

# Corrigés des exercices

L.M. Imbeau, *Statistiques sociales avec SPSS*

(Québec : PUL, 2008)

À l'usage exclusif des étudiants du cours POL14895 : Méthodes quantitatives.

©2008. Droits réservés.

## Chapitre 2 – Les statistiques et les variables

### *Analyse des données individuelles (WVS)*

#### **Question 1**

- a) ordinal
- b) 1 strong agree, 2 agree, 3 disagree, 4 strong disagree, 9 don't know
- c) 11,5
- d) 26,3
- e) 54,4
- f) 45

#### **Question 2**

- a) Ordinal
- b) Trop, Suffisamment, Trop peu, Ne sait pas
- c) 5,7%
- d) 27,7%
- e) 33,4%
- f) 68
- g) Qu'en concluez-vous?

La proportion des réponses « trop peu » laisse à penser qu'il existe un fort sentiment d'insatisfaction vis-à-vis des politiques du gouvernement concernant la pauvreté.

#### **Question 3**

- a) Ordinal
- b) very, rather, not very, not at all, don't know.
- c) Presque tous les répondants déclarent que la famille est très importante dans leur vie. Les répondants sont répartis inégalement, une faible proportion ayant répondu autre chose que « très importante », alors qu'une proportion négligeable ont répondu « peu » ou « pas du tout » importante.
- d) 1,2%
- e) 1487

**Question 4**

a) Nominal;

b) 1=Marié, 2=Comme marié, 3=Divorcé, 4=Séparé, 5=Veuf, 6=Célibataires, 9=Ne sais pas/Pas de réponse;

c) 20,0%

d) Décrivez le diagramme :

Le diagramme nous indique la forte présence des personnes mariées dans l'échantillon, suivie des personnes célibataires. Les veufs, les divorcés et les personnes vivant maritalement constituent une proportion semblable de cas. Les personnes séparées sont les moins nombreuses dans l'échantillon en question.

**Question 5**

a) Intervalles/Ratio

b) Continue

*Analyse de données agrégées***Question 1**

a) Ordinal.

b) libre, partiellement libre, non libre.

c) libre.

**Question 2**

a) intervalles/ratio

b) discrète

c) Chine

d) Liechtenstein

# Chapitre 3

## Les distributions de fréquences et de pourcentages

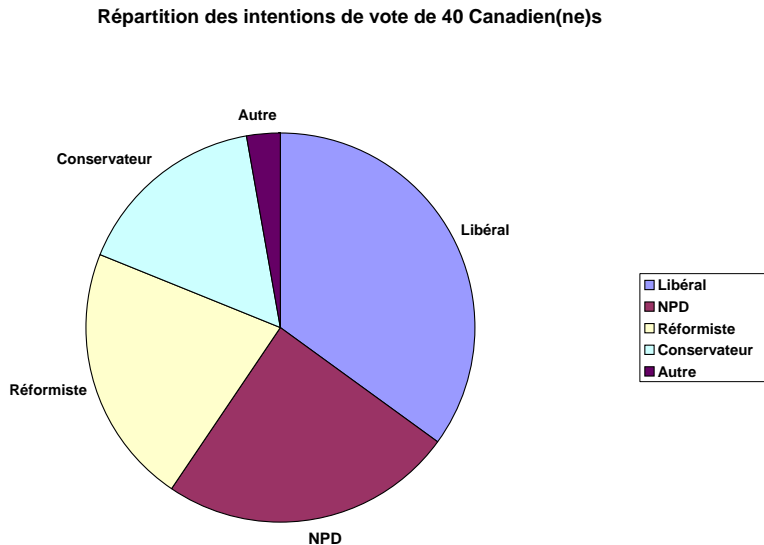
### *Exercices papier-crayon*

#### **Question 1**

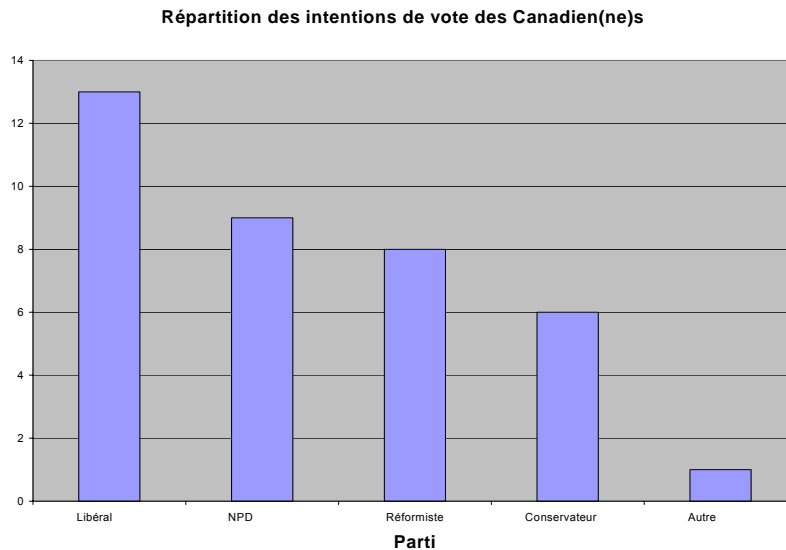
NPD – 9, Libéral – 13, Conservateur – 6, Réformiste – 8, Refus – 1, Invalide – 2, Autre – 1

Nous fusionnons 'Refus' et 'Invalide' comme données manquantes tout en gardant 'Autre' car cette catégorie inclut les intentions de vote pour les petits partis. C'est donc une donnée importante pour décrire les préférences des électeurs.

a) Diagramme circulaire pour la distribution de la variable 'intention de vote' :



b) La distribution des intentions de vote de 40 Canadiens



c) Tableau de la distribution des intentions de vote des Canadiens

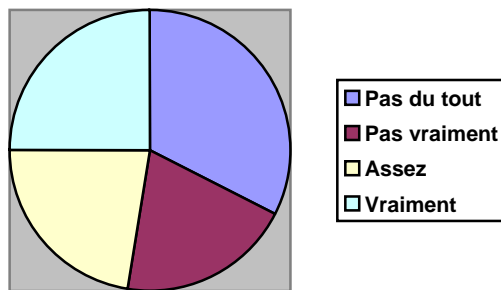
**Préférence pour le parti**

|          |              | Fréquence | Pourcentage | Pourcentage valide | Pourcentage cumulatif |
|----------|--------------|-----------|-------------|--------------------|-----------------------|
| Valide   | Autre        | 1         | 2,5         | 2,7                | 2,7                   |
|          | Conservateur | 6         | 15,0        | 16,2               | 18,9                  |
|          | Libéral      | 13        | 32,5        | 35,1               | 54,1                  |
|          | NPD          | 9         | 22,5        | 24,3               | 78,4                  |
|          | Réformiste   | 8         | 20,0        | 21,6               | 100,0                 |
|          | Total        | 37        | 92,5        | 100,0              |                       |
| Manquant |              | 3         | 7,5         |                    |                       |
| Total    |              | 40        | 100         |                    |                       |

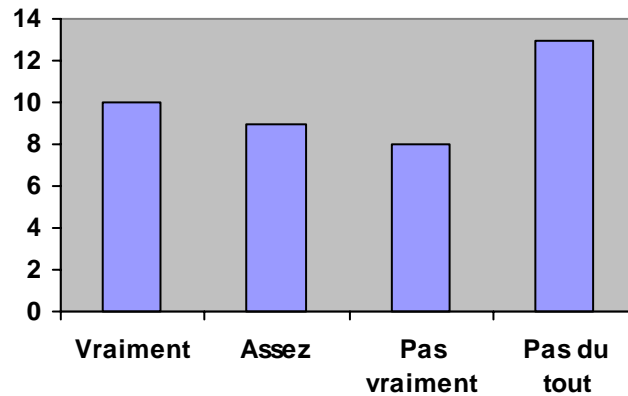
d) Ces quarante Canadiens sont plus nombreux à vouloir voter pour le parti libéral (35,1%) que pour tout autre parti. Les néo-démocrates (NPD) se situent 10,8 points derrière les libéraux à 24,3% dans leurs intentions de vote. Les réformistes puis les conservateurs suivent. En fusionnant les résultats des partis réformiste et conservateur pour obtenir une mesure du vote pour la droite, on voit que la distribution des intentions de vote dans cet échantillon est légèrement concentrée à la droite du spectre idéologique avec 37,8% à droite, 35,1% au centre et 24,3% à gauche.

**Question 2**

a) Graphique 1 : Importance de la religion dans la vie



b) Graphique 2 : Importance de la religion dans la vie



c)

Tableau 1 : Importance de la religion dans la vie des répondants

| Importance de la religion | Pourcentages |
|---------------------------|--------------|
| Vraiment importante       | 25           |
| Assez importante          | 23           |
| Pas vraiment importante   | 20           |
| Pas du tout importante    | 32           |
| Total                     | 100          |
| (N)                       | (40)         |

d) Les répondants sont distribués de manière quasi-égale sur les quatre valeurs. Pour la moitié des 40 répondants, la religion est assez ou vraiment importante, alors que pour l'autre moitié la religion n'est pas vraiment ou n'est pas du tout importante. Une seule valeur se démarque légèrement des autres, soit les répondants pour qui la religion n'est pas du tout importante, avec 13 répondants sur 40 (32 %).

*Analyse des données individuelles (WVS)*

**Question 1**

a) Ordinale

b) Discrète

c) 25-34 et 35-44 ans

d) Ceci dépend du but de la recherche. La nouvelle variable permet de distinguer les «personnes âgées» des autres adultes, et ces autres adultes entre eux par catégories d'âge. On peut ainsi voir plus clairement la répartition d'âge dans l'échantillon. Par contre les différences individuelles sont perdues car elles ont été agrégées dans les catégories plus larges. Il est à noter qu'il est possible, en fonction des finalités de la recherche, de regrouper les catégories de façon différente.

## Question 2

a) 21 ans et plus

b) 22,4

c) non, car des personnes différentes peuvent atteindre un niveau de scolarité donné à des âges différents.

## *Analyse de données agrégées*

### Question 1

a) 0,71

b) 82,4 %

### Question 2

Pour regrouper les valeurs d'une variable le plus important est de **réfléchir** (Quelles catégories semblent cohérentes ? Quelles catégories se prêtent mieux à l'analyse ? Quelles catégories conservent le plus d'information en rendant les données plus facilement interprétables ?)

Fox nous donne quelques conseils pour regrouper les catégories d'une variable :

- fusionner les valeurs d'une variable de façon à ce que cela soit cohérent à l'aune de nos objectifs ou de nos questions de recherche
- créer des catégories qui sont collectivement exhaustives et mutuellement exclusives (mais on peut exclure des données manquantes)
- ne pas conserver un nombre trop élevé de catégories (limiter à 6 ou 7)
- conserver des catégories homogènes, culturellement intelligibles et naturelles
- garder la même grandeur pour les catégories des variables d'intervalles/ratio
- respecter les conventions culturelles en créant, dans le cas des variables d'intervalles/ratio, des catégories en base 5 ou en base 10 (placer des catégories ouvertes à l'une des extrémités ou aux deux pour inclure les cas extrêmes)
- créer des catégories contenant un nombre sensiblement égal de cas.

Étant donné que sur 177 pays, le nombre maximal de pays ayant la même valeur était de trois, ce qui était ailleurs très rare, et que, généralement, chaque pays a un score unique, nous avons regroupé les pays en huit catégories, de façon à ce que les critères mentionnés ci-dessus soient respectés, et particulièrement en cherchant à obtenir une distribution équilibrée des cas, des catégories incluant les extrémités, ainsi qu'en respectant les conventions culturelles. Voici une façon de faire :

### STAB\_POL Indice de stabilité politique

|         |                    | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|--------------------|-----------|---------|---------------|--------------------|
| Valid   | 1.00 -1.5 et moins | 8         | 4.5     | 5.1           | 5.1                |
|         | 2.00 -1 à -1.49    | 13        | 7.3     | 8.3           | 13.4               |
|         | 3.00 -0.5 à -0.99  | 25        | 14.1    | 15.9          | 29.3               |
|         | 4.00 -0.01 à -0.49 | 25        | 14.1    | 15.9          | 45.2               |
|         | 5.00 0 à 0.49      | 31        | 17.5    | 19.7          | 65.0               |
|         | 6.00 0.5 à 0.99    | 26        | 14.7    | 16.6          | 81.5               |
|         | 7.00 1 à 1.49      | 26        | 14.7    | 16.6          | 98.1               |
|         | 8.00 1.5 et plus   | 3         | 1.7     | 1.9           | 100.0              |
|         | Total              | 157       | 88.7    | 100.0         |                    |
| Missing | -999.00            | 20        | 11.3    |               |                    |
| Total   |                    | 177       | 100.0   |               |                    |

Pour notre analyse, nous avons choisi l'Asie :

### STAB\_POL Indice de stabilité politique : Asie

|       |                 | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-----------------|-----------|---------|---------------|--------------------|
| Valid | 1 -1.5 et moins | 2         | 11.8    | 11.8          | 11.8               |
|       | 3 -0.5 à -0.99  | 1         | 5.9     | 5.9           | 17.6               |
|       | 4 -0.01 à -0.49 | 4         | 23.5    | 23.5          | 41.2               |
|       | 5 0 à 0.49      | 4         | 23.5    | 23.5          | 64.7               |
|       | 6 0.5 à 0.99    | 3         | 17.6    | 17.6          | 82.4               |
|       | 7 1 à 1.49      | 3         | 17.6    | 17.6          | 100.0              |
|       | Total           | 17        | 100.0   | 100.0         |                    |

L'analyse nous démontre que, parmi les 17 pays asiatiques analysés, la plupart ont un indice de stabilité politique positif. Cependant, les valeurs les plus fréquentes sont groupées dans les catégories autour de zéro.

# Chapitre 4

## Les mesures de tendance centrale

### *Exercices papier-crayon*

#### Question 1

| Variable | Niveau de mesure  | Mode        | Médiane | Moyenne |
|----------|-------------------|-------------|---------|---------|
| V216     | Intervalles/Ratio | pas de mode | 44      | 45,1    |
| V218     | Ordinale          | 10          | 10      | NP      |
| V219     | Nominale          | 2           | NP      | NP      |
| V220     | Nominale          | 1           | NP      | NP      |
| V221     | Nominale          | 3           | NP      | NP      |

#### Question 2

Médiane : 2,7

### *Analyse des données individuelles (WVS)*

#### Question 1

a)

Nominal

Mode : Catholiques

Médiane : NP

Moyenne : NP

b)

Intervalle/Ratio

Mode : 2

Médiane : 2

Moyenne : 2,43

c)

Ordinale

Mode : 2

Médiane : 2,26

Moyenne : NP



## Question 2

|         | ON    | QC    |
|---------|-------|-------|
| Mo      | 21    | 21    |
| Md      | 18.75 | 18.50 |
| Moyenne | 18.49 | 18.24 |

a) C'est en Ontario que les répondants ont étudié le plus longtemps puisque ils ont terminé leurs études plus tard, en moyenne (18,49 ans) qu'au Québec (18,24 ans). Cette conclusion est confirmée par les médianes (18,75 ans pour les Ontariens contre 18,50 ans pour les Québécois).

## Question 3

a) Il n'y a pas vraiment de différence entre l'opinion des hommes et l'opinion des femmes envers l'intervention gouvernementale. Les statistiques pour les deux groupes sont similaires. Pour 740 hommes, la moyenne est de 3,31, la médiane de 3, alors que pour les 715 femmes, la moyenne est de 3,42, la médiane de 3.

b) Selon les données de ces échantillons, les répondants de la Colombie-Britannique sont légèrement plus favorables que ceux du Québec à l'intervention gouvernementale dans l'économie. La moyenne pour la Colombie-Britannique (3,76) est légèrement plus élevée que pour le Québec (3,31).

c) La comparaison des mesures de tendance centrale nous indique que les Albertains sont plus en faveur du marché (moyenne = 2,93) que les habitants de la Colombie-Britannique (moyenne = 3,76). Le diagramme en bâtons nous apprend que la distribution est bi-modale en Colombie-Britannique alors qu'elle est uni-modale avec une forte asymétrie positive en Alberta : les répondants de la Colombie-Britannique sont aussi nombreux (23,8%) à répondre 1 (position de droite) que 5 (position mitoyenne) (23,2%), alors que plus de 40% des Albertains répondent «1». Dans les deux provinces, il y a très peu de répondants qui soutiennent qu'il devrait y avoir plus d'entreprises nationalisées. Les répondants de l'Alberta sont donc plutôt à droite sur cette question alors que ceux de la Colombie-Britannique sont plutôt au centre-droit.

## Question 4

a) Intervalles

b) Moyenne chez les hommes = 5,06

Moyenne chez les femmes = 4,85

Médiane chez les hommes = 5,32

Médiane chez les femmes = 5,09

c) Les femmes sont moins en faveur de l'euthanasie, mais la différence ne semble pas très importante.

## *Analyse des données agrégées (Pays2000)*

### **Question 1**

- a) Moyenne – 55,93 ; Médiane – 56,35
- b) La distribution est assez symétrique car les mesures de tendance centrale, à savoir la médiane et la moyenne se rejoignent.
- c) Hong Kong, Îles Caïman, Monaco, Singapour.
- d) Rwanda

### **Question 2**

- a) Moyenne : 1,62 ; Médiane : 1,75
  
- b) La distribution est légèrement asymétrique vers la gauche, car la moyenne est légèrement plus faible que la médiane. La statistique skewness (- 0,498) et l'histogramme confirment cette asymétrie vers la gauche. Quelques valeurs plus faibles tirent la valeur de la moyenne vers elles.
  
- c) L'Italie.
  
- d) La limite inférieure de la boîte tout comme la moustache inférieure sont plus éloignés de la médiane que la limite supérieure et la moustache supérieure, ce qui confirme l'asymétrie négative. Les données inférieures à la médiane sont plus étendues que les données supérieures. Par contre, il n'y a pas de valeur aberrante, toute la distribution est à l'intérieur des deux moustaches (aucune valeur ne se trouve à plus d'une fois et demie l'étendue interquartile en deçà du premier quartile ou au-delà du troisième).

# Chapitre 5

## Les mesures de variation

### *Exercices papier-crayon*

#### Question 1

| Cas             | Nombre d'heures ( $X_i$ ) | $(X_i - \bar{X})$ | $(X_i - \bar{X})^2$ |
|-----------------|---------------------------|-------------------|---------------------|
| 1               | 7                         | 1,4               | 1,96                |
| 2               | 2                         | -3,6              | 12,96               |
| 3               | 11                        | 5,4               | 29,16               |
| 4               | 5                         | -0,6              | 0,36                |
| 5               | 6                         | 0,4               | 0,16                |
| 6               | 2                         | -3,6              | 12,96               |
| 7               | 10                        | 4,4               | 19,36               |
| 8               | 3                         | -2,6              | 6,76                |
| 9               | 4                         | -1,6              | 2,56                |
| 10              | 6                         | 0,4               | 0,16                |
| Somme           | 56                        | 0                 | 86,40               |
| Moyenne         |                           |                   |                     |
| $\bar{X} = 5,6$ |                           |                   |                     |

Écart-type : 3,098      Variance : 9,6

#### Question 2

| Cas       | Nombre d'enfants ( $X_i$ ) | $(X_i - \bar{X})$ | $(X_i - \bar{X})^2$ |
|-----------|----------------------------|-------------------|---------------------|
| 1         | 1                          | -1,4              | 1,96                |
| 2         | 4                          | 1,6               | 2,56                |
| 3         | 3                          | ,6                | ,36                 |
| 4         | 2                          | -,4               | ,16                 |
| 5         | 0                          | -2,4              | 5,76                |
| 6         | 3                          | ,6                | ,36                 |
| 7         | 2                          | -,4               | ,16                 |
| 8         | 3                          | ,6                | ,36                 |
| 9         | 3                          | ,6                | ,36                 |
| 10        | 3                          | ,6                | ,36                 |
| Somme     | 24                         | 0                 | 12,4                |
| $\bar{X}$ | 2,4                        |                   |                     |

$s = 1,17 ; s^2 = 1,38$

### Question 3

- a)  $-0,34$  ; b)  $1,32$

### Question 4

- a) erreur-type =  $0,025$   
b) intervalle 99% =  $2,56$  à  $2,69$   
c) intervalle 95% =  $2,57$  à  $2,67$

### Question 5

| Cas | Score Z |
|-----|---------|
| 1   | 1,10    |
| 2   | - 1,57  |
| 3   | - 0,81  |
| 4   | - 0,42  |
| 5   | - 1,57  |
| 6   | - 1,19  |
| 7   | - 0,04  |
| 8   | 0,34    |
| 9   | - 0,81  |
| 10  | - 1,19  |

### Question 6

| Cas | Score Z |
|-----|---------|
| 1   | - 0,16  |
| 2   | 0,25    |
| 3   | - 0,99  |
| 4   | - 0,58  |
| 5   | 1,08    |
| 6   | 0,67    |
| 7   | - 0,16  |
| 8   | - 0,16  |
| 9   | 1,08    |
| 10  | - 0,99  |

### Question 7

Moyenne :  $75,5$       Écart-type :  $18,57$       Erreur-type :  $1,07$

### *Analyse des données individuelles (WVS)*

### Question 1

- a) Intervalles; Mo = 9; Md = 5,84; Moyenne = 5,75; s = 2,979;  $s^2 = 8,875$   
b) Nominal; Mo = 2; Md = NP; Moyenne = NP; s = NP;  $s^2 = NP$   
c) Intervalles/ratio; Mo = 0; Md = 1,43; Moyenne = 1,52; s = 1,384;  $s^2 = 1,916$   
d) Ordinal; Mo = 3; Md = 3; Moyenne = NP; s = NP;  $s^2 = NP$   
e) Nominal; Mo = 0 (not applicable); Md = NP; Moyenne = NP; s = NP;  $s^2 = NP$

## Question 2

- a) Écart-type : 2,69 ; Variance : 7,24
- b) Écart-type : 5,64 ; Variance : 31,76
- c) Les cas sont plus dispersés chez les Allemands. Les écarts-types sont de 5,64 pour les Allemands et de 2,69 pour les Belges.

## Question 3

- a) 2,83
- b) 3,08
- c) 1,02
- d) 1,19
- e) Les cas ne sont pas vraiment plus dispersés chez les hommes que chez les femmes, les coefficients de variation étant semblables : 36% chez les hommes, 38% chez les femmes.

## Question 4

- a) Le niveau de satisfaction est identique dans les deux cas
- b) Les moyennes de la satisfaction sont très proches (7,70 et 7,72)
- c) Chez les femmes mais la différence est minimale
- d) La variation chez les femmes atteint 3,43 versus 3,4 chez les hommes. Si on regarde le coefficient de variation il est de 23,98% chez les femmes contre 23,95% pour les hommes.

## Analyse des données agrégées (Pays2000)

### Question 1

Tableau 1 – Taux de stabilité politique dans les régions du monde

|                           | OCDE  | Europe de l'Est | Asie – Pacifique | Afrique | Moyen Orient | Amérique latine |
|---------------------------|-------|-----------------|------------------|---------|--------------|-----------------|
| a) Moyenne                | 1,24  | 0,16            | 0,13             | -0,64   | -0,07        | 0,04            |
| b) Écart-type             | 0,24  | 0,54            | 0,92             | 0,66    | 0,73         | 0,68            |
| c) Variance               | 0,06  | 0,29            | 0,85             | 0,44    | 0,54         | 0,47            |
| d) Indice de symétrie     | -0,27 | -0,78           | -0,97            | -0,03   | -0,09        | -0,31           |
| e) Indice d'aplatissement | -0,42 | -0,04           | 1                | -0,15   | -0,47        | -0,69           |
| f) CV                     | 19,35 | 337,5           | 707,69           | 103,13  | 1042,86      | 1700            |

- g) La région de l'OCDE est celle qui a la plus haute moyenne et le moins de dispersion. La moyenne la plus faible se retrouve en Afrique. Au niveau de l'homogénéité, trois groupes se distinguent : l'OCDE est dans une classe à part, l'Europe de l'Est et l'Afrique ont une dispersion moyenne tandis que l'Asie-Pacifique, le Moyen Orient et l'Amérique latine sont très hétérogènes.

S'il y a beaucoup de variations entre ces pays, il ne semble pas y avoir de valeurs extrêmes comme en Asie-Pacifique ou en Europe de l'Est, dont la distribution montre une asymétrie marquée vers la gauche. Une analyse détaillée des distributions permettrait de déterminer quels pays tirent la distribution vers la gauche dans les régions 2 et 3.

h) C'est dans les pays de l'OCDE que l'on retrouve le plus de stabilité politique, et c'est en Afrique que l'on en trouve le moins. Les autres régions du monde ont des niveaux de stabilité politique semblables. C'est aussi parmi les pays de l'OCDE que l'on trouve le plus d'homogénéité quant à la stabilité politique alors que l'Amérique latine, le Moyen-orient et l'Asie/Pacifique présentent une extrême variabilité.

### Question 2

|    | Europe de l'Est | OCDE |
|----|-----------------|------|
| a) | 2,5             | 1    |
| b) | 1,70            | 0    |
| c) | 2,89            | 0    |
| d) | 68%             | 0%   |

Les pays de l'OCDE présentent une constante au niveau de la liberté politique. C'est là que l'on trouve le niveau le plus élevé (niveau maximum) de liberté politique. Les pays de l'Europe de l'Est ont un niveau de liberté politique plus faible (la moyenne plus élevée indique moins de liberté politique). Dans les pays de l'Est il existait des pays qui n'avaient pas encore établi des structures nécessaires à la reconnaissance des libertés politiques. Afin d'en savoir davantage sur les niveaux de liberté politique, il serait nécessaire de regarder la distribution des scores à l'intérieur de l'échantillon et de comparer les pays ce qui mériterait une analyse plus détaillée.

# Chapitre 6

## L'analyse des tableaux bivariés

### *Exercices papier-crayon*

#### **Question 1**

- a) Variable indépendante : importance de la religion dans la vie
- b) Variable dépendante : fréquence à laquelle on pense au sens de la vie
- c) L'une des questions fondamentales auxquelles les religions cherchent à répondre concerne le sens de la vie : D'où venons-nous ? Où allons-nous ? On peut faire l'hypothèse que ceux qui considèrent la religion comme importante se préoccupent plus du sens de leur vie que les autres, soit parce qu'ils y sont prédisposés, soit parce qu'ils fréquentent des lieux de cultes où ces questions sont souvent abordées.

Tableau 6,1 : La fréquence des pensées sur le sens de la vie selon l'importance de la religion dans la vie (en pourcentage)

| Sens de la vie | Importance de la religion |              |             |             |
|----------------|---------------------------|--------------|-------------|-------------|
|                | pas du tout               | pas vraiment | assez       | vraiment    |
| 1 Souvent      | 12.5                      | <b>60.0</b>  | 22.2        | 25.0        |
| 2 Parfois      | <b>50.0</b>               | 26.7         | <b>55.6</b> | 25.0        |
| 3 Rarement     | <b>25.0</b>               | 6.7          | 11.1        | <b>25.0</b> |
| 4 Jamais       | 12.5                      | 6.7          | 11.1        | <b>25.0</b> |
| Total          | 100.0                     | 100.0        | 100.0       | 100.0       |
| (N)            | (8)                       | (15)         | (9)         | (8)         |

La relation est curvilinéaire. Elle est modérée. Parmi les personnes qui sont les plus tièdes vis-à-vis de la religion («pas du tout», «pas vraiment») la relation entre les deux variables est positive : ceux qui accordent plus d'importance à la religion pensent plus souvent au sens de la vie. En revanche, parmi les personnes qui sont plus religieuses («pas vraiment», «assez», «vraiment») la relation est négative.

#### **Question 2**

- a) Variable indépendante : le nombre d'heures consacrées à des activités physiques chaque semaine.
- b) Variable dépendante : les résultats scolaires.
- c) Il est bien connu que l'activité physique améliore l'état de santé d'une personne. Les personnes en santé peuvent fournir des efforts intellectuels plus intenses sur de plus longues périodes que les personnes fragiles. Les premières obtiennent ainsi de meilleurs résultats scolaires. Nous croyons donc que l'activité physique a un impact sur les résultats scolaires.

d)

| Résultats scolaires<br>(en %) | Nombre d'heures consacrées à des activités physiques chaque semaine |              |                  |
|-------------------------------|---|--------------|------------------|
|                               | Moins de 3 heures   | 3 à 6 heures | Plus de 6 heures |
| 90 et plus                    | 5   | 10           | <b>20</b>        |
| 75 à 89                       | 20  | <b>65</b>    | 50               |
| 60 à 74                       | <b>60</b>   | 20           | 20               |
| 59 et moins                   | <b>15</b>   | 5            | 10               |
| Total                         | 100   | 100          | 100              |
| (N)                           | (20)  | (20)         | (10)             |

Les pourcentages les plus élevés à l'intérieur des catégories de la variable dépendante (résultats scolaires) nous indiquent que la relation entre les deux variables est positive (plus on consacre un grand nombre d'heures à l'activité physique, meilleurs sont les résultats scolaires). La différence en points de pourcentage (jusqu'à 45 points de pourcentage) indique que la relation est forte.

### Question 3

- Variable indépendante : le sexe
- Variable dépendante : la valeur de travail
- Les hommes et les femmes n'accordent pas la même importance au travail pour des raisons qui sont sans doute liées à l'éducation et à la socialisation. En général, on s'attend à ce que les hommes accordent plus d'importance au travail que les femmes. (Il ne viendrait à l'esprit de personne de soutenir que l'importance accordée au travail - variable indépendante - détermine le sexe - variable dépendante !)

d)

| L'importance du travail (V8) | Le sexe     |             |
|------------------------------|-------------|-------------|
|                              | Homme       | Femme       |
| Vraiment                     | <b>62,6</b> | 54,5        |
| Assez                        | 26,8        | <b>32,6</b> |
| Pas vraiment                 | 4,8         | <b>7,6</b>  |
| Pas du tout                  | 5,6         | 5,3         |
| Total                        | 100,0       | 100,0       |
| (N)                          | (747)       | (736)       |

Les différences de pourcentage varient entre 0,3 points et 8,1 points. Ces différences sont trop faibles pour indiquer une relation quelconque entre les deux variables.



#### Question 4

- a) Variable indépendante : le statut socio-économique.  
b) Variable dépendante : l'acceptation de voisins homosexuels.  
c) Il est permis de penser que le statut socio-économique d'une personne a un impact sur son niveau de tolérance en général. Pour des raisons liées à la scolarisation (la fréquentation des écoles et des universités expose à une multiplicité de points de vue contribuant ainsi au développement de la tolérance) et à l'expérience de vie (les personnes plus fortunées voyagent plus, lisent plus, s'informent plus ; elles sont donc plus exposées à la multiplicité des points de vue), on s'attend à ce que les personnes à statut socio-économique plus élevé soient plus tolérantes envers les homosexuels.

d)

| Voisins<br>homosexuels<br>(n'aimerait<br>pas) (V60) | Statut socio-économique    |                     |              |              |
|---|----------------------------|---------------------|--------------|--------------|
|   | Ouvrier<br>Non<br>qualifié | Ouvrier<br>Qualifié | Moyen        | Élevé        |
| Mentionné   | 29                         | <b>40</b>           | 25           | 21           |
| Non<br>mentionné                                    | 71                         | 60                  | 75           | <b>79</b>    |
| Total<br>(N)  | 100<br>(183)               | 100<br>(558)        | 100<br>(589) | 100<br>(303) |

Telle qu'elle est construite, la variable dépendante mesure l'intolérance : les répondants qui ont mentionné les homosexuels parmi les personnes qu'ils n'aimeraient pas avoir comme voisin sont plus intolérants envers l'homosexualité que ceux qui n'ont pas mentionné les homosexuels. Les personnes plus intolérantes sont plus nombreuses chez les ouvriers qualifiés alors que les plus tolérantes sont plus nombreuses dans la classe élevée. Il y a donc une relation négative entre le niveau d'intolérance et le statut socio-économique. Cette relation est faible (différences de pourcentage de 19 points).

#### Question 5

- a) Variable indépendante : le milieu de résidence  
b) Variable dépendante : l'acceptation de voisins atteints du Sida  
c) Le lieu de résidence peut avoir un impact sur le niveau d'intolérance d'une personne. Parce qu'elles sont exposées à une variété d'expériences et de points de vue, les personnes qui vivent en ville sont susceptibles d'être moins intolérantes que les personnes vivant dans un milieu agricole. On peut donc s'attendre à ce que les premières manifestent plus d'ouverture envers les personnes atteintes du Sida.

d) Pourcentages :

| Voisins atteints<br>du sida<br>(n'aimerait pas)<br>(V58) | Milieu de résidence |       |           |
|--|---------------------|-------|-----------|
|  | Agricole            | Rural | Urbain    |
| Mentionné  | <b>29</b>           | 28    | 18        |
| Non mentionné  | 71                  | 72    | <b>82</b> |
| Total  | 100                 | 100   | 100       |
| (N)  | -41                 | -386  | -1267     |

En postulant que la variable «Milieu de résidence» est ordinale (c'est-à-dire qu'il y a un ordre intrinsèque entre les catégories de la variable), on arrive à la conclusion qu'il y a une relation négative entre l'intolérance et le caractère urbain du milieu de résidence : plus le milieu de résidence est urbain, plus l'intolérance est faible. La relation entre les deux variables est faible (différence de 11 points de pourcentage).

*Analyse de données individuelles*

**Question 1**

a) Variable indépendante : âge à la fin des études

b) Variable dépendante : importance de la religion

c) En exposant une personne à un grand nombre de points de vue, l'éducation formelle révèle le caractère relatif d'un grand nombre de «vérités» professées par les religions. On peut donc s'attendre à ce que les personnes qui ont étudié plus longtemps considèrent la religion comme moins importante.

L'importance de la religion selon l'âge à la fin des études

| La religion est-elle<br>importante? (V9) | Âge à la fin des études (V218) |           |                |
|--|--------------------------------|-----------|----------------|
|  | Moins de 16 ans                | 16-18 ans | Plus de 18 ans |
| 3 très                                   | <b>15.8</b>                    | 13.1      | 13.8           |
| 2 assez                                  | <b>34.9</b>                    | 30.3      | 22.3           |
| 1 pas ou peu                             | 49.3                           | 56.6      | <b>64.0</b>    |
| Total                                    | 100.0                          | 100.0     | 100.0          |
| (N)                                      | (272)                          | (350)     | (283)          |

La relation est négative (les personnes qui ont étudié plus longtemps sont plus nombreuses à dire que la religion est pas ou peu importante que les personnes qui ont étudié moins longtemps) et relativement faible (la différence des pourcentages est de 14 points).

## Question 2

- a) Variable indépendante : la langue parlée lors de l'entrevue.
- b) Variable dépendante : l'intention de combattre pour sa patrie en cas de guerre.
- c) Les francophones du Canada s'identifient-ils moins à l'État canadien que les anglophones au point d'être moins chauds à l'idée de combattre en cas de guerre ? C'est ce que veut découvrir l'hypothèse voulant que l'intention de combattre pour sa patrie en cas de guerre varie selon la langue parlée lors de l'entrevue. Ce serait l'appartenance à un groupe linguistique (variable indépendante) qui aurait un impact sur l'opinion concernant la volonté de participer à la guerre.
- d)

| Volonté de combattre (V110) | Langue parlée lors de l'interview (V235) |             |
|-----------------------------|--|-------------|
|                             | Anglais                                  | Français    |
| Oui                         | <b>75,7</b>                              | 41,4        |
| Non                         | 24,3                                     | <b>58,6</b> |
| Total                       | 100,0                                    | 100,0       |
| (N)                         | (936)                                    | (302)       |

Il y a une forte relation entre la langue et la volonté de participer à une guerre. Le fait de passer des répondants anglophones aux répondants francophones fait diminuer de 33 points de pourcentage le nombre de gens qui se disent prêt à se battre dans une guerre. Par ailleurs, il n'est pas pertinent de se demander si la relation est positive ou négative car nous ne pouvons ordonner les catégories nominales «anglophone» et «francophone».

## Question 3

- a) Variable indépendante : la province d'origine
- b) Variable dépendante : la tendance à l'enrôlement (volonté de combattre pour sa patrie)
- c) Les francophones du Canada s'identifient-ils moins à l'État canadien que les anglophones au point d'être moins chauds à l'idée de combattre en cas de guerre ? Comme les francophones sont concentrés dans la province de Québec, cette hypothèse veut que l'intention de combattre pour sa patrie en cas de guerre varie selon la province : les répondants du Québec devraient être moins en faveur de l'idée de combattre que les répondants résidant en Ontario ou en Colombie Britannique. La province de résidence (variable indépendante) aurait un impact sur l'opinion concernant la volonté de participer à la guerre.

d)

| Volonté de combattre<br>(V110) | Province (V234) |         |              |
|--------------------------------|-----------------|---------|--------------|
|                                | Québec          | Ontario | Colombie Br. |
| Oui                            | 42,5            | 77,4    | 68,8         |
| Non                            | 57,5            | 22,6    | 31,3         |
| Total                          | 100,0           | 100,0   | 100,0        |
| (N)                            | (320)           | (465)   | (128)        |

La relation entre la province de résidence et la volonté de combattre est relativement forte (différence de 34,9 points de pourcentage). Les répondants du Québec sont moins en faveur de l'idée de combattre pour défendre leur pays que les répondants des deux autres provinces. Par ailleurs, il n'est pas pertinent de se demander si la relation est positive ou négative car nous ne pouvons ordonner les catégories nominales «Québec» «Ontario» et «Colombie Britannique».

- e) Au Québec : 42,5%  
En Ontario : 77,4%  
En Colombie Britannique : 68,8%

Comme il est suggéré au paragraphe c) plus haut, le fait que les francophones canadiens soient concentrés au Québec explique ces différences : les francophones s'identifient moins à l'État canadien que les anglophones. Par conséquent, ils sont moins prêts à exposer leur vie pour le défendre.

### *Analyse de données agrégées (Pays2000)*

#### **Question 1**

a) Selon la théorie économique classique, le développement économique d'une nation repose sur l'initiative des personnes qui, en poursuivant leur propre intérêt, favorisent la création de richesse collective. Pour que cette initiative personnelle soit possible, il doit y avoir un minimum de liberté individuelle. De ce postulat découle l'hypothèse voulant que là où la liberté est plus grande, le développement économique soit plus grand

b) Variable indépendante : niveau de liberté de la population.

c) Variable dépendante : niveau de développement économique (PIB per capita).

d)

Le produit intérieur brut/hab. (\$) selon l'indice global de liberté

| PIB_TÊTE | Produit intérieur brut/hab. (\$) | LIBERTÉ Indice global de liberté |                     |             |
|----------|----------------------------------|----------------------------------|---------------------|-------------|
|          |                                  | Non libre                        | Partiellement libre | Libre       |
| 4        | 3000 et plus                     | 38.1                             | 36.8                | <b>94.0</b> |
| 3        | 2000-2999                        | 11.9                             | <b>17.5</b>         | 3.0         |
| 2        | 1000-1999                        | <b>31.0</b>                      | 26.3                | 1.5         |
| 1        | < 1000                           | 19.0                             | 19.3                | 1.5         |
| Total    |                                  | 100.0                            | 100.0               | 100.0       |
| (N)      |                                  | (42)                             | (57)                | (67)        |

e) Nos données nous permettent aussi de confirmer l'hypothèse d'Adejumobi. La relation entre les deux variables est positive, comme le prédisait l'hypothèse (le PIB per capita est d'autant plus élevé que l'indice global de liberté est élevé) et elle est très forte avec des différences de pourcentage atteignant 57,2 points.

f) Voici une hypothèse possible : le PIB per capita est plus élevé là où les mesures de contrôle de la corruption sont les plus développées.

Justification : Le développement économique est le fait de l'élite et non des masses. Ce n'est donc pas tant la liberté de la population que le contrôle des élites, en particulier le contrôle de la corruption, qui a un impact sur le développement économique. Le développement économique n'est possible que si les institutions permettent un contrôle efficace de la corruption.

# Chapitre 7

## Le test du khi-carré

### Exercices papier-crayon

#### Question 1

a) Fréquences anticipées :

| Résultats scolaires (en %) | Nombre d'heures consacrées à des activités physiques hebdomadairement |                        |                       | N  |
|----------------------------|---|------------------------|-----------------------|----|
|                            | Moins de 3 heures   | 3 à 6 heures           | Plus de 6 heures      |    |
| 90 et plus                 | 1<br>( <u>  2  </u> )   | 2<br>( <u>  2  </u> )  | 2<br>( <u>  1  </u> ) | 5  |
| 75 à 89                    | 4<br>( <u> 8,8 </u> )   | 13<br>( <u> 8,8 </u> ) | 5<br>( <u> 4,4 </u> ) | 22 |
| 60 à 74                    | 12<br>( <u> 7,2 </u> )  | 4<br>( <u> 7,2 </u> )  | 2<br>( <u> 3,6 </u> ) | 18 |
| 59 et moins                | 3<br>( <u>  2  </u> )   | 1<br>( <u>  2  </u> )  | 1<br>( <u>  1  </u> ) | 5  |
| Total                      | 20  | 20                     | 10                    | 50 |

b)  $H_0 : \chi^2 = 0$ , i.e., dans la population, les résultats scolaires ne varient pas en fonction du nombre d'heures consacrées chaque semaine à des activités physiques ; ou bien : il n'y a pas de relation entre ces deux variables.

c) khi-carré : 12,54 (12,5378788); dl=6; p: n.s.

Nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle il n'y a pas de lien dans la population entre les résultats scolaires et le nombre d'heures consacrées chaque semaine à l'activité physique. Avec 6 degrés de liberté un et seuil de signification de 0,05, nous devons avoir un khi-carré d'au moins 12,592 pour rejeter l'hypothèse nulle.

d) Beaucoup de fréquences anticipées sont inférieures à 5, en fait 8 sur 12 (67 %), ce qui est largement supérieur au seuil de 20 % des cellules contenant une fréquence anticipée inférieure à 5. Dans un tel cas, le test du khi-carré n'est pas valide. C'est pourquoi avons indiqué «n.a» pour «ne s'applique pas» au lieu du seuil de signification.

Si nous fusionnons les catégories de la variable dépendante en deux groupes (76 et plus, 75 et moins), il ne restera qu'une seule cellule sur 6 contenant une fréquence anticipée inférieure à 5. Le test serait alors valide.

## Question 2

a) Fréquences anticipées :

| L'importance<br>du travail<br>(V8) | Le sexe                 |                         | N    |
|------------------------------------|-------------------------|-------------------------|------|
|                                    | Masculin                | Féminin                 |      |
| Vraiment                           | 540<br>( <u>503,3</u> ) | 473<br>( <u>509,7</u> ) | 1013 |
| Assez                              | 230<br>( <u>255,4</u> ) | 284<br>( <u>258,6</u> ) | 514  |
| Pas vraiment                       | 41<br>( <u>52,7</u> )   | 65<br>( <u>53,3</u> )   | 106  |
| Pas du tout                        | 45<br>( <u>44,7</u> )   | 45<br>( <u>45,3</u> )   | 90   |
| Total                              | 856                     | 867                     | 1723 |

a) Fréquences anticipées : Le sexe M/F

b)  $H_0 : \chi^2 = 0$ , i.e., dans la population, il n'y a pas de relation entre le sexe et l'importance accordée au travail.

c) khi-carré : 15,51 dl = 3 ;  $p < 0,01$

À un seuil de signification de 0,01, nous pouvons rejeter l'hypothèse nulle selon laquelle il n'y a pas de lien dans la population entre le sexe et l'importance accordée au travail. Les 3 degrés de liberté un et seuil de signification de 0,01 nous demande de disposer d'un khi-carré d'au moins 11,345 pour rejeter l'hypothèse nulle ce que nous pouvons faire. La relation entre le sexe et l'importance accordée au travail est donc significative au seuil 0,01.

## Question 3

a) Fréquences anticipées :

| Voisins<br>homosexuels<br>(n'aimerait pas)<br>(V60) | Statut socio-économique |                       |                       |                       | N    |
|---|-------------------------|-----------------------|-----------------------|-----------------------|------|
|   | Ouvrier<br>non qualifié | Ouvrier<br>qualifié   | Statut<br>moyen       | Statut<br>élevé       |      |
| Mentionné   | 53<br>( <u>55</u> )     | 221<br>( <u>167</u> ) | 150<br>( <u>176</u> ) | 64<br>( <u>91</u> )   | 488  |
| Non mentionné                                       | 130<br>( <u>128</u> )   | 337<br>( <u>391</u> ) | 439<br>( <u>413</u> ) | 239<br>( <u>212</u> ) | 1145 |
| Total   | 183                     | 558                   | 589                   | 303                   | 1633 |

- b)  $H_0 : \chi^2 = 0$ , i.e., dans la population, il n'y a pas de relation entre le statut socio-économique et le sentiment envers les voisins homosexuels.
- c)  $(f_o - f_a)^2 / f_a = 0,07 ; 0,07 ; 17,46 ; 7,46 ; 3,84 ; 1,64 ; 8,01 ; 3,44$ .  
 Khi-carré = 41,99 ; dl = 3 ;  $p < 0,001$ .

Oui, l'hypothèse nulle devrait être rejetée. L'acceptation des voisins homosexuels est reliée au statut socio-économique, et cette relation est significative au seuil 0,001.

#### Question 4

a) Fréquences anticipées :

| Voisins atteints du sida (V58) | Milieu de résidence   |                         |                           | N    |
|--------------------------------|-----------------------|-------------------------|---------------------------|------|
|                                | Agricole              | Rural                   | Urbain                    |      |
| N'aimerait pas                 | 12<br>( <u>8,5</u> )  | 107<br>( <u>80,2</u> )  | 233<br>( <u>263,3</u> )   | 352  |
| Indifférent                    | 29<br>( <u>32,5</u> ) | 279<br>( <u>305,8</u> ) | 1034<br>( <u>1003,7</u> ) | 1342 |
| Total                          | 41                    | 386                     | 1267                      | 1694 |

b)  $H_0 : \chi^2 = 0$ , i.e., l'acceptation de voisins atteints du SIDA n'est pas reliée au milieu de résidence.

c) khi-carré = 17,5; dl = 2;  $p < 0,001$

Avec un khi-carré de 17,5 et 2 degrés de liberté, nous pouvons rejeter l'hypothèse nulle au seuil de 0,001. Cela signifie que la relation entre le milieu de résidence et l'acceptation de voisins atteints du SIDA est statistiquement significative au seuil 0,001.

#### *Analyse de données individuelles*

##### Question 1

Pour 3 catégories d'âge et trois catégories de la variable v9 (selon les transformations suggérées au chapitre précédent) :

Khi-carré : 13,96 ; dl : 4 ; sig 0,007.

##### Question 2

Khi-carré : 122,60 ; dl : 1 ; sig : 0,001.

##### Question 3



Khi-carré : 101,755 ; dl : 2 ; sig : 0,001.

*Analyse de données agrégées*

**Question 1**

a) Plus le gouvernement est fragmenté, plus le taux de croissance du PIB est faible.

b)  $H_0: \chi^2 = 0$ , i.e., il n'y a pas de relation entre la fragmentation du gouvernement et le taux de croissance du PIB.

**Case Processing Summary**

|                   | Cases |         |         |         |       |         |
|-------------------|-------|---------|---------|---------|-------|---------|
|                   | Valid |         | Missing |         | Total |         |
|                   | N     | Percent | N       | Percent | N     | Percent |
| croiss2 * gofrac2 | 143   | 80,8%   | 34      | 19,2%   | 177   | 100,0%  |

c)

**croiss2 \* gofrac2 Crosstabulation**

|         |                  |                  | gofrac2 |                      |             | Total  |
|---------|------------------|------------------|---------|----------------------|-------------|--------|
|         |                  |                  | ,00 0   | 1,00 0,1<br>- 0,4999 | 2,00 > 0,50 |        |
| croiss2 | 3,00 > 4,1       | Count            | 0       | 0                    | 1           | 1      |
|         |                  | % within gofrac2 | ,0%     | ,0%                  | 2,7%        | ,7%    |
|         | 2,00 0 à 4       | Count            | 65      | 34                   | 30          | 129    |
|         |                  | % within gofrac2 | 94,2%   | 91,9%                | 81,1%       | 90,2%  |
|         | 1,00 -1,5 à -0,1 | Count            | 4       | 3                    | 6           | 13     |
|         |                  | % within gofrac2 | 5,8%    | 8,1%                 | 16,2%       | 9,1%   |
| Total   |                  | Count            | 69      | 37                   | 37          | 143    |
|         |                  | % within gofrac2 | 100,0%  | 100,0%               | 100,0%      | 100,0% |

**Chi-Square Tests**

|                                 | Value              | df | Asymp. Sig.<br>(2-sided) |
|---------------------------------|--------------------|----|--------------------------|
| Pearson Chi-Square              | 6,269 <sup>a</sup> | 4  | ,180                     |
| Likelihood Ratio                | 5,849              | 4  | ,211                     |
| Linear-by-Linear<br>Association | 1,502              | 1  | ,220                     |
| N of Valid Cases                | 143                |    |                          |

a. 5 cells (55,6%) have expected count less than 5. The minimum expected count is ,26.

d) Khi-carré = 6,269; dl = 4; sig. = 0,180

e) L'analyse montre qu'il n'y pas de relation entre les deux variables. Les différences de pourcentage indiquent une relation négative faible. Le test du Khi-carré montre que cette relation n'est pas significative. 55,6% des cellules ont une fréquence anticipée inférieure à 5. Les résultats du test du khi-carré sont donc peu convaincants.

## Question 2

a) Plus il y a de médecins, plus le taux de mortalité infantile est faible.

b)  $H_0: \chi^2 = 0$ , i.e., il n'y a pas de relation entre le nombre de médecins et le taux de mortalité infantile.

c)

Tableau 7,2: Mortalité infantile selon le nombre de médecins (pourcentages)

| Mortalité infantile | Médecins pour mille habitants |             |             |
|---------------------|-------------------------------|-------------|-------------|
|                     | < 0,95                        | 1 à 2,7     | > 2,8       |
| >50                 | <b>72,3</b>                   | 6,7         | 0           |
| 10 à 49             | 26,5                          | <b>61,7</b> | 52,9        |
| < 10                | 1,2                           | 31,7        | <b>47,1</b> |
| Total               | 100                           | 100         | 100         |
| (N)                 | (83)                          | (60)        | (34)        |

La relation entre le nombre de médecins et le taux de mortalité infantile est négative (le taux de mortalité infantile est d'autant plus élevé que le nombre de médecins est faible). Cette relation est très forte avec des différences de pourcentages allant de 35,2 à 72,3 points.

d) Khi-carré = 98,112 ; dl = 4 ; sig < 0,001.

e) L'analyse montre qu'il y a une forte relation négative entre le taux de mortalité infantile et le nombre de médecins et que cette relation est statistiquement significative au seuil 0,001.

# Chapitre 8

## Mesures d'association

### pour des données de tableau croisé

#### *Exercices papier-crayon*

#### **Question 1**

a) Khi-carré = 12,53 ; V = 0,35 ; Gamma = 0,50

b) La relation est non significative, ce qui veut dire qu'on ne peut affirmer que la relation trouvée dans l'échantillon est généralisable à la population. Pour qu'elle soit significative à un seuil de 0,05, le khi-carré aurait dû être d'au moins 12,592, alors qu'il est de 12,53.

La relation dans l'échantillon étudié est positive et relativement forte. Le gamma est plus approprié parce que nous étudions deux variables ordinales. Mais comme la relation n'est pas significative, nous la considérons comme nulle.

#### **Question 2**

a) Khi-carré= 15,47 ; V=0,10 ; Gamma= 0,15 (le gamma est non pertinent parce que l'une des variables est nominale)

b) La relation est statistiquement significative avec le seuil de 0,01 car le khi-carré est de 15,47 pour 3 degrés de liberté. L'intensité de la relation est très faible. Il n'est pas pertinent de parler de la direction de la relation parce que V0 (SEX) est une variable nominale.

#### **Question 3**

a)  $\chi^2=41,82$  ; V=0,16 (puisque V60, qui mesure la tolérance envers les homosexuels, est une variable ordinale, le V de Cramer est non pertinent)  
G = 0,21

b) La relation entre l'intolérance et le statut socio-économique est négative. Les répondants au statut socio-économique élevé ont tendance à montrer moins d'intolérance. L'intensité, avec un G de - 0,21, est faible. La relation est significative au seuil 0,001.

#### **Question 4**

a) Khi-carré = 17,48 ; V = 0,10 ; Gamma = - 0,26

b) La relation entre les deux variables est négative (les urbains sont moins intolérants) et significative au seuil 0,001, mais elle est plutôt faible avec un Gamma de - 0,26.

## *Analyse de données individuelles (WVS)*

### **Question 1**

Importance de la religion (V9) selon l'âge à la fin des études (V218) (France)

a) Transformations effectuées :

Nous avons fusionné les scores d'âge en créant quatre catégories (pour avoir un tableau carré) ce qui a donné : 1 pour 12 ans et moins, 2 pour 13-16 ans, 3 pour 17-19 et 4 pour 20 ans et plus. Du côté de la variable dépendante nous avons gardé les quatre catégories.

b) Valeur du khi-carré : 35,442 ; dl=9.

c) La relation est statistiquement significative au seuil 0,001.

d) Comme les deux variables sont ordinales et puisque nous savons que V9 est la variable dépendante, nous choisissons une mesure directionnelle (ou non symétrique), le D de Somers.  $D = 0,11$ .

### **Question 2**

L'enrôlement (V110) selon la langue (V235) (Canada)

a) Transformations effectuées :

Nous avons considéré qu'il était inutile de procéder à des transformations, le nombre de catégories de chaque variable nous paraissant approprié.

b) Valeur du khi-carré : 122,6 ; dl=1

c) La relation est statistiquement significative au seuil 0,001.

d) Puisqu'il s'agit d'un tableau carré avec deux colonnes et deux rangées et qu'au moins une variable est nominale, nous avons choisi le « fi » qui est égal à 0,32.

### **Question 3**

L'enrôlement selon la province (Canada)

a) Transformations effectuées :

Comme dans les chapitres précédents, nous avons choisi trois provinces canadiennes : le Québec, l'Ontario et la Colombie britannique. L'autre variable étant dichotomique (réponses Oui/Non) elle est demeurée inchangée.

b) Valeur du khi-carré : 101,76 ; ddl=2

c) La relation est statistiquement significative au seuil 0,001.

d) Puisque nous avons affaire à au moins une variable nominale, nous avons retenu le V de Cramer dont la valeur est égale à 0,33. La relation est relativement forte.

### **Question 4**

a) L'âge à la fin des études.

b) La tolérance aux fraudes fiscales.

- c) H : les répondants plus instruits ont tendance à trouver les fraudes fiscales justifiées plus souvent que les répondants moins instruits.
- d) Comme les deux variables sont ordinales et puisque nous savons que V194 est la variable dépendante, nous choisissons une mesure directionnelle, le D de Somers.
- e)  $D_{yx}=0,17$ .
- f) La relation est positive de faible intensité.
- g)  $\chi^2=53,89$  ;  $dl=6$  ;  $p<0,001$ . La relation est statistiquement significative au seuil 0,001 car le khi-carré calculé est supérieur au khi-carré critique donné dans le tableau de la distribution du khi-carré pour 6 degrés de liberté et un niveau de signification de 0,001 (22,457).
- h) Oui, l'hypothèse de départ est confirmée. Le D de Somers indique une relation positive faible (les répondants plus instruits sont plus tolérants à la fraude fiscale) et le test du khi-carré indique que la relation est significative au seuil 0,001.

### Question 5

- a) L'importance de la religion dans la vie
- b) L'appui au recours à l'avortement.
- c) Plus la religion est importante dans la vie d'un individu, moins il appuie le recours à l'avortement.
- d) Comme les deux variables sont ordinales et puisque nous savons que V199 est la variable dépendante, nous choisissons une mesure directionnelle, le D de Somers.
- e)  $D = - 0,42$ .
- f) La relation est négative et relativement forte. Plus la religion est importante dans la vie d'un individu, moins il aura tendance à considérer que le recours à l'avortement est justifiable.
- g) Avec un khi-carré de 380,127 la relation est statistiquement significative au seuil 0,001 car le khi-carré calculé est supérieur au khi-carré critique donné dans le tableau de la distribution du khi-carré pour 9 degrés de liberté et un niveau de signification de 0,001 (27,877).
- h) Oui, l'hypothèse de départ est confirmée. Le D de Somers indique une relation négative relativement forte (les répondants qui accordent plus d'importance à la religion appuient moins le recours à l'avortement) et le test du khi-carré indique que la relation est significative au seuil 0,001.

### Question 6

- a) Le statut socio-économique
- b) Le sexisme quant à l'accès au travail.
- c) Les répondants à statut socio-économique élevé sont moins sexistes quant à l'accès au travail.

- d) Comme les deux variables sont ordinales et puisque nous savons que V61 est la variable dépendante, nous choisissons une mesure directionnelle, le  $D_{yx}$  de Somers.
- e)  $D_{yx} = -,08 [-,076]$ .
- f) La relation est négative et très faible voire nulle.
- g) Non, le khi-carré obtenu (14,06) est plus petit que le khi-carré critique donné dans le tableau de la distribution du khi-carré pour 8 degrés de liberté et un niveau de signification de 0,05 (15,507).
- h) Non, l'hypothèse de départ n'est pas confirmée car la relation n'est pas statistiquement significative au seuil 0,05 et l'intensité de la relation est presque nulle. Il n'y a pas de relation entre les variables.

# Chapitre 9

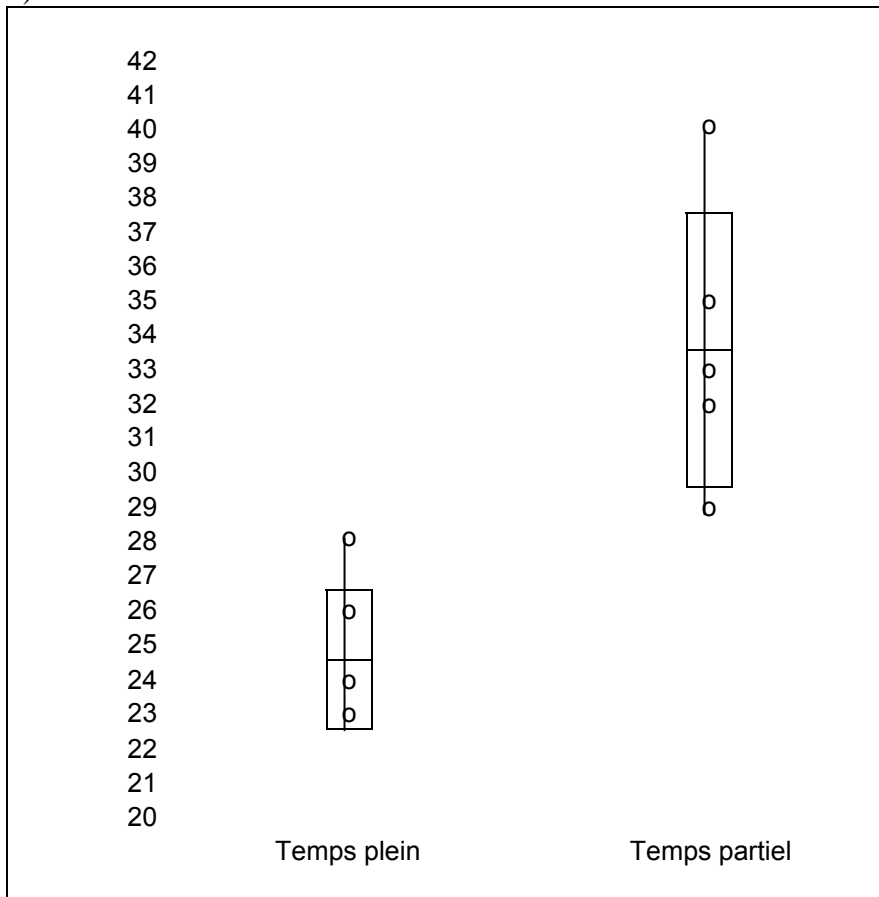
## Comparaison de moyennes et test $t$

### Exercices papier-crayon

*Note : les résultats présentés pour les exercices papier-crayon sont ceux des calculs faits à la main. Si vous créez votre propre matrice et procédez à des calculs avec SPSS vous arriverez à un «  $t$  » légèrement supérieur, ce qui est dû à une différence dans les algorithmes de calcul utilisés par le logiciel.*

### Question 1

a)

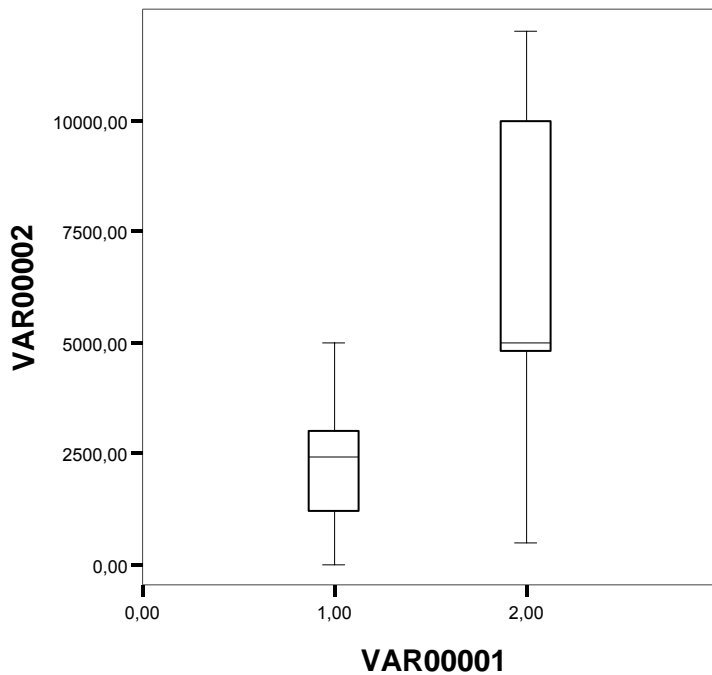


- b) Moyennes : 33,8 et 25,4  
Différence de moyennes : 8,4  
Variances : 16,7 et 3,8  
 $t = 3,7105$  ;  $dl = 8$  ;  $p < 0,01$

c) Les données permettent de voir que les travailleurs à temps plein sont significativement plus rapides que les travailleurs à temps partiel. La différence entre les deux groupes (8,4 secondes) est significativement différente de zéro au seuil 0,01.

## Question 2

a)



b) Moyennes : 2320 et 6460

Différence entre les moyennes : - 4140

Variances : 3 572 000 et 20 908 000

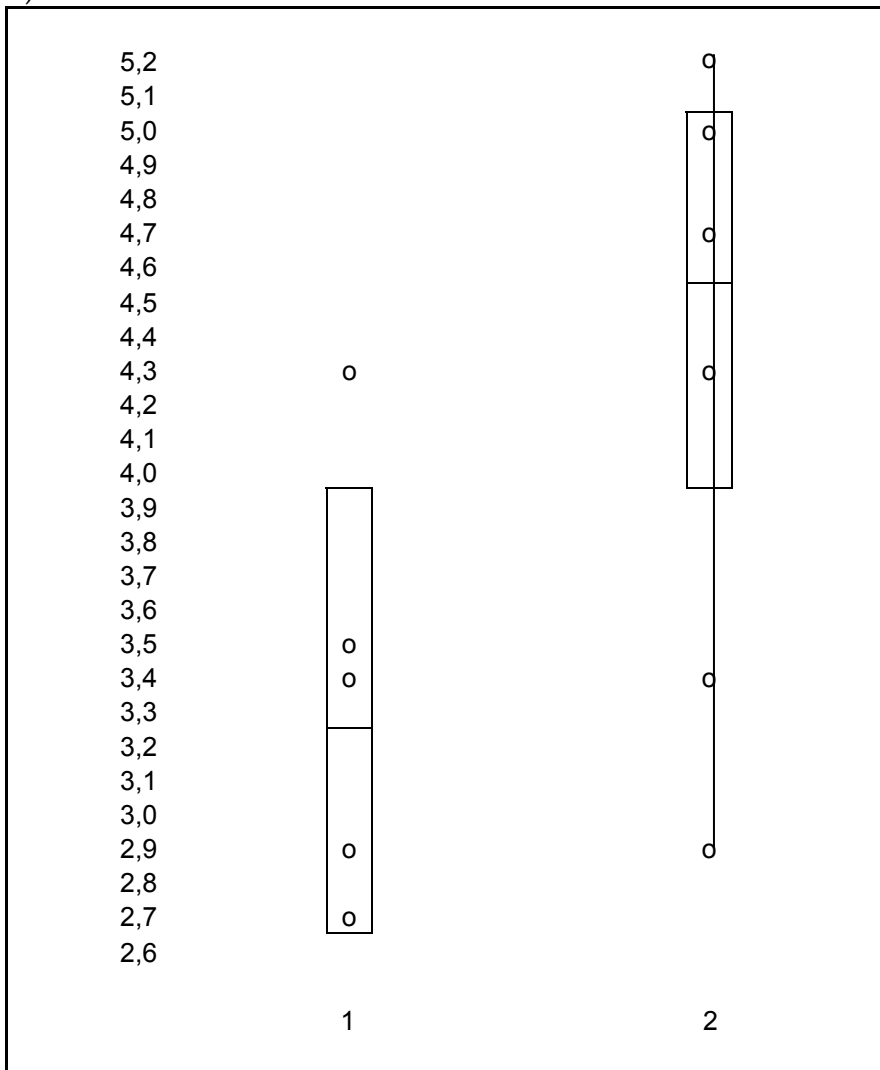
$t = - 1,673$ ;  $dl = 8$ ; n.s.

L'endettement des étudiants vivant chez leurs parents est plus faible que celui des étudiants vivant à l'extérieur. La différence entre les deux groupes s'établit à 4140\$ dans notre échantillon. Le test  $t$  nous indique que cette différence n'est pas significativement différente de zéro.



### Question 3

a)

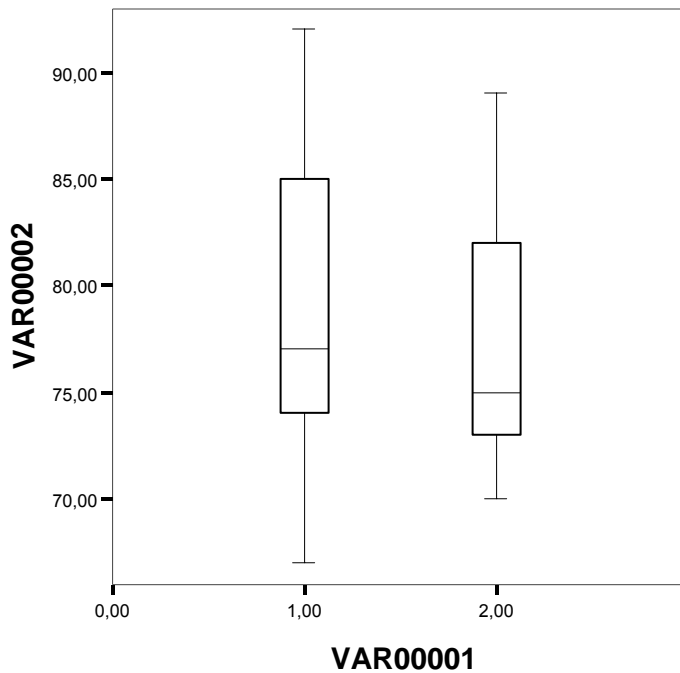


b) Moyennes : 3,34 et 4,62  
Différence entre les moyennes : - 1,28  
Variances : 0,374 et 0,255  
 $t = - 3,23; dl=8; p < 0,02$

c) Dans cet échantillon, les étudiants qui assistent à au moins 75% des cours ont, en moyenne, des résultats plus faibles que ceux qui assistent moins souvent aux cours. Cette différence dans les résultats scolaires est significative au seuil 0,05.

#### Question 4

a)



b) Moyennes : 79 et 77,8

Différence entre les moyennes = 1,2

$t=0,194$

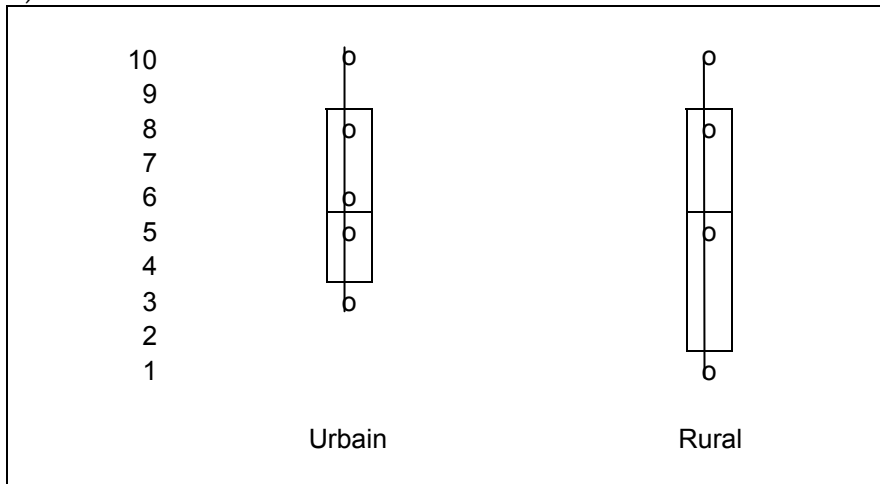
$dl=8$

$p=n.s.$

c) Le professeur accorde des notes légèrement plus élevée en moyenne que l'auxiliaire. Mais la différence (1,2 points) n'est pas significativement différente de zéro.

### Question 5

a)



b) Moyennes : 6,4 et 5,8.

Différence entre les moyennes : 0,6

Variances : 7,3 et 11,7

$t = 0,276$ ;  $dl = 8$ ; n.s.

c) Les personnes de l'échantillon vivant en ville disent accorder plus d'importance à Dieu que les personnes vivant en campagne. Mais la différence entre les deux groupes (0,6 points) n'est pas significativement différente de zéro.

### *Analyse de données agrégées*

#### Question 1

a)  $H_0: \mu_1 - \mu_2 = 0$ , i.e., il n'y a pas de relation entre l'idéologie du gouvernement et le niveau des dépenses en éducation.

b) Groupe 1 : 4,47                      Groupe 2 : 5,46

c) Différence entre les moyennes = - 0,99

d) Le test de Levene n'est pas significatif ( $F = 3,868$  ;  $p > 0,05$ ) donc nous assumons l'égalité des variances dans les groupes ce qui veut dire que nous devons lire la ligne supérieure du test  $t$ .

e) Selon les résultats obtenus pour le  $t$  de Student ( $t = - 2,450$  ;  $dl = 86$  ;  $p = 0,016$ ), on devrait rejeter  $H_0$  selon laquelle il n'y a pas de différence entre les moyennes. On peut donc généraliser les résultats de l'échantillon à la population.

## Question 2

- a)  $H_0: \mu_1 - \mu_2 = 0$ , i.e., il n'y a pas de différence entre les budgets alloués à la défense par les gouvernements de gauche et par ceux de droite
- b) Groupe 1 = 2,98 ; Groupe 2 = 3,09
- c) Différence entre les moyennes : - 0,11
- d) Les variances sont égales. Il faut lire la première ligne.
- e) Le test  $t$  indique que la différence entre les moyennes n'est pas significativement différente de zéro. Il faut donc supposer que les moyennes dans la population sont égales.
- f) Les gouvernements de gauche ont le même comportement que les gouvernements de droite pour ce qui est de la défense. Les comportements diffèrent cependant quand il s'agit d'éducation. Ces résultats confirment l'idée que les différences idéologiques concernent les dépenses d'État providence (dont l'éducation fait partie).

# Chapitre 10

## L'analyse de variance

### *Exercices papier crayon*

#### Question 1

a) Tableau complété :

| Diplôme obtenu | Nombre de journaux lus | $X_i - \mu_t$ | $(X_i - \mu_t)^2$ | $X_i - \mu_g$ | $(X_i - \mu_g)^2$ |
|----------------|------------------------|---------------|-------------------|---------------|-------------------|
| Collégial      | 2                      | -2,071        | 4,289             | -3,333        | 11,109            |
| Collégial      | 7                      | 2,929         | 8,579             | 1,667         | 2,779             |
| Collégial      | 7                      | 2,929         | 8,579             | 1,667         | 2,779             |
| Prof.          | 2                      | -2,071        | 4,289             | 0,667         | 0,445             |
| Prof.          | 2                      | -2,071        | 4,289             | 0,667         | 0,445             |
| Prof.          | 0                      | -4,071        | 16,573            | -1,333        | 1,777             |
| Secondaire     | 3                      | -1,071        | 1,147             | -0,5          | 0,25              |
| Secondaire     | 3                      | -1,071        | 1,147             | -0,5          | 0,25              |
| Secondaire     | 1                      | -3,071        | 9,431             | -2,5          | 6,25              |
| Secondaire     | 7                      | 2,929         | 8,579             | 3,5           | 12,25             |
| Universitaire  | 7                      | 2,929         | 8,579             | 1,25          | 1,563             |
| Universitaire  | 2                      | -2,071        | 4,289             | -3,75         | 14,062            |
| Universitaire  | 7                      | 2,929         | 8,579             | 1,25          | 1,563             |
| Universitaire  | 7                      | 2,929         | 8,579             | 1,25          | 1,563             |
| <b>Totaux</b>  |                        |               | <b>96,928</b>     |               | <b>57,085</b>     |

b)  $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ , i.e., il n'y a pas de relation entre le niveau de scolarité et l'assiduité à la lecture des quotidiens.

c) Moyenne totale : 4,071

Moyenne secondaire : 3,5

Moyenne école professionnelle : 1,333

Moyenne collégial : 5,333

Moyenne universitaire : 5,75

d) Somme totale des carrés : 96,298

Somme des carrés intra-groupe : 57,085

Somme des carrés intergroupes : 39,857

e) OK, à quelques centièmes près en raison des arrondissements.

f)

Intergroupes =  $k - 1 = 3$

Intragroupe =  $N - k = 10$

g) Somme moyenne des carrés intergroupes : 13,286  
Somme moyenne des carrés intra-groupe : 5,708

h) 2,328

i) Pour 3 et 10 degrés de liberté, le tableau de la distribution du  $F$  donne un  $F$  critique de 3,71. Si on la calculait, on trouverait que la probabilité associée à un  $F$  de 2,328 pour le même nombre de degrés de liberté serait supérieure à 0,05 (SPSS donne une probabilité de 0,136). Cela signifie que la probabilité de tirer un échantillon comme celui que nous avons (i.e. ayant un  $F$  de 2,328) à partir d'une population où le  $F$  est égal à 0 (selon l'hypothèse nulle) est supérieure à 0,05 (ou égale à 0,136 nous dit SPSS).

j) Nous ne pouvons pas rejeter l'hypothèse nulle car la probabilité d'obtenir un  $F$  égal ou supérieure à 2,328, si  $H_0$  est vraie, est supérieure à 0,05.

j)

Analyse de la variance de la relation entre le niveau de scolarité  
et l'intérêt pour la lecture d'un quotidien

| Origine      | Somme des carrés | dl | Somme moyenne des carrés | F     | p    |
|--------------|------------------|----|--------------------------|-------|------|
| Intergroupes | 39,857           | 3  | 13,286                   | 2,328 | n.s. |
| Intra-groupe | 57,085           | 10 | 5,708                    |       |      |
| Total        | 96,942           | 13 |                          |       |      |

Ce tableau nous indique que la relation entre le niveau de scolarité et l'intérêt pour la lecture d'un quotidien n'est pas significative ( $F = 2,328$ ;  $dl = 3$  et  $13$ ).

### *Analyse de données individuelles*

#### **Question 1**

- a) 18 à 24: 1,36  
25 à 34: 1,49  
35 à 44: 1,81  
45 à 54: 2,19  
55 à 64: 2,53  
65 et plus: 2,66

b)

| Source de variation | Somme des carrés | dl  | Somme moyenne des carrés | F     |
|---------------------|------------------|-----|--------------------------|-------|
| Intergruppes        | 145,73           | 5   | 29,25                    | 26,52 |
| Intra-groupe        | 994,56           | 905 | 1,1                      |       |
| Total               | 1140,29          | 910 |                          |       |

$\text{Eta}^2 : 0,13$

c)

La valeur de la statistique  $F$  pour 5 et 905 degrés de liberté nous indique que la relation entre l'âge et la taille de la famille est significative au seuil de 0,001. Ainsi, nous pouvons dire qu'il y a une relation significative entre l'âge et le nombre d'enfants. Connaissant le groupe d'âge des répondants, nous pouvons réduire nos erreurs de prédiction du nombre d'enfants de 13%. L'âge explique 13% de la variance dans le nombre d'enfants.

## Question 2

a) Moyenne du nombre d'enfants idéal pour chaque groupe d'âge

18-24 ans : 2,5

25-34 ans : 2,41

35-44 ans : 2,32

45-54 ans : 2,38

55-64 ans : 2,6

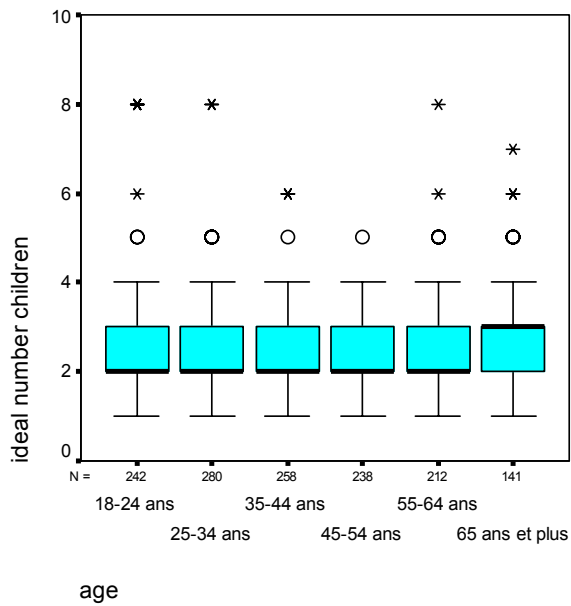
65 ans et plus : 2,75

Moyenne du nombre d'enfants idéal selon les groupes d'âge pour les Italiens : Tableau ANOVA

| Source de variation | Somme des carrés | dl   | Somme moyenne des carrés | F    |
|---------------------|------------------|------|--------------------------|------|
| Intergruppes        | 23,67            | 5    | 4,73                     | 6,02 |
| Intra-groupe        | 1073,76          | 1365 | ,787                     |      |
| Total               | 1097,43          | 1370 |                          |      |

$E^2 = 0,02$

La relation entre l'âge et le nombre idéal d'enfants est significative au seuil 0,001 ( $F = 6,02$  ;  $dl = 5$  et  $1365$ ). Connaissant le groupe d'âge des répondants, nous pouvons réduire nos erreurs de prédiction du nombre idéal d'enfants de 2%. L'âge explique 2% de la variance dans le nombre idéal d'enfants.



Description du diagramme en boîtes :

Le nombre idéal d'enfants ne semble pas varier d'un groupe d'âge à l'autre, sauf en ce qui concerne les 65 ans et plus. Partout ailleurs la médiane est à 2 enfants (elle est à 3 enfants chez les plus âgés). Le nombre idéal d'enfants va jusqu'à 8 dans trois groupes d'âge. À la lumière de ce diagramme, la relation entre les deux variables est nulle.

### Analyse de données agrégées

#### Question 1

| Variable dépendante      | <u>Moyenne par région économique</u> |                 |                |              |              | Amérique Latine |
|--------------------------|--------------------------------------|-----------------|----------------|--------------|--------------|-----------------|
|                          | OCDE                                 | Europe de l'Est | Asie Pacifique | Afrique      | Moyen Orient |                 |
| Espérance de vie moyenne | <u>77,48</u>                         | <u>70,50</u>    | <u>67,65</u>   | <u>49,25</u> | <u>70,88</u> | <u>70,74</u>    |

#### Analyse de Variance de la relation entre l'espérance de vie et la région économique

| Origine      | Somme des carrés | dl  | Somme moyenne des carrés | F     | P     |
|--------------|------------------|-----|--------------------------|-------|-------|
| Intergroupes | 15078,20         | 5   | 3015,64                  | 96,43 | 0,001 |
| Intra-groupe | 3940,32          | 126 | 31,27                    |       |       |
| Total        | 19018,52         | 131 |                          |       |       |



$\hat{\epsilon}^2$  : 0,79

Conclusion générale :

Le tableau qui présente les moyennes d'espérance de vie par région économique montre qu'il existe une variation importante entre les différentes régions. En effet, les pays de l'Afrique ont une moyenne de 49,25 alors que les pays de l'OCDE ont une moyenne de 77,48. La valeur de la statistique  $F$  (96,43 pour 5 et 126 degrés de liberté) nous indique que cette relation est statistiquement significative au seuil de 1 pour mille. De plus, l' $\hat{\epsilon}^2$  à 0,79 nous indique que le fait de connaître la région économique d'un pays nous permet de réduire de 79 % nos erreurs de prédiction sur l'espérance de vie. La région économique explique donc 79 % de la variance dans le niveau d'espérance de vie d'un pays.

## Question 2

Taux de fécondité selon la religion prédominante

|                |      |
|----------------|------|
| 1 Anglicane    | 1,7  |
| 2 Animiste     | 5,48 |
| 3 Baptiste     | 2,6  |
| 4 Bouddhiste   | 2,86 |
| 5 Catholique   | 2,79 |
| 6 Hindou       | 2,95 |
| 7 Juive        | 2,7  |
| 8 Musulmane    | 4,27 |
| 9 Orthodoxe    | 1,6  |
| 10 Protestante | 2,22 |
| 11 Taoïste     | 1,75 |

Taux de fécondité selon la religion prédominante : ANOVA Table

| Source de variation | Somme des carrés | dl  | Somme moyenne des carrés | F    | p    |
|---------------------|------------------|-----|--------------------------|------|------|
| Intergroupes        | 179,72           | 10  | 17,97                    | 8,57 | ,000 |
| Intra-groupe        | 314,49           | 150 | 2,1                      |      |      |
| Total               | 494,2            | 160 |                          |      |      |

$E^2 = 0,36$

Le degré de signification de 0,000 nous permet de rejeter l'hypothèse nulle. La relation est significative et son intensité est modérée.

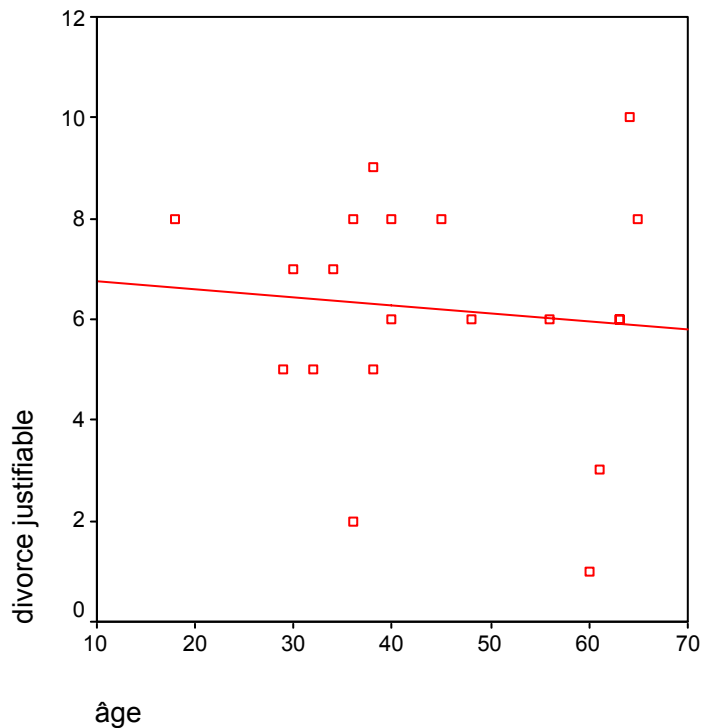
# Chapitre 11

## La régression et la corrélation

### Exercices papier-crayon

#### Question 1

| Répon-<br>dants | $X_i$ | $X_i - \bar{X}$ | $(X_i - \bar{X})^2$ | $Z_x$   | $Y_i$ | $Y_i - \bar{Y}$ | $(Y_i - \bar{Y})^2$ | $Z_y$   | $(X_i - \bar{X})^*$<br>$(Y_i - \bar{Y})$ | $Z_x Z_y$ |
|-----------------|-------|-----------------|---------------------|---------|-------|-----------------|---------------------|---------|--|-----------|
| 1               | 40    | -4,8            | 23,04               | -0,3381 | 8     | 1,8             | 3,24                | 0,7878  | -8,64                                    | -0,2664   |
| 2               | 36    | -8,8            | 77,44               | -0,6199 | 2     | -4,2            | 17,64               | -1,8381 | 36,96                                    | 1,1394    |
| 3               | 56    | 11,2            | 125,44              | 0,78893 | 6     | -0,2            | 0,04                | -0,0875 | -2,24                                    | -0,0691   |
| 4               | 60    | 15,2            | 231,04              | 1,0707  | 1     | -5,2            | 27,04               | -2,2757 | -79,04                                   | -2,4366   |
| 5               | 18    | -26,8           | 718,24              | -1,8878 | 8     | 1,8             | 3,24                | 0,7878  | -48,24                                   | -1,4871   |
| 6               | 63    | 18,2            | 331,24              | 1,28202 | 6     | -0,2            | 0,04                | -0,0875 | -3,64                                    | -0,1122   |
| 7               | 64    | 19,2            | 368,64              | 1,35246 | 10    | 3,8             | 14,44               | 1,663   | 72,96                                    | 2,2492    |
| 8               | 48    | 3,2             | 10,24               | 0,22541 | 6     | -0,2            | 0,04                | -0,0875 | -0,64                                    | -0,0197   |
| 9               | 32    | -12,8           | 163,84              | -0,9016 | 5     | -1,2            | 1,44                | -0,5252 | 15,36                                    | 0,4735    |
| 10              | 34    | -10,8           | 116,64              | -0,7608 | 7     | 0,8             | 0,64                | 0,3501  | -8,64                                    | -0,2664   |
| 11              | 40    | -4,8            | 23,04               | -0,3381 | 6     | -0,2            | 0,04                | -0,0875 | 0,96                                     | 0,0296    |
| 12              | 29    | -15,8           | 249,64              | -1,113  | 5     | -1,2            | 1,44                | -0,5252 | 18,96                                    | 0,5845    |
| 13              | 45    | 0,2             | 0,04                | 0,01409 | 8     | 1,8             | 3,24                | 0,7878  | 0,36                                     | 0,0111    |
| 14              | 38    | -6,8            | 46,24               | -0,479  | 5     | -1,2            | 1,44                | -0,5252 | 8,16                                     | 0,2516    |
| 15              | 36    | -8,8            | 77,44               | -0,6199 | 8     | 1,8             | 3,24                | 0,7878  | -15,84                                   | -0,4883   |
| 16              | 63    | 18,2            | 331,24              | 1,28202 | 6     | -0,2            | 0,04                | -0,0875 | -3,64                                    | -0,1122   |
| 17              | 65    | 20,2            | 408,04              | 1,4229  | 8     | 1,8             | 3,24                | 0,7878  | 36,36                                    | 1,1209    |
| 18              | 30    | -14,8           | 219,04              | -1,0425 | 7     | 0,8             | 0,64                | 0,3501  | -11,84                                   | -0,365    |
| 19              | 38    | -6,8            | 46,24               | -0,479  | 9     | 2,8             | 7,84                | 1,2254  | -19,04                                   | -0,587    |
| 20              | 61    | 16,2            | 262,44              | 1,14114 | 3     | -3,2            | 10,24               | -1,4005 | -51,84                                   | -1,5981   |
| Totaux          |       | 0               | 3829,2              | 0       |       | 0               | 99,2                | 0       | -63,2                                    | -1,9483   |



b) Âge : Moyenne : 44,8 Écart-type : 14,20  
 Divorce Moyenne : 6,2 Écart-type : 2,28

c) Droite de régression :  $Y = 6,9394 - 0,0165 X$

d) Voir le diagramme en a).

e)  $r = -0,10$

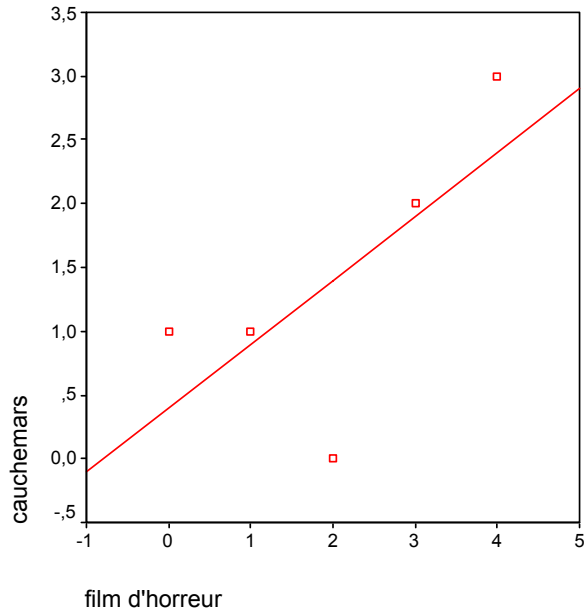
f)  $F = 0,182$

Avec 1 et 18 degrés de liberté, le  $F$  critique est de 4,41 à un seuil de 0,05. Le score  $F$  de notre  $r$  de Pearson est largement inférieur à ce score. Nous ne pouvons donc pas rejeter l'hypothèse nulle à l'effet que  $r = 0$ .

g) Il n'y a pas de relation entre l'âge et l'opinion concernant l'acceptabilité du divorce. La corrélation entre les deux variables est presque nulle et non significative ( $r = 0,10$  ; n.s.).

## Question 2

a) Nombre de cauchemars en fonction du nombre de films d'horreur regardés



b)

| Enfant    | X | X - u <sub>x</sub> | (X-u <sub>x</sub> ) <sup>2</sup> | Z <sub>x</sub> | Y | Y - u <sub>y</sub> | (Y-u <sub>y</sub> ) <sup>2</sup> | Z <sub>y</sub> | (X-u <sub>x</sub> )*(Y-u <sub>y</sub> ) | Z <sub>x</sub> Z <sub>y</sub> |
|-----------|---|--------------------|----------------------------------|----------------|---|--------------------|----------------------------------|----------------|---|-------------------------------|
| Noémie    | 2 | 0                  | 0                                | 0              | 0 | -1,4               | 1,96                             | -1,228         | 0                                       | 0                             |
| Pascal    | 4 | 2                  | 4                                | 1,266          | 3 | 1,6                | 2,56                             | 1,404          | 3,2                                     | 1,78                          |
| Olivier   | 1 | -1                 | 1                                | -0,633         | 1 | -0,4               | 0,16                             | -0,351         | 0,4                                     | 0,22                          |
| Alexandra | 0 | -2                 | 4                                | -1,266         | 1 | -0,4               | 0,16                             | -0,351         | 0,8                                     | 0,44                          |
| Marie-Ève | 3 | 1                  | 1                                | 0,633          | 2 | 0,6                | 0,36                             | 0,526          | 0,6                                     | 0,33                          |
| Total     |   |                    | 10                               | 0              |   |                    | 5,2                              | 0              | 5                                       | 2,78                          |

Films d'horreur : Moyenne = 2 Écart-type = 1,58

Cauchemars : Moyenne = 1,4 Écart-type = 1,14

c)  $Y = 0,4 + 0,5X$

d) Voir graphique en a).

e)  $r = 0,70$

f)  $F = 2,882$ .

Avec 1 et 3 degrés de liberté, le  $F$  critique est de 10,13 à un seuil de 0,05. Le score  $F$  de notre  $r$  de Pearson est largement inférieur à ce score. Nous ne pouvons donc pas rejeter l'hypothèse nulle à l'effet que  $r = 0$ .

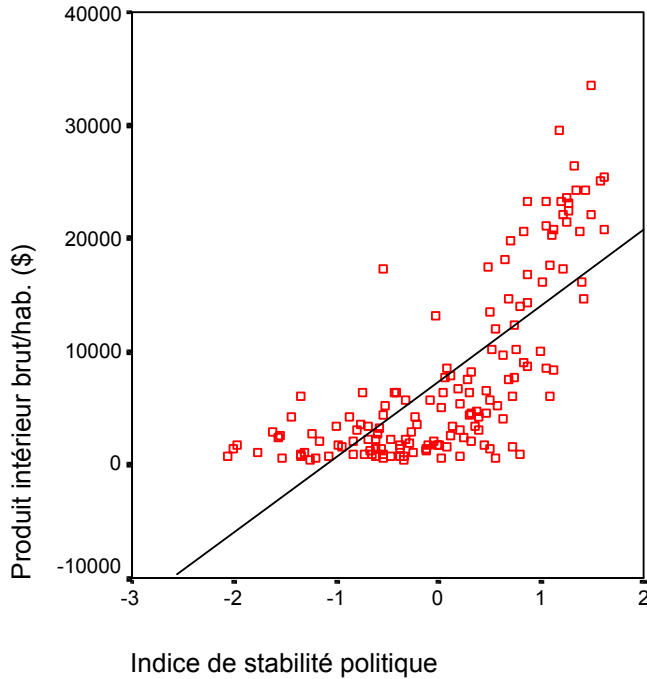
g) Il n'y a pas de relation entre le nombre de films d'horreur regardé et le nombre de cauchemars. Nous avons trouvé, dans l'échantillon, une relation positive et relativement forte entre les films d'horreurs et les cauchemars. Mais cette relation

n'est pas significative au seuil 0,05. Même si la relation est relativement forte, elle ne peut être généralisée à la population entière.

### Analyse de données agrégées (Pays2000)

#### Question 1

a)



Il s'agit d'une relation positive forte.

b)  $Y = 7379,49 + 6666,66X$

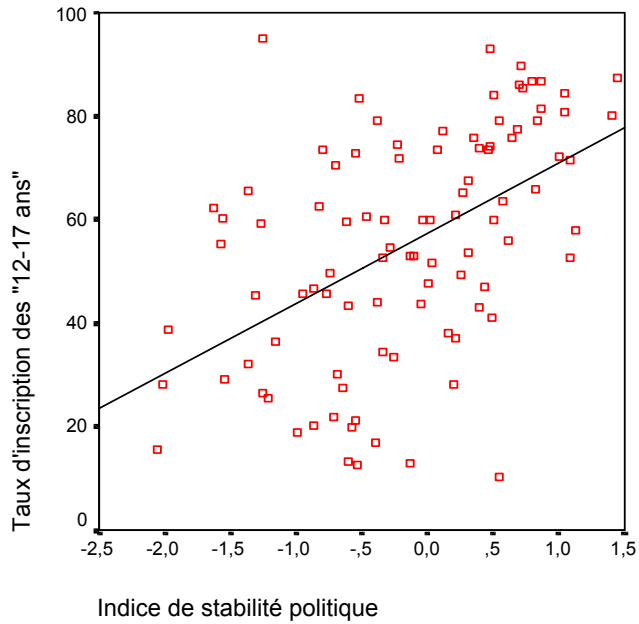
c) 0,74

d) 0,55 donc on peut expliquer 55% de la variation de la variable dépendante

e) L'hypothèse de départ est confirmée. Il y a une corrélation positive significative entre l'indice de stabilité politique et le PIB per capita ( $r = 0,74$  ;  $p < 0,001$ ). Les pays qui jouissent d'une plus grande stabilité politique ont un PIB per capita plus élevé.

## Question 2

a)



C'est une relation linéaire positive modérée.

b)  $Y = 57,39 + 13,60X$

c) 0,51

d) 0,25

e) Oui, plus un pays est politiquement stable, plus le taux de scolarisation est élevé. Le coefficient de corrélation de Pearson ( $r=0,51$ ,  $p < 0,001$ ) indique qu'il existe une relation positive significative entre les deux variables et qu'elle est modérée.

# Chapitre 12

## L'analyse tabulaire multivariée

### *Exercices papier-crayon*

#### **Question 1**

a) Même si elle est significative au seuil 0,05 la relation entre l'opinion concernant le rôle de l'homme et de la femme et le revenu est si faible ( $\Gamma = 0,11$  ; différence de pourcentage = 8,3 points) qu'elle doit être considérée comme nulle. Si un si petit  $\Gamma$  est significatif, c'est à cause de la grande taille de l'échantillon.

b) Variable indépendante : Revenu familial

c) Variable dépendante : Opinion à propos de l'affirmation selon laquelle l'homme et la femme doivent contribuer à faire vivre la famille (Égalité des sexes quant à la contribution au revenu familial).

d) Variable de contrôle : la scolarité (mesurée par l'âge à la fin des études).

e) Reproduction. Aucun des  $\Gamma$ s des tableaux partiels n'est significatif. La relation est donc nulle, quel que soit le niveau de scolarité. Il n'y a pas plus de relation dans les tableaux partiels que dans le tableau bi-varié. Les résultats des tableaux partiels renforcent notre décision de considérer comme nulle la relation entre le revenu familial et l'opinion concernant le rôle des femmes.

f) Il y a trois explications possibles : 1- Il n'y a pas de différence d'opinion entre les personnes qui ont un revenu élevé et les autres concernant le rôle de l'homme et de la femme en ce qui concerne la contribution au revenu familial. C'est pourquoi nous n'avons pas trouvé de relation entre les deux variables, même en contrôlant l'effet du niveau d'instruction. 2- Il y a une relation mais elle est dissimulée par une variable antécédente inconnue et cette variable n'est pas l'âge à la fin des études. 3- Notre échantillon n'est pas représentatif ou nous n'avons pas des mesures valides.

#### **Question 2**

a) La relation est significative au seuil  $p < 0,001$ . Son intensité, révélée par le  $\Gamma$  de 0,55, est modérée. De plus, une analyse du tableau nous permet de constater que la relation est positive : les répondants croyants ont tendance à être plus attachés au Canada que les non-croyants. Donc, la relation entre le niveau de croyance religieuse et l'attachement au Canada est positive et modérée, et peut être généralisée à l'ensemble de la population.

b) Variable indépendante : Niveau de croyance religieuse



- c) Variable dépendante : Attachement au Canada
- d) Variable de contrôle : Scolarité (mesurée par le dernier diplôme reçu)
- e) Reproduction. Le gamma dans les trois catégories partielles est sensiblement identique à celui de la relation primitive. Toutes les relations partielles sont modérées, positives et statistiquement significatives.
- f) Le niveau de croyance peut être interprété comme une mesure de conservatisme : toutes choses égales par ailleurs, les personnes qui se disent croyantes sont plus attachées aux valeurs traditionnelles. Il est fait l'hypothèse que ces personnes auront tendance à être plus attachée au Canada, car le pays est le lieu d'attachement traditionnel (par opposition à la localité ou au monde). L'analyse bi-variée montre que la relation entre les deux variables est modérée et significative.

Le niveau de scolarité est une variable antécédente puisqu'un répondant fait ses études avant de répondre à l'une ou l'autre question. En théorie, il est permis de penser que les répondants les plus scolarisés, parce qu'ils ont développé des habiletés intellectuelles plus sophistiquées, feront la différence entre leur croyance religieuse et leur opinion politique, soit parce qu'ils considèrent que leur religion dépasse les frontières, soit parce qu'ils croient que leur rapport à un être suprême est hautement personnel et n'a donc pas de rapport avec la politique. La relation devrait varier d'un niveau de scolarité à l'autre (spécification).

L'analyse multivariée révèle que le niveau de scolarité n'a pas d'impact sur la relation entre le niveau de croyance et l'attachement au Canada.

### *Analyse de données individuelles (WVS)*

#### **Question 1**

- a) Plus le revenu est élevé, meilleur est l'état de santé. Justification : les gens plus fortunés ont une meilleure alimentation, plus de loisirs, font plus d'exercice et ont une meilleure hygiène, tous ces facteurs contribuant à un meilleur état de santé.
- b) Oui, avec un  $X^2$  de 38,40 et 4 degrés de liberté, la relation est statistiquement significative au seuil de 0,001. Elle est positive, ce que prouve l'analyse du tableau ainsi que les mesures d'association. Les différences des pourcentages de même que les résultats pour les mesures d'association montrent que l'intensité de la relation est faible si on retient le D de Somers (puisque nous connaissons la variables dépendante) .
- c)  $D_{yx} = 0,12$ .

- d) Bien que l'intensité de la relation entre les deux variables soit faible, notre hypothèse de départ a été confirmée. Les personnes ayant un revenu plus élevé jouissent d'une meilleure santé.

Tableau 12,1: État de santé selon le revenu  
(pourcentages)

| Santé    | Revenu |       |       |
|----------|--------|-------|-------|
|          | Faible | Moyen | Élevé |
| Bonne    | 73,9   | 87,2  | 90,3  |
| Moyenne  | 22,6   | 11,2  | 8,0   |
| Mauvaise | 3,5    | 1,6   | 1,7   |
| Total    | 100,0  | 100,0 | 100,0 |
| (N)      | (368)  | (250) | (349) |

$$D_{yx} = 0,12 ; p < 0,001$$

$$\chi^2 = 38,394 ; dl = 4 ; p < 0,001.$$

- e) VD = niveau de santé;  
VI = revenu ;  
Variable de contrôle = niveau d'instruction ; c'est une variable antécédente.
- f) Voici les résultats de l'analyse multivariée :  
relation primitive VI-VD :  $D_{yx}$  de Somers = 0,12 ( $p < 0,001$ ) ;  
chez les «Niveau d'instruction faible»,  $D_{yx} = 0,19$  ( $p < 0,001$ ) ;  
chez les «Niveau d'instruction moyen»,  $D_{yx} = 0,07$  ( $p < 0,01$ ) ;  
chez les «Niveau d'instruction élevé»,  $D_{yx} = 0,06$  (n.s.).  
Nous avons ici un cas de spécification car la relation VI-VD s'intensifie dans une catégorie de la variable de contrôle («Niveau d'instruction faible») et disparaît dans les autres catégories.
- g) La règle générale voulant que le niveau de revenu soit relié à l'état de santé se vérifie chez les moins instruits mais non chez les personnes à niveau d'instruction moyen ou élevé. Il est possible que les habitudes de vie menant à une meilleure santé soient plus répandues chez les personnes plus instruites. Cela expliquerait pourquoi l'hypothèse est confirmée chez les moins instruits seulement.

Tableau 12.2: État de santé selon le revenu contrôlant l'effet du niveau d'instruction (pourcentages)

| Santé    | Niveau d'instruction          |       |       |                              |       |       |                        |       |       |
|----------|-------------------------------|-------|-------|------------------------------|-------|-------|------------------------|-------|-------|
|          | Faible                        |       |       | Moyen                        |       |       | Élevé                  |       |       |
|          | Revenu                        |       |       | Revenu                       |       |       | Revenu                 |       |       |
|          | Faible                        | Moyen | Élevé | Faible                       | Moyen | Élevé | Faible                 | Moyen | Élevé |
| Bonne    | 62,7                          | 75,0  | 90,9  | 80,3                         | 88,3  | 90,2  | 77,1                   | 93,6  | 90,1  |
| Moyenne  | 33,3                          | 20,0  | 9,1   | 16,1                         | 11,1  | 8,1   | 20,8                   | 4,3   | 7,7   |
| Mauvaise | 4,0                           | 5,0   | 0,0   | 3,6                          | 0,6   | 1,7   | 2,1                    | 2,1   | 2,1   |
| Total    | 100,0                         | 100,0 | 100,0 | 100,0                        | 100,0 | 100,0 | 100,0                  | 100,0 | 100,0 |
| (N)      | (126)                         | (40)  | (33)  | (193)                        | (162) | (173) | (48)                   | (47)  | (142) |
|          | $D_{yx} = 0,19$ ; $p < 0,001$ |       |       | $D_{yx} = 0,07$ ; $p < 0,01$ |       |       | $D_{yx} = 0,06$ ; n.s. |       |       |
|          | $X^2 = 11,251$ ; $p < 0,05$   |       |       | $X^2 = 10,142$ ; $p < 0,05$  |       |       | $X^2 = 9,015$ ; n.s.   |       |       |

## Question 2

- a) La volonté de s'engager dans l'armée est moindre au Québec du fait de facteurs politiques et historiques spécifiques notamment, le caractère francophone de la province, le mouvement de sécession, la crise de conscription.
- b) Nous avons fusionné toutes les réponses provenant des 9 provinces à majorité anglophone en une seule catégorie «Reste du Canada».
- c) La relation est statistiquement significative au seuil 0,001 car le khi-carré de 121,383 est supérieur à la valeur critique donnée par le tableau de la distribution du khi-carré pour ce niveau de signification et 1 degré de liberté (10,827).
- d) Le V de Cramer est de 0,31 signe d'une relation relativement forte.
- e) L'hypothèse de départ est confirmée. Les Québécois affichent une volonté significativement plus faible de s'enrôler dans l'armée en cas de guerre que les Canadiens des autres provinces.

## Tableaux croisés

### Récapitulatif du traitement des observations

|  | Observations |          |           |          |       |          |
|--|--------------|----------|-----------|----------|-------|----------|
|  | Valide       |          | Manquante |          | Total |          |
|  | N            | Pourcent | N         | Pourcent | N     | Pourcent |
| V110 willing to fight in war * V234 region | 1238         | 83,1%    | 252       | 16,9%    | 1490  | 100,0%   |

### Tableau croisé V110 willing to fight in war \* V234 region

|                              |       |                    | V234 region       |          | Total  |
|------------------------------|-------|--------------------|-------------------|----------|--------|
|                              |       |                    | 1 Reste du Canada | 2 Québec |        |
| V110 willing to fight in war | 1 yes | Effectif           | 698               | 136      | 834    |
|                              |       | % dans V234 region | 76,0%             | 42,5%    | 67,4%  |
|                              | 2 no  | Effectif           | 220               | 184      | 404    |
|                              |       | % dans V234 region | 24,0%             | 57,5%    | 32,6%  |
| Total                        |       | Effectif           | 918               | 320      | 1238   |
|                              |       | % dans V234 region | 100,0%            | 100,0%   | 100,0% |

### Tests du Khi-deux

|                                   | Valeur               | ddl | Signification asymptotique (bilatérale) | Signification exacte (bilatérale) | Signification exacte (unilatérale) |
|-----------------------------------|----------------------|-----|---|-----------------------------------|------------------------------------|
| Khi-deux de Pearson               | 121,383 <sup>b</sup> | 1   | ,000                                    |                                   |                                    |
| Correction pour la continuité     | 119,863              | 1   | ,000                                    |                                   |                                    |
| Rapport de vraisemblance          | 116,289              | 1   | ,000                                    |                                   |                                    |
| Test exact de Fisher              |                      |     |   | ,000                              | ,000                               |
| Association linéaire par linéaire | 121,285              | 1   | ,000                                    |                                   |                                    |
| Nombre d'observations valides     | 1238                 |     |   |                                   |                                    |

a. Calculé uniquement pour un tableau 2x2

b. 0 cellules (,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 104,43.

### Mesures directionnelles

|                           |        |  | Valeur | Erreur standard asymptotique <sup>a</sup> | T approximé <sup>b</sup> | Signification approximée |
|---------------------------|--------|--|--------|---|--------------------------|--------------------------|
| Nominal par Nominal       | Lambda | Symétrique                             | ,066   | ,024                                      | 2,691                    | ,007                     |
|                           |        | V110 willing to fight in war dépendant | ,119   | ,042                                      | 2,691                    | ,007                     |
|                           |        | V234 region dépendant                  | ,000   | ,000                                      | . <sup>c</sup>           | . <sup>c</sup>           |
| Tau de Goodman et Kruskal |        | V110 willing to fight in war dépendant | ,098   | ,018                                      |                          | ,000 <sup>d</sup>        |
|                           |        | V234 region dépendant                  | ,098   | ,018                                      |                          | ,000 <sup>d</sup>        |

a. L'hypothèse nulle n'est pas considérée.

b. Utilisation de l'erreur standard asymptotique dans l'hypothèse nulle.

c. Calcul impossible /: l'erreur standard asymptotique est égale à zéro.

d. Basé sur une approximation du Khi-deux

### Mesures symétriques

|                               |  |                            | Valeur | Erreur standard asymptotique <sup>a</sup> | T approximé <sup>b</sup> | Signification approximée |
|-------------------------------|--|----------------------------|--------|---|--------------------------|--------------------------|
| Nominal par Nominal           |  | Phi                        | ,313   |   |                          | ,000                     |
|                               |  | V de Cramer                | ,313   |   |                          | ,000                     |
|                               |  | Coefficient de contingence | ,299   |   |                          | ,000                     |
| Ordinal par Ordinal           |  | Gamma                      | ,622   | ,042                                      | 10,239                   | ,000                     |
| Nombre d'observations valides |  |                            | 1238   |   |                          |                          |

a. L'hypothèse nulle n'est pas considérée.

b. Utilisation de l'erreur standard asymptotique dans l'hypothèse nulle.

- e) VD : La tendance à l'enrôlement en cas de guerre.  
VI : La province habitée.

VC : La fierté d'être Canadien. En postulant que la province habitée est la province d'origine, on peut considérer la variable «Fierté d'être Canadien» comme une variable intermédiaire.

f) Parmi les gens qui se disent fiers d'être canadiens, 46,6% des Québécois se disent prêts à s'enrôler dans l'armée en cas de guerre contre 77% des répondants du reste du Canada. Cette différence est significative au seuil 0,001. La relation entre la province habitée et la volonté de s'enrôler est modérée (V de Kramer = 0,28 ;  $\chi^2 = 88,816$  ;  $p < 0,001$ ).

g) Parmi les répondants qui répondent «pas beaucoup» ou «pas du tout » à la question sur le sentiment de fierté d'être Canadien, les résultats mènent à la même conclusion : 24,5% des Québécois contre 59,1% des autres Canadiens se disent prêts à s'enrôler ; la relation entre les deux variables est toujours modérée et significative (V de Kramer = 0,33 ;  $\chi^2 = 8,200$  ;  $p < 0,01$ ).

h) Comme les mesures d'association sont semblables dans les tableaux partiels et dans la relation primitive, il faut conclure que la variable-contrôle ne relie pas la variable indépendante à la variable dépendante. L'impact des facteurs historiques et culturels captés par la variable indépendante («province habitée») ne passe pas par la fierté d'être canadien.

i) La variable «Fierté d'être Canadien» ne relie pas les variables indépendante et dépendante. D'autres variables sont susceptibles de faire cette liaison.

## Tableaux croisés

Récapitulatif du traitement des observations

|   | Observations |          |           |          |       |          |
|---|--------------|----------|-----------|----------|-------|----------|
|   | Valide       |          | Manquante |          | Total |          |
|   | N            | Pourcent | N         | Pourcent | N     | Pourcent |
| V110 willing to fight in war * V234 region * V205 proud to be | 1218         | 81,7%    | 272       | 18,3%    | 1490  | 100,0%   |

**Tableau croisé V110 willing to fight in war \* V234 region \* V205 proud to be**

| V205 proud to be Canadian |                              |          |                    | V234 region       |          | Total  |
|---------------------------|------------------------------|----------|--------------------|-------------------|----------|--------|
|                           |                              |          |                    | 1 Reste du Canada | 2 Québec |        |
| 1 oui                     | V110 willing to fight in war | 1 yes    | Effectif           | 678               | 122      | 800    |
|                           |                              |          | % dans V234 region | 77,0%             | 46,6%    | 70,0%  |
|                           | 2 no                         | Effectif | 203                | 140               | 343      |        |
|                           |                              |          | % dans V234 region | 23,0%             | 53,4%    | 30,0%  |
|                           | Total                        |          | Effectif           | 881               | 262      | 1143   |
|                           |                              |          | % dans V234 region | 100,0%            | 100,0%   | 100,0% |
| 2 non                     | V110 willing to fight in war | 1 yes    | Effectif           | 13                | 13       | 26     |
|                           |                              |          | % dans V234 region | 59,1%             | 24,5%    | 34,7%  |
|                           | 2 no                         | Effectif | 9                  | 40                | 49       |        |
|                           |                              |          | % dans V234 region | 40,9%             | 75,5%    | 65,3%  |
|                           | Total                        |          | Effectif           | 22                | 53       | 75     |
|                           |                              |          | % dans V234 region | 100,0%            | 100,0%   | 100,0% |

**Tests du Khi-deux**

| V205 proud to be Canadian |                                   | Valeur              | ddl | Signification asymptotique (bilatérale) | Signification exacte (bilatérale) | Signification exacte (unilatérale) |
|---------------------------|-----------------------------------|---------------------|-----|---|-----------------------------------|------------------------------------|
| 1 oui                     | Khi-deux de Pearson               | 88,816 <sup>b</sup> | 1   | ,000                                    | ,000                              | ,000                               |
|                           | Correction pour la continuité     | 87,375              | 1   | ,000                                    |                                   |                                    |
|                           | Rapport de vraisemblance          | 83,536              | 1   | ,000                                    |                                   |                                    |
|                           | Test exact de Fisher              |                     |     |   |                                   |                                    |
|                           | Association linéaire par linéaire | 88,738              | 1   | ,000                                    |                                   |                                    |
|                           | Nombre d'observations valides     | 1143                |     |   |                                   |                                    |
| 2 non                     | Khi-deux de Pearson               | 8,200 <sup>c</sup>  | 1   | ,004                                    | ,007                              | ,005                               |
|                           | Correction pour la continuité     | 6,745               | 1   | ,009                                    |                                   |                                    |
|                           | Rapport de vraisemblance          | 7,985               | 1   | ,005                                    |                                   |                                    |
|                           | Test exact de Fisher              |                     |     |   |                                   |                                    |
|                           | Association linéaire par linéaire | 8,090               | 1   | ,004                                    |                                   |                                    |
|                           | Nombre d'observations valides     | 75                  |     |   |                                   |                                    |

a. Calculé uniquement pour un tableau 2x2

b. 0 cellules (,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 78,62.

c. 0 cellules (,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 7,63.

### Mesures directionnelles

| V205 proud to be Canadian |                           |        |  | Valeur | Erreur standard asymptotique <sup>a</sup> | T approximé <sup>b</sup> | Signification approximée |
|---------------------------|---------------------------|--------|--|--------|---|--------------------------|--------------------------|
| 1 oui                     | Nominal par Nominal       | Lambda | Symétrique                             | ,030   | ,026                                      | 1,113                    | ,266                     |
|                           |                           |        | V110 willing to fight ir war dépendant | ,052   | ,046                                      | 1,113                    | ,266                     |
|                           |                           |        | V234 region dépendant                  | ,000   | ,000                                      | . <sup>c</sup>           | . <sup>c</sup>           |
|                           | Tau de Goodmar et Kruskal | Lambda | V110 willing to fight ir war dépendant | ,078   | ,017                                      |                          | ,000 <sup>d</sup>        |
|                           |                           |        | V234 region dépendant                  | ,078   | ,017                                      |                          | ,000 <sup>d</sup>        |
|                           |                           |        |  |        |   |                          |                          |
| 2 non                     | Nominal par Nominal       | Lambda | Symétrique                             | ,083   | ,174                                      | ,466                     | ,641                     |
|                           |                           |        | V110 willing to fight ir war dépendant | ,154   | ,166                                      | ,857                     | ,391                     |
|                           |                           |        | V234 region dépendant                  | ,000   | ,232                                      | ,000                     | 1,000                    |
|                           | Tau de Goodmar et Kruskal | Lambda | V110 willing to fight ir war dépendant | ,109   | ,076                                      |                          | ,004 <sup>d</sup>        |
|                           |                           |        | V234 region dépendant                  | ,109   | ,076                                      |                          | ,004 <sup>d</sup>        |
|                           |                           |        |  |        |   |                          |                          |

a. L'hypothèse nulle n'est pas considérée.

b. Utilisation de l'erreur standard asymptotique dans l'hypothèse nulle.

c. Calcul impossible /: l'erreur standard asymptotique est égale à zéro.

d. Basé sur une approximation du Khi-deux

### Mesures symétriques

| V205 proud to be Canadian |                     |       |                               | Valeur | Erreur standard asymptotique <sup>a</sup> | T approximé <sup>b</sup> | Signification approximée |
|---------------------------|---------------------|-------|-------------------------------|--------|---|--------------------------|--------------------------|
| 1 oui                     | Nominal par Nominal | Phi   | Phi                           | ,279   |   |                          | ,000                     |
|                           |                     |       | V de Cramer                   | ,279   |   |                          | ,000                     |
|                           |                     |       | Coefficient de contingence    | ,269   |   |                          | ,000                     |
|                           | Ordinal par Ordinal | Gamma | Gamma                         | ,586   | ,048                                      | 8,478                    | ,000                     |
|                           |                     |       | Nombre d'observations valides | 1143   |   |                          |                          |
|                           |                     |       |                               |        |   |                          |                          |
| 2 non                     | Nominal par Nominal | Phi   | Phi                           | ,331   |   |                          | ,004                     |
|                           |                     |       | V de Cramer                   | ,331   |   |                          | ,004                     |
|                           |                     |       | Coefficient de contingence    | ,314   |   |                          | ,004                     |
|                           | Ordinal par Ordinal | Gamma | Gamma                         | ,633   | ,161                                      | 2,750                    | ,006                     |
|                           |                     |       | Nombre d'observations valides | 75     |   |                          |                          |
|                           |                     |       |                               |        |   |                          |                          |

a. L'hypothèse nulle n'est pas considérée.

b. Utilisation de l'erreur standard asymptotique dans l'hypothèse nulle.

## Corrélations partielles

- - - P A R T I A L C O R R E L A T I O N C O E F F I C I E N T S  
- -

Zero Order Partial

|      | V110                        | V234                        | V205                        |
|------|-----------------------------|-----------------------------|-----------------------------|
| V110 | 1,0000<br>( 0)<br>P= .      | ,3155<br>( 1216)<br>P= ,000 | ,1818<br>( 1216)<br>P= ,000 |
| V234 | ,3155<br>( 1216)<br>P= ,000 | 1,0000<br>( 0)<br>P= .      | ,2621<br>( 1216)<br>P= ,000 |
| V205 | ,1818<br>( 1216)<br>P= ,000 | ,2621<br>( 1216)<br>P= ,000 | 1,0000<br>( 0)<br>P= .      |

(Coefficient / (D.F.) / 2-tailed Significance)

" , " is printed if a coefficient cannot be computed

- - - P A R T I A L C O R R E L A T I O N C O E F F I C I E N T S  
- -

Controlling for.. V205

|      | V110                        | V234                        |
|------|-----------------------------|-----------------------------|
| V110 | 1,0000<br>( 0)<br>P= .      | ,2823<br>( 1215)<br>P= ,000 |
| V234 | ,2823<br>( 1215)<br>P= ,000 | 1,0000<br>( 0)<br>P= .      |

(Coefficient / (D.F.) / 2-tailed Significance)

" , " is printed if a coefficient cannot be computed



# Chapitre 13

## Les régressions et corrélations multiples

### *Analyse de données individuelles*

#### **Question 1**

a) Le sentiment de liberté est d'autant plus grand que les répondants ont un revenu élevé, qu'ils sont des hommes, qu'ils sont âgés et qu'ils se perçoivent comme faisant partie de la classe sociale supérieure.

*Note* : Nous avons transformé la variable «Revenu familial» (V227) suivant les instructions du cahier d'exercices. Nous avons aussi transformé la variable sexe en attribuant la valeur 0 aux femmes et la valeur 1 aux hommes. Nous avons créé trois variables factices à partir des catégories de la variable statut socio-économique (V226), la catégorie de référence étant le statut élevé (« upper ») :

MOYSUP : 1 : upmid; 0 : autres

MOYINF : 1 : lowmid; 0 : autres

COLBLEU : 1 : working; 0 : autres

Nous avons enfin codé la catégorie «other» de la variable V226 dans les valeurs manquantes.

Le modèle à estimer est donc :

$V66 = a + b_1(V227) + b_2(V214) + b_3(V216) - b_4(MOYSUP) - b_5(MOYINF) - b_6(COLBLEU)$ . Par hypothèse,  $b_1, b_2$  et  $b_3 > 0$  et  $b_4, b_5$  et  $b_6 < 0$ . Nous nous attendons aussi à ce que le sentiment de liberté soit d'autant plus faible que les répondants se placent dans une classe inférieure. Par conséquent, par hypothèse,  $b_4 < b_5 < b_6$ .

b) 2% ( $r^2$  ajusté = 0,015). L'ensemble des variables indépendantes a donc très peu d'impact sur la variable dépendante.

c) Le coefficient de corrélation multiple ( $R = 0,14$ ) est statistiquement significatif au seuil de 0,001 (sig. = 0,001).

d) Seule la variable INCOME (v227) a un effet statistiquement significatif sur la variable dépendante ( $p < 0,02$ ).

e) 0,087 ( $p < 0,02$ )

f) - 0,032 (n.s.)

g) Col bleu : -0,068 (n.s.) ; Moyen inférieur : -0,056 (n.s.) ; Moyen supérieur : -0,024 (n.s.).

h) La variable qui a l'effet le plus grand sur la variable dépendante est la variable REVENU car son Bêta est le plus grand. C'est la seule variable significative du modèle. Le signe du coefficient correspond à l'hypothèse : les répondants qui gagnent plus, se sentent significativement plus libres (signe positif).

i)

Tableau 13,1: Explication du sentiment de liberté chez les Canadiens (analyse de régression)

|                       | b         | Bêta   | t         |
|-----------------------|-----------|--------|-----------|
| Constante             | 7,579     |        | 26,193*** |
| Revenu (V227)         | 0,0000081 | 0,087  | 2,486*    |
| Sexe (V214)           | -0,113    | -0,31  | -1,073    |
| Age ( V216)           | -004      | -0,032 | -1,060    |
| MOYSUP                | -0.090    | -0,024 | -0,573    |
| MOYINF                | -0,214    | -0,056 | -1,301    |
| COLBLEU               | -0,391    | -0,068 | -1,704    |
| R <sup>2</sup> ajusté | 0,015     |        |           |
| N                     | 1189      |        |           |

\*\*\* p < 0,001; \*\* p < 0,01; \* p < 0,05 (test bilatéral)

Une seule hypothèse est confirmée, celle qui relie le revenu au sentiment de liberté : les Canadiens se disent d'autant plus libres qu'ils ont un revenu élevé. Le coefficient de cette variable est positif et significatif. À une augmentation de 10000\$ de revenu correspond une augmentation d'un peu moins d'un centième de point (0,081) dans le sentiment de liberté (0,0000081 points pour chaque dollar de revenu). Aucune des autres hypothèses n'est confirmée faute d'un coefficient significatif.

## Question 2

a) Notre hypothèse est la suivante : les répondants sont d'autant plus tolérants envers l'homosexualité qu'ils sont jeunes, éduqués et de sexe féminin.

Le modèle à estimer est le suivant :

$$V197 = a + b_1(V214) - b_2(V216) + b_3(V218).$$

Pour en faire une variable factice, nous avons transformé la variable sexe (V214) de la façon suivante :

Homme = 0

Femme = 1

Par hypothèse,  $b_1$  et  $b_3 > 0$  ;  $b_2 < 0$ .

- b) 14 % ( $R^2 = 0,143$ ).
- c) Le coefficient de corrélation multiple ( $R = 0,38$ ) est significatif au seuil 0,001.
- d) Sexe (V214),  $p < 0,01$ .  
Age (V216),  $p < 0,001$ .  
Age à la fin des études (V218),  $p < 0,001$ .
- e) + 0,210.
- f) - 0,277.
- g) C'est l'âge qui a l'impact le plus important avec le coefficient standardisé le plus élevé en valeur absolue (0,277). Suivent l'âge à la fin des études (0,210) et le sexe (0,074).

h)

Tableau 13,2: Explication de la tolérance à l'homosexualité des Italiens (analyse de régression)

|                                | <b>b</b> | <b>Béta</b> | <b>t</b>   |
|--------------------------------|----------|-------------|------------|
| Constante                      | 2,391    |             | 4,640***   |
| Sexe (V214)                    | 0,463    | 0,074       | 3,009**    |
| Age ( V216)                    | -0,053   | -0,277      | -11,029*** |
| Age à la fin des études (V218) | 0,261    | 0,21        | 8,367***   |
| R <sup>2</sup> ajusté          | 0,14     |             |            |
| N                              | 1409     |             |            |

\*\*\*  $p < 0,001$ ; \*\*  $p < 0,01$ ; \*  $p < 0,05$  (test bilatéral)

Les trois hypothèses sont confirmées ( $b_1$  et  $b_3$  sont positifs,  $b_2$  est négatif et les trois coefficients sont significatifs). Les Italiens sont d'autant plus tolérants à l'homosexualité qu'ils sont jeunes, scolarisés et qu'ils sont des femmes. Ces trois facteurs expliquent un mince 14% de la variance dans la variable dépendante. Sur une échelle allant de 1 (moins tolérant) à 10 (plus tolérant), les femmes ont un niveau de tolérance supérieur de un demi point (0,463) à celui des hommes. À chaque augmentation d'une année d'âge, correspond une diminution de cinq centièmes de point (0,053) du niveau de tolérance. Enfin, à chaque année additionnelle de scolarité correspond un niveau de tolérance supérieur de un quart de point (0,261).

## Régression

### Variables introduites/éliminées<sup>a</sup>

| Modèle | Variables introduites  | Variables éliminées | Méthode    |
|--------|--|---------------------|------------|
| 1      | V218 age finished school, V214 Gender, <sup>a</sup> V216 age | .                   | Introduire |

a. Toutes variables requises introduites

b. Variable dépendante : V197 homosexuality

### Récapitulatif du modèle

| Modèle | R                 | R-deux | R-deux ajusté | Erreur standard de l'estimation |
|--------|-------------------|--------|---------------|---------------------------------|
| 1      | ,381 <sup>a</sup> | ,145   | ,143          | 2,883                           |

a. Valeurs prédites : (constantes), V218 age finished school, V214 Gender, V216 age

### ANOVA<sup>b</sup>

| Modèle |            | Somme des carrés | ddl  | Carré moyen | F      | Signification     |
|--------|------------|------------------|------|-------------|--------|-------------------|
| 1      | Régression | 1980,868         | 3    | 660,289     | 79,452 | ,000 <sup>a</sup> |
|        | Résidu     | 11676,316        | 1405 | 8,311       |        |                   |
|        | Total      | 13657,184        | 1408 |             |        |                   |

a. Valeurs prédites : (constantes), V218 age finished school, V214 Gender, V216 age

b. Variable dépendante : V197 homosexuality

### Coefficients

| Modèle |                          | Coefficients non standardisés |                 | Coefficients standardisés | t       | Signification | Corrélations       |           |        |
|--------|--------------------------|-------------------------------|-----------------|---------------------------|---------|---------------|--------------------|-----------|--------|
|        |                          | B                             | Erreur standard | Bêta                      |         |               | Corrélation simple | Partielle | Partie |
| 1      | (constante)              | 2,391                         | ,515            |                           | 4,640   | ,000          |                    |           |        |
|        | V214 Gender              | ,463                          | ,154            | ,074                      | 3,009   | ,003          | ,055               | ,080      | ,074   |
|        | V216 age                 | -,053                         | ,005            | -,277                     | -11,029 | ,000          | -,314              | -,282     | -,272  |
|        | V218 age finished school | ,261                          | ,031            | ,210                      | 8,367   | ,000          | ,257               | ,218      | ,206   |

a. Variable dépendante : V197 homosexuality

## Analyse de données agrégées

### Question 1

a) Le PIB per capita est d'autant plus grand que l'indice de développement humain est élevé et que la législation est fiable.

Le modèle à estimer est :

$$\text{PIB\_TETE} = a + b_1(\text{IDH}) + b_2(\text{LOIS})$$

Par hypothèse,  $b_1$  et  $b_2 > 0$ .

b)  $\text{PIB\_TETE} = -2667,47 + 14570,51 (\text{IDH}) + 5313,09 (\text{LOIS})$

c) 78% ( $R^2$  ajusté = 0,783).

d) Le coefficient de corrélation multiple est statistiquement significatif au seuil 0,001.

e)

Tableau 13,3: Explication du niveau de développement économique (analyse de régression)

|  | <b>b</b>  | <b>Béta</b> | <b>t</b>  |
|--|-----------|-------------|-----------|
| Constante                                  | -2667,472 |             | -1,599    |
| IDH Indice de développement humain         | 14570,518 | 0,325       | 6,081***  |
| LOIS Indice de fiabilité de la législation | 5313,087  | 0,629       | 11,748*** |
| $R^2$ ajusté                               | 0,78      |             |           |
| N  | 149       |             |           |

\*\*\*  $p < 0,001$ ; \*\*  $p < 0,01$ ; \*  $p < 0,05$  (test bilatéral)

Les hypothèses de départ sont confirmées. Les pays ayant un plus grand niveau de développement humain et de fiabilité de l'appareil législatif ont un PIB per capita plus élevé. Plus précisément, à une augmentation d'un point du niveau de développement humain correspond une augmentation de 14570,51\$ du PIB per capita lorsque l'indice de fiabilité de l'appareil législatif est maintenu constant. De même, une augmentation d'un point du niveau de fiabilité de l'appareil législatif est accompagnée d'une augmentation de 5313,09\$ du PIB per capita lorsque l'indice de développement humain est maintenu constant. Les deux coefficients de régression sont significatifs au seuil 0,001.

## Question 2

a) Hypothèse : L'espérance de vie est d'autant plus élevée que le nombre de médecins est grand, le nombre de calories disponibles en pourcentage des besoins alimentaires est élevé et que le pays est classé dans la région économique OCDE.

Nous avons créé des variables factices à partir des catégories de la variable « région », la catégorie de référence étant l'OCDE :

EUROPEST = 1 si REGION = 2; sinon, EUROPEST = 0

ASIEPAC = 1 si REGION = 3; sinon, ASIEPAC = 0

AFRIQ = 1 si REGION = 4; sinon, AFRIQ = 0

AMERLAT = 1 si REGION = 6; sinon, AMERLAT = 0

MOYENOR = 1 si REGION = 5; sinon, MOYENOR = 0

Le modèle à estimer est le suivant :

$$\text{ESP\_VIE} = a + b_1\text{MEDECINS} + b_2\text{CALORIES} - b_3\text{EUROPEST} - b_4\text{ASIEPAC} - b_5\text{AFRIQ} - b_6\text{AMERLAT} - b_7\text{MOYENOR}$$

Par hypothèse,  $b_1$  et  $b_2 > 0$ , et  $b_3$  à  $b_7 < 0$ .

b) L'équation de régression est :

$$\text{ESP\_VIE} = 55,767 + 1,880\text{MEDECINS}^{**} + 0,120\text{CALORIES}^{***} - 5,995\text{EUROPEST}^* - 2,770\text{ASIEPAC} - 18,387\text{AFRIQ}^{***} - 4,636\text{MOYENOR}^* - 1,934\text{AMERLAT} + v$$

(\*\*\*  $p < 0,001$  ; \*\*  $p < 0,01$  ; \*  $p < 0,05$ )

c) 86% ( $R^2$  ajusté = 0,856).

d) Oui, au seuil 0,001.

e)

Tableau 13,4: Explication de l'espérance de vie (analyse de régression)

|                       | <b>b</b> | <b>Béta</b> | <b>t</b>  |
|-----------------------|----------|-------------|-----------|
| Constante             | 55,767   |             | 12,534*** |
| MÉDECINS              | 1,880    | 0,208       | 2,927**   |
| CALORIES              | 0,120    | 0,214       | 3,488***  |
| EUROPE DE L'EST       | -5,995   | -0,100      | -2,335*   |
| ASIE-PACIFIQUE        | -2,770   | -0,701      | -1,295    |
| AFRIQUE               | -18,387  | -0,701      | -7,796*** |
| MOYEN-ORIENT          | -4,636   | -0,118      | -2,324*   |
| AMÉRÉRIQUE LATINE     | -1,934   | -0,0636     | -1,034    |
| R <sup>2</sup> ajusté | 0,856    |             |           |
| N                     | 93       |             |           |

\*\*\* p < 0,001; \*\* p < 0,01; \* p < 0,05 (test bilatéral)

Deux hypothèses sont confirmées. L'espérance de vie est significativement plus élevée là où il y a plus de médecins et suffisamment de nourriture. Plus précisément, à une augmentation de 1 médecin par mille habitants correspond une augmentation de l'espérance de vie de 1,9 année et à une augmentation des calories disponibles équivalant à dix points de pourcentage des besoins alimentaires correspond une augmentation de l'espérance de vie de 1,2 année (0,120 années pour 1 point de pourcentage). L'hypothèse concernant les pays de l'OCDE est confirmée en partie seulement. Les pays d'Afrique, du Moyen-Orient et de l'Europe de l'Est ont une espérance de vie significativement plus faible que les pays de l'OCDE. Au total, le modèle explique 86% de la variance dans la variable dépendante.