont une échelle qui va de 1, « mauvaise », à 6, « excellente ». Il leur est demandé d'effectuer cette estimation pour ce qui est de lire, de parler et d'écrire. À l'aide du logiciel *SPSS*, et en recourant au test de Friedman, déterminez s'il y a une différence entre ces trois registres d'auto-estimation. Les données se trouvent dans le fichier « Test de Friedman – aptitude à lire parler et écrire ».

- 2) Un intervenant social offre des séances de sensibilisation au racisme. Il veut découvrir si les séances qu'il propose ont quelque efficacité. Il se donne un échantillon de 18 personnes. Juste avant que ne débute une séance, celles-ci sont invitées à estimer leur niveau de racisme sur une échelle de 1 à 10, 1 signifiant « pas du tout raciste » et 10 signifiant « très raciste ». Elles referont le même exercice immédiatement après la séance et de nouveau six semaines plus tard. À l'aide du logiciel *SPSS*, et en employant le test de Friedman, vérifiez s'il y a une différence pour cette estimation entre les trois moments. Les données se trouvent dans le fichier « Test de Friedman – racisme ».

EXERCICES

18 - LA CORRÉLATION DE PEARSON

Effectuez les exercices suivants.

- 1) Prenons les joueurs de la Ligne nationale de hockey et retenons-les en fonction du nombre de points qu'ils ont obtenus durant la saison régulière 2014-2015, pourvu qu'ils aient joué au moins 70 des 82 matchs. Associons à cela le nombre de points qu'ils ont obtenus durant la saison régulière 2013-2014 selon les mêmes conditions. Conservons les 45 meilleurs pointeurs de cet ensemble pour la saison 2014-2015. Avec l'aide du logiciel *SPSS*, déterminez la valeur de la corrélation de Pearson entre ces deux séries de données. Les données se trouvent dans le fichier « Corrélation de Pearson – Hockey ».

- 2) On a un échantillon d'élèves qui en sont à la dernière année de leurs études de niveau secondaire. Pour chacun d'eux, on connaît la moyenne des résultats scolaires, dans l'ensemble, pour la première demie de l'année scolaire. Chacun d'eux, par ailleurs, a estimé pendant combien d'heures, d'une façon générale, il s'adonnait, hebdomadairement, à des jeux vidéo. À l'aide du logiciel *SPSS*, trouvez quelle est la corrélation de Pearson entre ces deux séries de données. Les données se trouvent dans le fichier « Corrélation de Pearson – vidéo ».

EXERCICES

19 - LA CORRÉLATION DE SPEARMAN

Effectuez les exercices suivants.

- 1) Prenons les joueurs de la Ligne nationale de hockey et retenons-les en fonction de leur rang au registre du nombre de points qu'ils ont obtenus durant la saison régulière 2014-2015, pourvu qu'ils aient joué au moins 70 des 82 matchs. Associons à cela leur rang comme pointeur durant la saison régulière 2013-2014 selon les mêmes conditions. Conservons les 45 meilleurs pointeurs de cet ensemble pour la saison 2014-2015. Avec l'aide du logiciel *SPSS*, déterminez la valeur de la corrélation de Spearman pour ces deux séries de données. Les données se trouvent dans le fichier « Corrélation de Spearman – Hockey ».

- 2) Une sociopsychologue souhaite vérifier s'il y a quelque corrélation entre la classe de revenu et le niveau de bonheur. Le revenu se présente sur une échelle ordinale dont les modalités vont de 1, « moins de 10 000 par année », à 16, « 150 000 et plus par année ». Le bonheur apparaît sur une échelle à 10 niveaux qui sont des regroupements de scores d'une échelle qui va de 1 à 100 ; on a ainsi, dans le regroupement 1, « de 1 à 9 » et, dans le regroupement 10, de « 90 à 100 ». On a ces deux séries de données pour un échantillon aléatoire de 121 individus. À l'aide du logiciel *SPSS*, et en employant la corrélation de Spearman, déterminez si la sociopsychologue a raison de croire en cette corrélation. Les données se trouvent dans le fichier « Corrélation de Spearman – classe de revenu et bonheur ».

EXERCICES

20 - LA CORRÉLATION PARTIELLE ET SEMI-PARTIELLE

Effectuez l'exercice suivant.

Des étudiants ont été soumis à 3 tests. Les résultats de ces tests se trouvent dans le fichier « Corrélation partielle et semi-partielle – 3 tests ». À l'aide du logiciel *SPSS*, effectuez une analyse de corrélation partielle et une autre de corrélation semi-partielle entre les tests 2 et 3, puis interprétez les résultats.

EXERCICES

21 - L'ANALYSE DE RÉGRESSION

Effectuez les exercices suivants.

- 1) Un sociopsychologue s'intéresse aux déterminants de l'amitié. Il croit que le nombre d'amis qu'une personne compte dans son cercle dépend de la confiance qu'elle peut accorder aux autres. Il soumet un questionnaire à un échantillon d'individus âgés d'au moins 15 ans. Le questionnaire comporte la question suivante :

Combien d'amis proches avez-vous (c'est-à-dire de personnes avec qui vous n'êtes pas parent, mais avec qui vous êtes à l'aise, vous pouvez dire ce que vous pensez et à qui vous pouvez demander de l'aide)?

L'échelle va de 0 à 20, la valeur de 20 signifiant « 20 ou plus ». Le questionnaire comprend aussi des questions sur la confiance. Ces questions invitent les participants à se prononcer sur la confiance envers les membres de la famille, les gens du voisinage, les personnes avec qui on travaille ou va à l'école, et les inconnus. Avec cet ensemble de questions, le sociopsychologue crée un indice dont la valeur de 4 correspond à une « grande méfiance » et celle de 20, à une « grande confiance ». À l'aide du logiciel *SPSS*, et en effectuant une analyse de régression simple, observez si le sociopsychologue vérifie son hypothèse. Les données se trouvent dans le fichier « Régression linéaire – amis et confiance »¹.

- 2) Un sociopsychologue s'intéresse aux déterminants de l'amitié. Il croit que le nombre d'amis qu'une personne compte dans son cercle dépend de la confiance qu'elle peut accorder aux autres et de la satisfaction à l'endroit de la fréquence de la communication qu'elle peut avoir avec des membres de sa famille. Il soumet un questionnaire à un échantillon d'individus âgés d'au moins 15 ans. Le questionnaire comporte la question suivante :

Combien d'amis proches avez-vous (c'est-à-dire de personnes avec qui vous n'êtes pas parent, mais avec qui vous êtes à l'aise, vous pouvez dire ce que vous pensez et à qui vous pouvez demander de l'aide)?

L'échelle va de 0 à 20, la valeur de 20 signifiant « 20 ou plus ». Le questionnaire comprend aussi des questions sur la confiance. Ces questions invitent les participants à se prononcer sur la confiance envers les membres de la famille, les gens du voisinage, les personnes avec

¹ Les données relatives aux questions 1 à 4 de cette section proviennent de *L'enquête sociale générale, cycle 22 [Canada] 2008 : les réseaux sociaux*, de statistique Canada.

qui on travaille ou va à l'école, et les inconnus. Avec cet ensemble de questions, le sociopsychologue crée un indice dont la valeur de 4 correspond à une « grande méfiance » et celle de 20, à une « grande confiance ». Le questionnaire comprend aussi cette question :

En général, à quel point êtes-vous satisfait-e du nombre de fois que vous communiquez avec les membres de votre famille?

Une échelle de Likert à cinq niveaux est attachée à cette question. Elle va de 1, pour « très satisfait-e », à 5, pour « très insatisfait ». À l'aide du logiciel *SPSS*, et en utilisant une analyse de régression multiple à entrée forcée, observez si le sociopsychologue vérifie son hypothèse. Les données se trouvent dans le fichier « Régression linéaire – amis confiance famille ».

- 3) Un sociopsychologue s'intéresse aux déterminants de l'amitié. Il croit que le nombre d'amis qu'une personne compte dans son cercle dépend de la confiance qu'elle peut accorder aux autres et du fait qu'elle appartienne ou non à une minorité visible. Il soumet un questionnaire à un échantillon d'individus âgés d'au moins 15 ans. Le questionnaire comporte la question suivante :

Combien d'amis proches avez-vous (c'est-à-dire de personnes avec qui vous n'êtes pas parent, mais avec qui vous êtes à l'aise, vous pouvez dire ce que vous pensez et à qui vous pouvez demander de l'aide)?

L'échelle va de 0 à 20, la valeur de 20 signifiant « 20 ou plus ». Le questionnaire comprend aussi des questions sur la confiance. Ces questions invitent les participants à se prononcer sur la confiance envers les membres de la famille, les gens du voisinage, les personnes avec qui on travaille ou va à l'école, et les inconnus. Avec cet ensemble de questions, le sociopsychologue crée un indice dont la valeur de 4 correspond à une « grande méfiance » et celle de 20, à une « grande confiance ». Le questionnaire invite aussi le participant à indiquer s'il appartient à une minorité visible. Les données ont été codifiées de telle manière que la valeur de 0 marque la « non-appartenance à une minorité visible » alors que celle de 1 signale cette appartenance. À l'aide du logiciel *SPSS*, et en utilisant une analyse de régression multiple à entrée forcée, observez si le sociopsychologue vérifie son hypothèse. Les données se trouvent dans le fichier « Régression linéaire – amis confiance minorité ».

- 4) Un sociopsychologue s'intéresse aux déterminants de l'amitié. Il croit que le nombre d'amis qu'une personne compte dans son cercle dépend de plusieurs facteurs, liés à la satisfaction à l'égard de la fréquence de la communication avec les membres de la famille, à l'état de santé mentale et à la santé dans son ensemble, et au fait d'appartenir ou non à une minorité visible.

Il soumet un questionnaire à un échantillon d'individus âgés d'au moins 15 ans. Le questionnaire comporte la question suivante :

Combien d'amis proches avez-vous (c'est-à-dire de personnes avec qui vous n'êtes pas parent, mais avec qui vous êtes à l'aise, vous pouvez dire ce que vous pensez et à qui vous pouvez demander de l'aide)?

L'échelle va de 0 à 20, la valeur de 20 signifiant « 20 ou plus ».

Le questionnaire comprend aussi cette question, qui permet de mesurer la communication avec les membres de la famille :

En général, à quel point êtes-vous satisfait-e du nombre de fois que vous communiquez avec les membres de votre famille?

Une échelle de Likert à cinq niveaux est attachée à cette question. Elle va de 1, pour « très satisfait-e », à 5, pour « très insatisfait-e ».

Pour observer la santé mentale, le sociopsychologue construit un indice de confiance à partir de diverses questions qui invitent les participants à se prononcer sur la confiance envers les membres de la famille, les gens du voisinage, les personnes avec qui on travaille ou va à l'école et les inconnus. Avec cet ensemble de questions, il crée un indice dont la valeur de 4 correspond à une « grande méfiance » et celle de 20, à une « grande confiance ». Il pose aussi cette question : « En général, diriez-vous que votre santé mentale est... ? », à côté de quoi se trouve une échelle à cinq niveaux dont les pôles sont 1, pour « excellente », et 5, pour « mauvaise ». Le questionnaire interroge aussi le répondant de la sorte : « À l'aide d'une échelle de 1 à 10, quel sentiment éprouvez-vous à l'égard de votre vie? », où 1 veut dire « très insatisfait-e » et 10, « très satisfait-e ». En plus de cela, il demande « Vous décririez-vous comme étant habituellement (heureux-se et intéressé-e à la vie)? » ; l'échelle des réponses débute avec la valeur de 1 pour « heureux-se et intéressé-e à la vie », et se termine avec la valeur de 5 pour « si malheureux-se que la vie ne vaut pas la peine d'être vécue ».

Le sociopsychologue se donne les moyens d'observer la santé dans son ensemble à partir de ce libellé : « En général, diriez-vous que votre santé est... ? » ; les réponses possibles s'échelonnent de 1, pour « excellente », à 5, pour « mauvaise ».

Le questionnaire, enfin, invite le participant à indiquer s'il appartient à une minorité visible. Les données ont été codifiées de telle manière que la valeur de 0 marque la « non-appartenance à une minorité visible » alors que celle de 1 signale cette appartenance.

À l'aide du logiciel *SPSS*, en utilisant une analyse de régression multiple et en recourant à une technique d'entrée et d'exclusion combinée des variables indépendantes, découvrez le modèle que le sociopsychologue obtient. Les données se trouvent dans le fichier « Régression linéaire sélective – amis ».

- 5) Une spécialiste de la question sait que plus un immigrant compte d'années de scolarité, meilleur tend à être le revenu qu'il peut tirer de son emploi. Elle s'attend toutefois à ce que l'incidence des années d'instruction subisse l'influence de la connaissance de la langue du pays d'accueil. Pour vérifier son hypothèse, elle compose un échantillon de 58 immigrants. Pour chacun, elle connaît le revenu annuel et le nombre d'années de scolarité. Pour chacun d'eux aussi, elle établit la qualité de la langue du pays d'accueil sur une échelle qui va de 1, pour « mauvaise », à 5, pour « excellente ». Pour examiner l'interaction de la scolarité et de la connaissance de cette langue, elle crée une variable qui en est le produit. À l'aide du logiciel *SPSS*, déterminez si l'hypothèse de la spécialiste est juste. Les données se trouvent dans le fichier « Régression linéaire interaction – immigration scolarité langue ».

EXERCICES

22 - L'ANALYSE DE CHEMINEMENT

Effectuez l'exercice suivant.

On a des variables qui théoriquement influencent l'aspiration éducationnelle chez les étudiants de 1^{re} année universitaire. On leur demande d'évaluer leur intention de poursuivre des études supérieures, sur une échelle de 1 à 7. On a aussi de l'information sur les variables suivantes :

- Le rendement scolaire : la moyenne des cours du secondaire, en pourcentage ;
- la scolarité des parents : la moyenne des années de scolarité des deux parents ;
- l'intérêt de l'étudiant pour la recherche : mesuré sur une échelle 1 à 7 ;
- l'encouragement : la tendance chez les parents à encourager l'étudiant à faire de la recherche, mesurée sur une échelle de 1 à 7.

À l'aide du logiciel *SPSS*, dessinez un cheminement causal entre l'aspiration éducationnelle, d'une part, et les 4 facteurs, d'autre part. Les données sont transformées en scores z et se trouvent dans le fichier « Analyse de cheminement – aspiration ».

EXERCICES

23 - LA RÉGRESSION LOGISTIQUE

Effectuez les exercices suivants.

- 1) Un sociologue s'intéresse au phénomène du don. Il a en sa possession des données sur un échantillon de personnes pour lesquelles on sait si, oui ou non, elles ont fait un don en argent à un organisme au cours des douze derniers mois. Le sociologue n'a pas d'hypothèse précise quant aux facteurs relatifs à ce don, mais ses travaux lui donnent à croire que l'âge jouerait quelque rôle, les plus vieux étant plus enclins que les plus jeunes à faire de tels dons. Il lui semble aussi que la religion inciterait les individus à la générosité. La matrice dont il dispose comporte deux énoncés : « Sur le plan religieux, je suis une personne croyante » et « Sur le plan religieux, je suis une personne pratiquante » ; les positions apparaissent sur une échelle de Likert à 5 niveaux dont les pôles sont 1, pour « pas du tout d'accord », et 5, pour « tout à fait d'accord ». Il lui semble encore que la spontanéité, la tendresse et la compassion devraient intervenir. Dans la matrice, le sociologue trouve 3 énoncés pertinents auxquels sont attachées des échelles de Likert à 5 points, de 1, pour « pas du tout d'accord », à 5, pour « tout à fait d'accord ». À l'aide du logiciel *SPSS*, et en recourant à l'analyse de régression logistique binaire avec la méthode descendante Wald, découvrez les résultats qui se présentent au sociologue. Les données se trouvent dans le fichier « Régression logistique binaire – don ».
- 2) Une spécialiste de l'aménagement urbain s'interroge sur les facteurs qui font qu'une personne préférera établir sa résidence en ville, en banlieue ou à la campagne. Ses hypothèses veulent que la banlieue interpelle principalement les personnes d'âge moyen et qui ont des enfants ; qu'une affection prononcée pour la nature tende à faire privilégier la campagne ; que cette campagne attire peu les jeunes. Pour vérifier ses hypothèses, la spécialiste se donne un échantillon stratifié selon les trois lieux de résidence et se renseigne sur l'âge de chacun des individus, sur leur appréciation de la nature et sur le fait qu'ils ont (1) ou non (0) des enfants de moins de 15 ans. L'âge est donné dans sa simple cardinalité. Un énoncé « J'aime la nature » est lié à une échelle à six niveaux qui vont de 1, pour « pas du tout », à 6, pour « beaucoup ». À l'aide du logiciel *SPSS*, et en employant l'analyse de régression logistique multinomiale, découvrez quels résultats obtient la spécialiste.

3) Un spécialiste en marketing est invité par une compagnie de production de musique classique à définir un profil des amateurs de ce type de musique. Pour y parvenir, il se donne un échantillon d'individus qu'il répartit en quatre ensembles selon qu'ils sont 1) des non-amateurs du genre, 2) de faibles amateurs, 3) des amateurs intermédiaires ou 4) de grands amateurs. Son objectif est de repérer les déterminants de cette hiérarchisation de l'appréciation. Il croit que le goût pour la musique classique a pour corollaire l'appréciation du jazz et que l'intérêt pour la chanson et celui pour le rock seront négativement corrélés avec l'appréciation de la musique classique. Il suppose que l'instruction et l'âge rendent plus probable le goût du classique. Il soupçonne que le sexe n'aura pas d'incidence, mais, incertain, il choisit de le vérifier. Les appréciations pour le jazz, la chanson et le rock apparaissent sur des échelles de 1 à 6 où la valeur la plus faible signifie « pas du tout » et la plus forte, « beaucoup ». L'instruction se présente dans un mode ordinal :

- 1 : secondaire non terminé,
- 2 : diplôme d'études secondaires,
- 3 : diplôme d'études collégiales,
- 4 : diplôme de 1^{er} cycle universitaire,
- 5 : diplôme d'études supérieures.

L'âge est la valeur cardinale du nombre d'années. Pour la variable sexe, le chiffre 1 renvoie à la modalité féminine et le chiffre 2, à la modalité masculine. À l'aide du logiciel *SPSS*, et en utilisant l'analyse de régression logistique ordinale, voyez quel modèle le spécialiste en marketing peut dessiner. Les données se trouvent dans le fichier « Régression logistique ordinale – musique classique ».

EXERCICES

24 - LA RÉGRESSION DE COX

Effectuez l'exercice suivant.

On veut connaître des facteurs qui permettraient de prédire le moment où les abonnés d'un journal donné annulent leur abonnement. On dispose des informations suivantes sur 150 abonnés :

- statut : a mis fin ou non à son abonnement durant le dernier mois ;
- scolarité : le niveau de scolarité des abonnés avec les modalités 1 = secondaire et 2 = universitaire ;
- nombre d'autres journaux que lit l'individu ;
- nombre de mois pendant lesquels l'individu est abonné au journal.

À l'aide du logiciel *SPSS*, effectuez une régression de Cox avec la méthode « ascendante : rapport de vraisemblance » et interprétez les résultats principaux. Les données se trouvent dans le fichier « Régression de Cox – journal ».

EXERCICES

25 - LE COEFFICIENT DE CONCORDANCE W DE KENDALL

Effectuez les exercices suivants.

- 1) On a proposé un questionnaire à un échantillon de 220 jeunes adultes âgés de 24 à 27 ans. Ce questionnaire comporte 5 énoncés qui ont tous trait à la qualité des relations avec les autres :
 - Ma relation avec ma mère est harmonieuse ;
 - Ma relation avec les personnes de la communauté est harmonieuse ;
 - Ma relation avec mon père est harmonieuse ;
 - Ma relation avec mes ami-e-s est harmonieuse ;
 - Ma relation avec mes frères et sœurs est harmonieuse.

Une échelle de Likert est associée à chacun de ces énoncés et ses valeurs vont de 1, pour « pas du tout d'accord », à 6, pour « tout à fait d'accord ». La chercheuse estime que dans l'ensemble, ces lieux d'harmonie devraient converger, meilleure étant la relation avec un type de personne, meilleures étant les relations avec les autres. À l'aide du logiciel *SPSS*, et en utilisant la technique du coefficient de concordance de Kendall, déterminez si l'hypothèse de la chercheuse se justifie. Les données se trouvent dans le fichier « W de Kendall – relations avec les autres ».

- 2) On a les résultats scolaires moyens, sur 20, pour un échantillon de 32 élèves pour 6 années d'études. À l'aide du logiciel *SPSS*, déterminez si ces résultats sont corrélés sur l'ensemble des 6 années. Les résultats se trouvent dans le fichier « W de Kendall – résultats scolaires ».

EXERCICES

26 - L'ANALYSE DISCRIMINANTE

Effectuez l'exercice suivant.

Le directeur d'un club de sports veut savoir quels facteurs jouent un rôle dans l'adhésion à son club. Il a identifié trois groupes de gens : 20 membres habituels de son club, 15 membres qui ne l'ont été que sur une courte durée et 15 membres d'un autre club. Les facteurs qu'il considère sont les suivants :

- l'âge de la personne ;
- la distance (en kilomètres) entre son centre de sport et le domicile de la personne ;
- le sexe de la personne ;
- le revenu annuel de la famille à laquelle appartient la personne (en dollars).

À l'aide du logiciel *SPSS*, effectuez une analyse discriminante avec entrée simultanée des variables afin de découvrir si les 4 facteurs envisagés par le directeur permettent de bien classer les personnes dans les 3 groupes et, le cas échéant, indiquez l'importance relative de ces facteurs dans la classification. Les données se trouvent dans le fichier « Analyse discriminante – club de sports ».

EXERCICES

27 - L'ANALYSE FACTORIELLE

Effectuez les exercices suivants.

- 1) Pour mesurer l'agressivité, l'inattention et l'anxiété chez des élèves du primaire, un chercheur demande aux enseignants d'évaluer les comportements suivants de 306 enfants sur une échelle de 1 à 7, 1 signifiant « pas du tout » et 7 signifiant « tout à fait » :
- Agressivité :
 1. l'enfant intimide souvent les autres enfants
 2. l'enfant frappe souvent les autres enfants
 - Inattention :
 1. l'enfant abandonne facilement
 2. l'enfant se trouve souvent dans la lune
 - Anxiété :
 1. l'enfant pleure souvent
 2. l'enfant est souvent triste

À l'aide du logiciel *SPSS*, effectuez une analyse factorielle confirmatoire pour valider ces correspondances. Les données se trouvent dans le fichier « Analyse factorielle – agressivité inattention anxiété ».

- 2) 517 parents ont évalué leurs comportements parentaux en utilisant un questionnaire qui comporte les énoncés suivants :
- V1 : Lorsque mon enfant se comporte mal, je le (la) menace de punition, mais, à la fin, je ne le fais jamais ;
 - V2 : Je cède souvent à la demande de mon enfant ;
 - V3 : Je n'encourage pas mon enfant à contester mes idées ;
 - V4 : Lorsqu'il y a un désaccord entre moi et mon enfant, c'est moi qui ai le dernier mot ;
 - V5 : J'encourage mon enfant à participer à la définition des règles de la famille ;
 - V6 : Lorsqu'il y a un conflit entre moi et mon enfant, je l'encourage à raisonner avec moi ;
 - V7 : Je fais comprendre à mon enfant la conséquence de ne pas faire ses devoirs ;
 - V8 : Je punis mon enfant lorsqu'il (elle) ne respecte pas les règles de la famille.

Les parents se sont prononcés sur les énoncés sur une échelle de 1 à 7, 1 signifiant « pas du tout vrai » et 7, « tout à fait vrai ».

À l'aide du logiciel *SPSS*, effectuez une analyse factorielle afin de savoir si ces comportements peuvent se grouper dans quelques catégorisations distinctes. Les données se trouvent dans le fichier « Analyse factorielle – parentage ».

EXERCICES

28 - L'ANALYSE DE REGROUPEMENT

Effectuez l'exercice suivant.

On veut savoir si on peut regrouper 365 élèves du secondaire selon le rendement scolaire (la moyenne de tous les cours), le quotient intellectuel, la participation aux activités sportives (en nombre d'heures par semaine) et la compétence sociale (mesurée à partir d'un questionnaire, en scores de 20 à 50). Les valeurs de ces 4 mesures sont converties en scores z. À l'aide du logiciel *SPSS*, effectuez d'abord une analyse de regroupement avec le nombre de groupes prédéterminé à 2 et en utilisant la technique « nuées dynamiques », puis une seconde analyse sans prédéterminer le nombre de groupes, cette fois en recourant à la technique « Cluster TwoStep ». Les données se trouvent dans le fichier « Analyse de regroupement – élèves ».