

CHAPITRE 9 LES RELATIONS ENTRE VARIABLES

TABLE DES MATIÈRES

1. [Schéma de relations](#)
 2. [Relation entre deux variables qualitatives](#)
 3. [Relation entre deux variables quantitatives](#)
 4. [Les accidents de la route](#)
- [Exercices supplémentaires](#)
 - [Dossier](#)

Lorsque deux variables « varient » de concert, on peut affirmer qu'elles sont reliées. Mais quelle est la nature réelle de leur relation? Est-ce une relation de cause à effet? Autrement dit, est-ce que l'une des variables (la *variable dépendante*) dépend de l'autre (la *variable indépendante*) et si oui, laquelle, pourquoi et comment? Ou bien sommes-nous en présence d'une simple corrélation? Dans ce cas, se pourrait-il que les deux variables en jeu dépendent d'une troisième variable?

Au terme de ce chapitre, vous devriez être en mesure de répondre aux questions suivantes :

- Comment, en examinant les chiffres, peut-on déceler s'il existe une relation entre deux variables?
- Comment, lorsque les chiffres proviennent d'une enquête, peut-on déterminer si la relation observée peut être attribuable au hasard de l'échantillonnage?
- Quelle est la nature de la relation? Est-ce une relation de cause à effet ou est-ce une simple corrélation?
- Quelle est la force de la relation? Peut-on utiliser les résultats observés pour faire des prédictions intéressantes?

1. SCHÉMA DE RELATIONS

Qui influence quoi?

Avant de faire de savants calculs sur la force éventuelle d'une relation, il est indispensable d'établir comment les variables sont reliées les unes aux autres. Pour cela, rien ne vaut un bon vieux dessin. Nous vous proposons ici deux procédés (schéma et courbe) permettant de clarifier les relations entre les variables et de faire ressortir certains aspects intéressants de ces relations.

1.1. Sagesse populaire

En Italie, une croyance populaire très ancienne veut que le vin soit un fortifiant idéal pour un préadolescent un peu anémique. Un verre quotidien, de *rouge* évidemment, et le sang retrouve ses couleurs. Au pays de Descartes, cette coutume fait sourire. Au pays de Victoria, elle scandalise. Enfin, au pays de Ti-Poil, on se contente de taxer le vin sans juger personne. Et si la sagesse populaire italienne avait raison envers et contre tous?

Qu'est-ce qui se cache derrière cette corrélation?

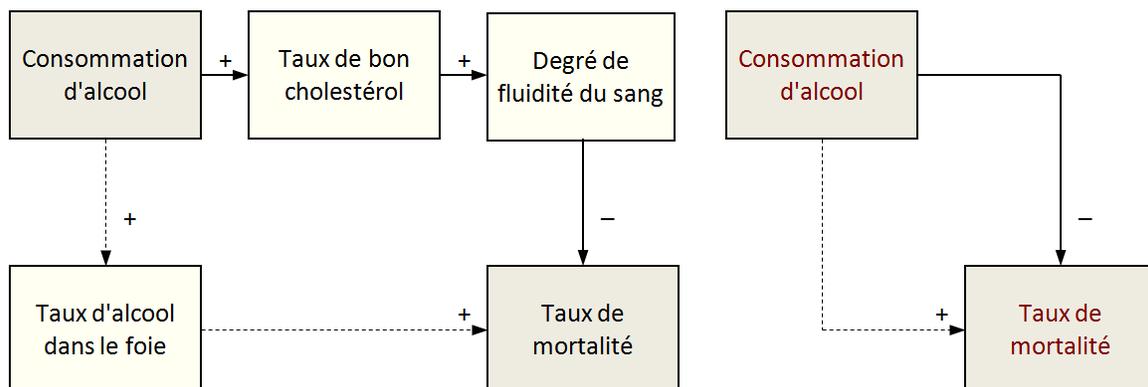
Selon une recherche citée par le magazine *Sciences et Avenir* de février 1995, on a observé une *corrélation négative* entre la consommation d'alcool et le taux de mortalité : *corrélation* parce que les deux variables varient en même temps, et *négative* parce qu'elles varient en sens inverse.

Il existe théoriquement trois manières de relier ces variables :

- 1) La consommation d'alcool (variable indépendante) exerce une influence sur le taux de mortalité (variable dépendante).
- 2) Le taux de mortalité (variable indépendante) exerce une influence sur la consommation d'alcool (variable dépendante).
- 3) La consommation d'alcool et le taux de mortalité sont toutes deux influencées par une troisième variable à découvrir.

La deuxième hypothèse doit être rejetée d'emblée parce qu'elle est absurde. Par ailleurs, c'est grâce à la découverte des chaînons manquants reliant les variables que nous pourrions finalement écarter la troisième hypothèse et retenir la première. C'est ce que nous avons voulu illustrer dans la figure 9.1.

FIGURE 9.1 – L'alcool, gage de santé jusqu'à un certain point



Comment construire un schéma de variables.

Le schéma de variables, dont nous avons déjà vu quelques exemples au [chapitre 6](#), illustre les relations de cause à effet (représentées par une flèche) entre les variables (représentées par des cases). Les signes qui accompagnent les flèches représentent le sens de la relation : un signe positif indique que les variables augmentent en même temps et diminuent en même temps (relation *directe*) ; un signe négatif indique que les variables varient en sens *inverse*. Évidemment, le signe ne s'applique pas pour des variables purement nominales dans lesquelles il n'y a ni ordre ni direction.

Dans le schéma, on constate qu'il y a une relation *directe* entre la consommation d'alcool et le taux de bon cholestérol : les deux variables varient dans le même sens, ce que nous indiquons par un signe positif. On note aussi qu'il existe une relation *inverse* entre le degré de fluidité du sang et le taux de mortalité : lorsque le premier augmente, le second diminue, ce que nous indiquons par un signe négatif.

Dans la partie droite de la figure 9.1, nous avons fait la synthèse de la relation en éliminant toutes les variables intermédiaires : la cause (consommation d'alcool) est reliée à l'effet (taux de mortalité) par une flèche surmontée d'un signe négatif. Ce signe est obtenu en combinant tous les signes successifs du schéma original, en prenant pour principe que la relation change de sens chaque fois que l'on rencontre un chiffre négatif.

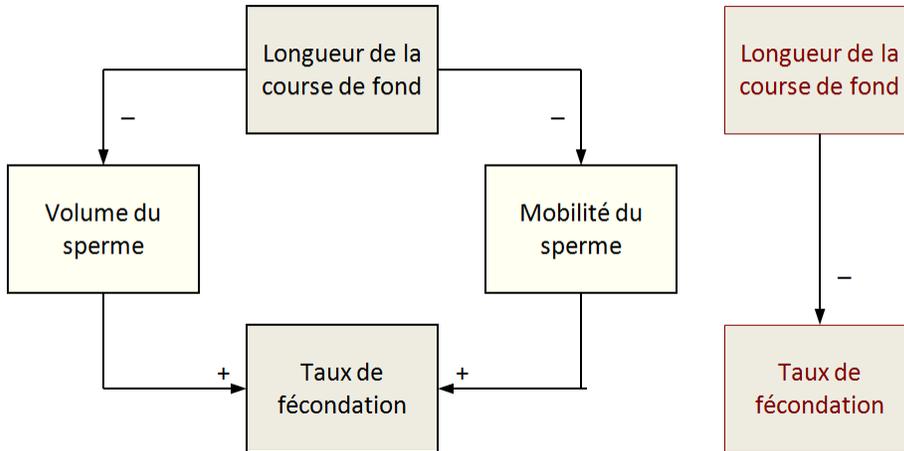
Alors, vite! Une bouteille et ça presse? Minute! Il est clair que nous nous sommes contentés d'étudier un seul aspect de la question. Passé un certain seuil, par exemple, la consommation d'alcool est reliée de façon *inverse* au taux de mortalité par l'intermédiaire de ce cher vieux foie que nous avons tous.

1.2. Tarzan au lit

Rien de tel qu'épouser une personne bardée de prix Nobel ou de médailles olympiques pour avoir des enfants beaux et forts, n'est-ce pas? Cependant, avant de faire votre demande en mariage, prenez note du fait suivant : on a observé, chez les hommes qui s'adonnent à la course de fond, que le sperme était moins volumineux et les spermatozoïdes moins mobiles que d'ordinaire. De quoi diminuer les chances de fécondation? On serait tenté de prétendre que le faible volume de sperme réduit la mobilité des spermatozoïdes, un peu à la manière des billots qui se traînent misérablement sur une rivière l'été. Il n'en est rien cependant. Les phénomènes sont seulement corrélés, car ils

dépendent d'une variable commune : la longueur de la course de fond, et l'état d'épuisement qui en résulte. C'est très sérieux! Nous l'avons également appris dans la revue *Sciences et Avenir* (voir le schéma de la figure 9.2).

FIGURE 9.2 - Rien ne sert de courir si on veut se reproduire



1.3. L'alcool au volant

Cette fois, l'alcool est plutôt nuisible.

Vous avez un ami qui a déjà conduit en état d'ivresse sans avoir d'accident. Cet ami en déduit que l'alcool et le volant font bon ménage, d'autant plus que son oncle, antialcoolique notoire, a déjà provoqué un carambolage monstre sur l'autoroute métropolitaine après avoir bu un verre d'eau.

Il n'en demeure pas moins que les conducteurs qui ont une prédilection pour la dive bouteille ont beaucoup plus de *chance* d'être impliqués dans un accident de la route que les non-buveurs. C'est ce qu'indique l'étude menée par Statistique Canada (voir la figure 9.3). Si un dessin vaut mille mots, une courbe vaut parfois mille calculs : nous avons tracé, sous la figure 9.3, des [courbes*](#) qui mettent en évidence la relation entre consommation d'alcool (sur l'axe horizontal) et taux d'accident (sur l'axe vertical).

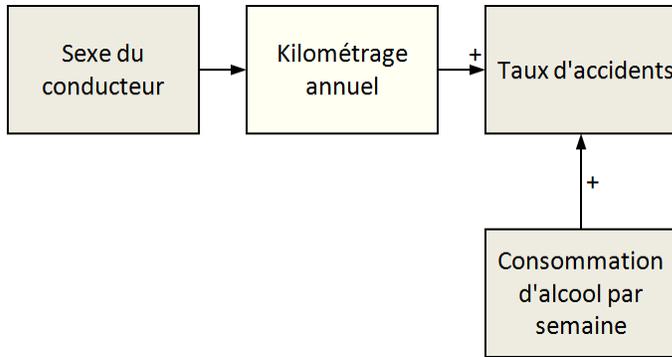
Ce genre de courbe n'est possible que si les échelles des variables sont, au minimum, des échelles ordinales.

FIGURE 9.3 - Accidents de la circulation et consommation d'alcool

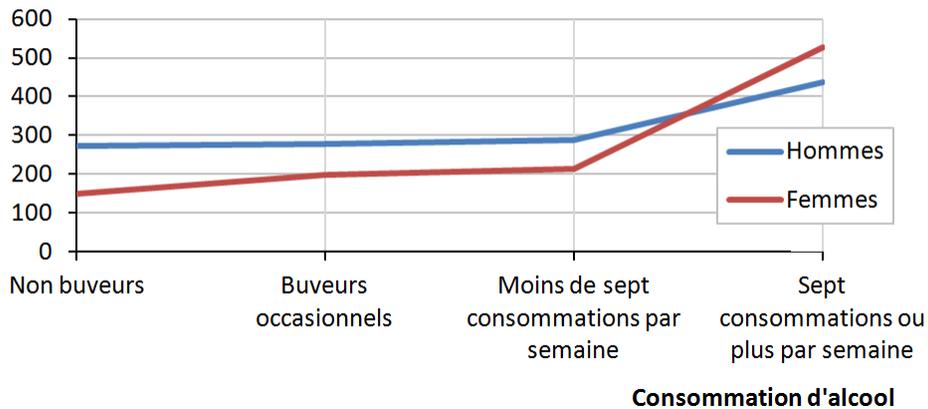
Taux d'accidents de la circulation pour 1000 habitants en 1988

| | Hommes | Femmes |
|---|--------|--------|
| Non buveurs | 274 | 149 |
| Buveurs occasionnels | 278 | 198 |
| Moins de sept consommations par semaine | 288 | 214 |
| Sept consommations ou plus par semaine | 437 | 526 |

Source : Statistique Canada, Les accidents au Canada, 1991. Données de 1988.



Taux d'accidents de la circulation pour 1000 habitants



Comment interpréter les différences entre hommes et femmes? Une fausse explication serait de dire que les femmes boivent moins que les hommes. Que cela soit vrai ou non, cela n'a pas d'influence ici puisqu'on nous donne des taux d'accidents par catégorie de buveurs, et non des fréquences brutes. Il se pourrait également que les femmes conduisent mieux que les hommes — sauf celles qui boivent le plus. Cette hypothèse n'est pas à rejeter, quoiqu'il soit facile de proposer une autre explication : il se peut que les femmes aient moins d'accidents parce qu'elles conduisent moins. Il ne reste plus qu'à trouver les chiffres qui permettent de confirmer ou de rejeter toutes ces hypothèses.

1.4. La télévision

Si la durée d'écoute hebdomadaire de la télévision est très élevée au Québec, elle commence néanmoins à décliner. Les méchantes langues diront que c'est parce que de nouvelles formes d'abrutissement ont été mises au point. Au fait, savez-vous qui regarde le plus la télévision : les hommes ou les femmes? Les « vrais » Québécois ou les « faux »? Encore une fois, des chiffres bien choisis vont pouvoir détruire (ou renforcer) quelques préjugés.

Dans la figure 9.4, nous avons choisi de représenter par des courbes la relation entre *heures d'écoute* (sur l'axe vertical, la position préférée de la variable dépendante) et *groupe d'âge* (sur l'axe horizontal)*. Les courbes mettent bien en évidence le fait que l'écoute de la télévision augmente avec l'âge. Cela dit, la véritable explication vient peut-être du fait que les vieux ont plus de temps libre

que les jeunes (voir le schéma de variables au centre de la figure 9.4). Pour vérifier cette hypothèse, on pourrait consulter une des nombreuses études sur le [temps libre](#)*. Cette influence de l'âge sur les heures d'écoute se retrouve systématiquement chez les deux sexes et auprès des deux groupes ethniques.

FIGURE 9.4 - Qui écoute la télévision?

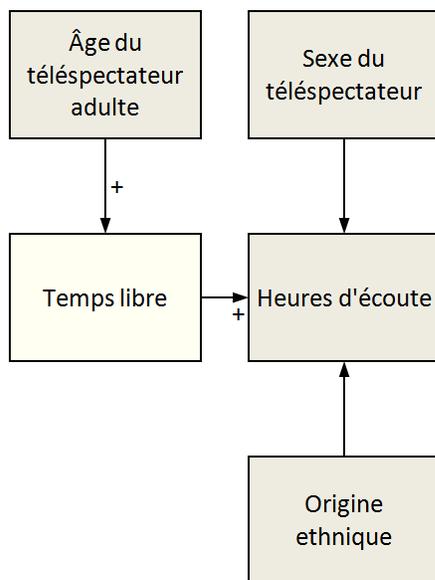
Heures d'écoute hebdomadaire moyenne de la télévision au Québec

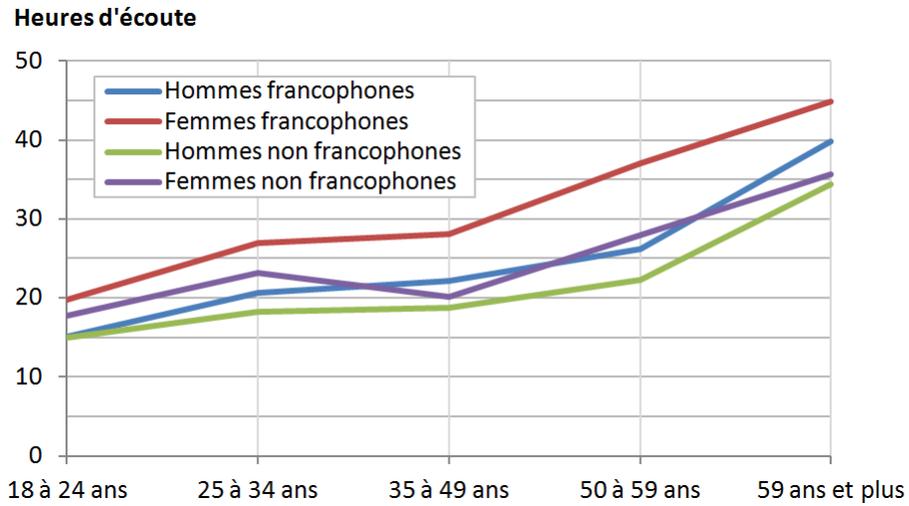
| | Francophones | | Non francophones | |
|---------------------------------------|--------------|-------------|------------------|-------------|
| | Hommes | Femmes | Hommes | Femmes |
| 18 à 24 ans | 15,1 | 19,8 | 15,0 | 17,7 |
| 25 à 34 ans | 20,6 | 27,0 | 18,3 | 23,2 |
| 35 à 49 ans | 22,1 | 28,1 | 18,7 | 20,1 |
| 50 à 59 ans | 26,2 | 37,1 | 22,3 | 27,9 |
| 59 ans et plus | 39,8 | 44,8 | 34,4 | 35,7 |
| Total adultes (18 ans et plus) | 24,3 | 31,8 | 22,3 | 25,8 |

Sources : Statistique Canada, L'écoute de la télévision, Automne 1994. Sondages BBM.

Il aurait été impossible de construire de telles courbes avec les autres variables (sexe et origine ethnique) qui appartiennent à des échelles nominales.

Par exemple : L'emploi du temps, Statistique Canada, 1991, SC 11-612F, no 4.

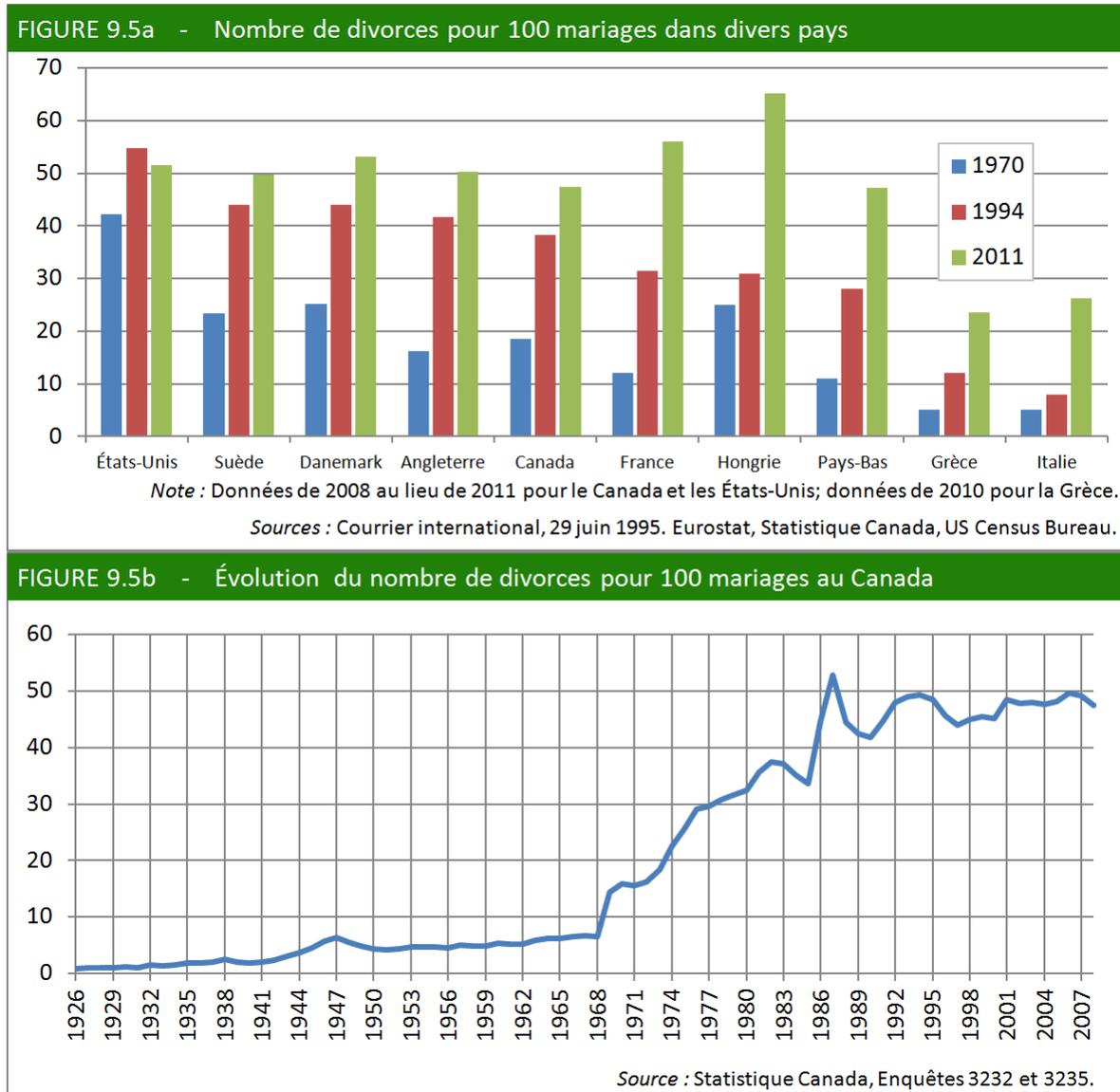




Encore une fois, l'utilisation de courbes et d'un schéma de variables a permis de mettre en évidence la relation et d'enrichir la description du phénomène.

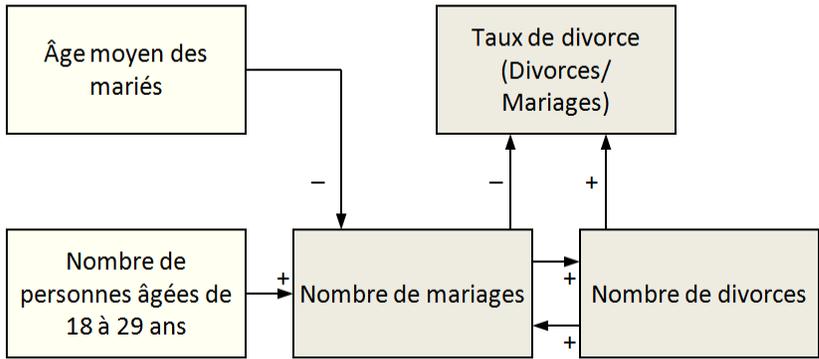
1.5. Un cas plus complexe : les deux variables s'influencent mutuellement

Le taux de divorce, c'est-à-dire le nombre de divorces par rapport au nombre de mariages, est en hausse dans tous les pays industrialisés. En observant la figure 9.5, on se demande si le divorce et le développement économique ne sont pas corrélés.



Nous utiliserons ce dernier exemple pour montrer qu'il n'est pas toujours facile de distinguer la variable dépendante de la variable indépendante dans une relation de cause à effet. Le nombre de divorces dépend, entre autres, du nombre de mariages : seuls les gens mariés peuvent divorcer. Mais le nombre de divorces peut, lui aussi, influencer le nombre de mariages : chaque divorcé est un futur marié potentiel remis « sur le marché ». Nous avons illustré cette influence réciproque à la figure 9.6, que nous avons par ailleurs enrichie en tenant compte de certains facteurs démographiques : lorsqu'une vague de jeunes arrive à l'âge de convoler, le nombre de mariages augmente naturellement, suivi, quelques années plus tard par une hausse du nombre de divorces.

FIGURE 9.6 - Interrelation entre divorce et mariage

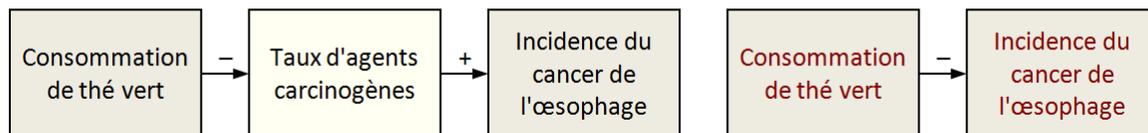


EXERCICES 1

1. Le thé, autre gage de santé

Répondez aux questions en vous référant au schéma de variable de la figure 9.7.

FIGURE 9.7 - Le thé vert est bon pour la santé



- Quelle est la nature de la relation entre la consommation de thé vert et l'incidence du cancer de l'œsophage (simple corrélation ou relation de cause à effet)?
- Si la relation en est une de cause à effet, quelle est la variable associée à la cause, quelle est la variable associée à l'effet et quel est le sens de la relation (direct ou inverse)?

2. Quand on n'en a pas on l'aime, quand on en a on s'ennuie

- Pour mettre en évidence la relation entre l'âge et l'organisation du temps des individus, tracez des courbes, à partir des données du tableau 9.1.

TABLEAU 9.1 - La répartition du temps des Canadiens

| (nombre d'heures) | Soins personnels | Activités productrices | Activités de loisir | Non précisé | Total |
|-------------------|------------------|------------------------|---------------------|-------------|-------------|
| 15 à 24 ans | 10,9 | 7,6 | 5,4 | 0,0 | 24,0 |
| 25 à 34 ans | 10,6 | 8,5 | 4,8 | 0,0 | 24,0 |
| 35 à 44 ans | 10,7 | 8,7 | 4,6 | 0,1 | 24,0 |
| 45 à 54 ans | 10,7 | 8,0 | 5,2 | 0,1 | 24,0 |
| 55 à 64 ans | 11,4 | 6,3 | 6,2 | 0,1 | 24,0 |
| 65 ans et plus | 12,4 | 3,8 | 7,7 | 0,1 | 24,0 |

Source : Statistique Canada, L'emploi du temps, 1991. Données de 1986.

Note : Soins personnels : dormir, se nourrir, se laver, se vêtir.

Activités productrices : travail rémunéré, activités éducatives, travaux ménagers, soin des enfants, etc.

Activités de loisir : télévision, lecture, divertissement, sport, passe-temps, bénévolat, etc.

- À propos de la [figure 9.4](#), nous avons émis une hypothèse selon laquelle les personnes plus âgées ont de bonnes raisons d'écouter la télévision, étant donné qu'elles disposent de plus de temps libre que les jeunes. Le tableau 9.1 confirme-t-il cette hypothèse?

2. RELATION ENTRE DEUX VARIABLES QUALITATIVES

Après avoir constaté la présence d'une relation entre deux variables, il reste à évaluer la force de cette relation. Par ailleurs, si les chiffres obtenus proviennent d'un simple échantillon, il faut prendre quelques précautions supplémentaires en s'assurant que les résultats observés ne sont pas l'effet du hasard.

2.1. Le tableau croisé

Chaque colonne du tableau croisé correspond à une catégorie d'une des deux variables et chaque ligne correspond à une catégorie de l'autre variable.

Lorsque les variables sont qualitatives, on a souvent recours à un croisement entre les catégories de chaque [variable](#)*. Dans le cas du tabagisme chez les jeunes, par exemple, on pourrait croiser la variable *sexe* (en colonnes) avec la variable *consommation de cigarettes* (en lignes) pour déterminer si le sexe a une influence sur le comportement (l'inverse est peu probable!). Chaque case du tableau croisé contiendra une fréquence, c'est-à-dire le nombre d'individus correspondant simultanément à la caractéristique de la colonne et à celle de la ligne.

Nous examinerons le cas des variables quantitatives dans la [section suivante](#).

Dans chaque case du tableau croisé, on inscrit les fréquences associées à la catégorie de la colonne et à celle de la ligne correspondantes.

Les données du [tableau 9.2](#) et du [tableau 9.3](#) proviennent de deux enquêtes effectuées dans les années 1990 sur les adolescents au Québec. Nous les avons extraites d'un article publié dans la *Revue québécoise de psychologie* dans lequel l'auteur cherche à montrer que la perception des gens à l'égard des adolescents est plus influencée par les préjugés que par la réalité. Notre propos est ici beaucoup plus modeste : nous utiliserons deux exemples de cette recherche pour étudier la présence et la force d'une relation entre deux variables qualitatives. Le premier exemple traite d'un comportement (« fumes-tu la cigarette? ») et le second d'une perception des choses (« mes parents se chicanent souvent entre eux »). Nous essaierons de déterminer si le sexe a une influence sur le comportement (dans le premier cas) ou sur les perceptions (dans le second).

Commençons par observer les données brutes de l'enquête sur le tabagisme. Dans le tableau 9.2a, on constate que 5580 élèves du secondaire ont été interrogés, dont 2650 garçons et 2930 filles. Sur les 2930 filles interrogées, 322 déclarent fumer régulièrement, 355 à l'occasion et 2253 jamais. Toutes proportions gardées, puisque l'échantillon contient davantage de filles que de garçons, ces dernières semblent plus enclines à fumer que leurs confrères.

TABLEAU 9.2 - Le tabagisme chez les élèves du secondaire au Québec : « Fumes-tu la cigarette? »

Tableau 9.2a - Fréquences observées

| | Garçons | Filles | Total |
|---------------|-------------|-------------|-------------|
| Régulièrement | 196 | 322 | 518 |
| À l'occasion | 217 | 355 | 572 |
| Jamais | 2237 | 2253 | 4490 |
| Total | 2650 | 2930 | 5580 |

Tableau 9.2b - Proportion de chaque case (en %)

| | Garçons | Filles | Total |
|---------------|-------------|-------------|--------------|
| Régulièrement | 3,5 | 5,8 | 9,3 |
| À l'occasion | 3,9 | 6,4 | 10,3 |
| Jamais | 40,1 | 40,4 | 80,5 |
| Total | 47,5 | 52,5 | 100,0 |

Tableau 9.2c - Fréquences théoriques

| | Garçons | Filles | Total |
|---------------|-------------|-------------|-------------|
| Régulièrement | 246 | 272 | 518 |
| À l'occasion | 272 | 300 | 572 |
| Jamais | 2132 | 2358 | 4490 |
| Total | 2650 | 2930 | 5580 |

Tableau 9.2d - Écarts au carré relatifs

| | Garçons | Filles | Total |
|---------------|--------------|--------------|--------------|
| Régulièrement | 10,16 | 9,19 | 19,36 |
| À l'occasion | 10,99 | 9,94 | 20,94 |
| Jamais | 5,14 | 4,65 | 9,78 |
| Total | 26,29 | 23,78 | 50,08 |

Source : données brutes tirées de l'article de Richard CLOUTIER, « L'image des adolescents rongée par les mythes », Revue québécoise de psychologie, vol. 16, n° 3, 1995.

Note : Données de 1991. Enquête menée au Québec auprès des élèves du secondaire.

Pour en avoir le cœur net, observons les proportions calculées dans le tableau 9.2b. Étant donné que les filles représentent 52,5 % de l'échantillon, on peut s'attendre à retrouver dans la colonne « filles » des proportions légèrement supérieures à celles de la colonne « garçons ». Or, on constate

que sur les 9,3 % d'élèves qui fument régulièrement, 5,8 % sont des filles et 3,5 % sont des garçons : on ne peut plus parler de « légères » différences.

Étant donné que 9,3 % des élèves fument régulièrement et que 47,5 % d'entre eux sont des garçons, on pourrait s'attendre à ce que 9,3 % x 47,5 % des fumeurs réguliers soient des garçons. Sur les 5580 élèves interrogés, il devrait donc y avoir « théoriquement » $9,3 \% \times 47,5 \% \times 5580 = 246$ garçons classés comme fumeurs réguliers. C'est ce qu'on appelle la fréquence théorique (tableau 9.2c), par opposition à la fréquence observée (tableau 9.2a).

$$\text{Fréquence théorique d'une case} = \text{Proportion de la ligne} \times \text{Proportion de la colonne} \times \text{Fréquence totale}$$

$$\begin{aligned} \text{Fréquence théorique de filles qui ne fument jamais} &= 80,5 \% \times 52,5 \% \times 5580 = 2358 \\ (\text{Ou encore : } 0,805 \times 0,525 \times 5580 &= 2358) \end{aligned}$$

Si le sexe n'exerçait aucune influence sur le tabagisme, la répartition des élèves entre les diverses catégories (tableau 9.2a) devrait ressembler à celle obtenue dans le tableau des fréquences théoriques (tableau 9.2c). Lorsque l'on compare ces fréquences théoriques aux fréquences observées dans l'enquête, on constate toutefois certains écarts, qui ne sont ni énormes ni négligeables. L'utilisation d'un outil assez répandu, le Khi carré, va nous permettre d'évaluer l'importance de cet écart et de l'interpréter.

Examinons la première case du tableau 9.2a : 196 individus y sont recensés alors qu'on s'attendait à en retrouver 246 (voir tableau 9.2c). On est donc en déficit de 50 individus (-50) sur un total de 246. Comme c'est l'écart absolu qui est important, nous nous débarrassons du signe en mettant cet écart au carré avant de le diviser par le total. Nous obtenons ainsi l'écart au carré relatif.

$$\text{Écart}^2 \text{ relatif} = (\text{Fréquence observée} - \text{Fréquence théorique})^2 / \text{Fréquence théorique}$$

$$\text{Écart}^2 \text{ relatif pour les garçons qui fument régulièrement} = (196 - 246)^2 / 246 = 2500 / 246 = 10,16$$

On retrouve ce dernier chiffre dans le tableau 9.2d. Le Khi carré n'est autre que la somme de tous ces écarts : il est ici de 50,08. Si les fréquences observées correspondaient exactement aux fréquences théoriques, le Khi carré aurait une valeur de 0. Il nous faut maintenant interpréter la valeur de 50,08 que nous avons obtenue.

2.2. L'écart peut-il être attribué au hasard?

Les données que nous venons d'utiliser proviennent d'une enquête. Faute d'information précise sur la population des élèves du secondaire dans son ensemble, nous devons nous contenter d'un échantillon. Mais quelles sont les chances que nous soyons tombés, par malheur, sur un échantillon non représentatif? Comme pour le test d'hypothèse (vu au [chapitre précédent](#)), nous allons nous donner un seuil de signification, mettons 0,05 (ou 5 %, ou 1/20). Cela signifie que nous ne voulons courir le risque de nous tromper qu'une fois sur 20 si jamais nous émettons l'hypothèse selon laquelle les écarts sont suffisamment grands pour ne pas être le simple effet du [hasard](#)*.

Nous aurions pu prendre un seuil de signification plus grand ou plus petit : le tout dépend du degré du risque que nous sommes prêts à accepter dans le contexte.

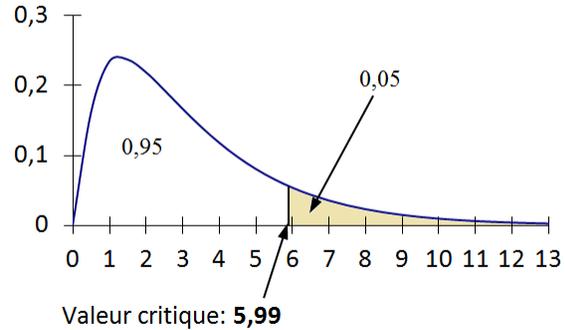
La figure 9.8 nous indique la valeur critique, c'est-à-dire la valeur minimale que doit atteindre notre Khi carré pour que nous puissions accepter notre hypothèse avec un risque de nous tromper

inférieur à 5 %. Mais comme le Khi carré est influencé par la taille du tableau (plus le tableau est grand, plus les écarts s'accumulent), il nous faut tenir compte du nombre de colonnes et lignes du tableau. C'est ce qu'on appelle le nombre de degrés de liberté.

FIGURE 9.8 - Distribution du Khi carré

Valeur critique selon le seuil de signification et le nombre de degrés de liberté

| Degrés de liberté | Seuil de signification | | |
|-------------------|------------------------|-------------|-------|
| | 0,1 | 0,05 | 0,01 |
| 1 | 2,71 | 3,84 | 6,64 |
| 2 | 4,61 | 5,99 | 9,21 |
| 3 | 6,25 | 7,82 | 11,34 |
| 4 | 7,78 | 9,49 | 13,28 |
| 5 | 9,24 | 11,07 | 15,09 |
| 6 | 10,64 | 12,59 | 16,81 |
| 7 | 12,02 | 14,07 | 18,48 |
| 8 | 13,36 | 15,51 | 20,09 |
| 9 | 14,68 | 16,92 | 21,67 |



$$\text{Degrés de liberté} = (\text{Nombre de colonnes} - 1) \times (\text{Nombre de lignes} - 1)$$

$$\text{Degrés de liberté} = (2 - 1) \times (3 - 1) = 1 \times 2 = 2$$

Dans la table de distribution du Khi carré, la valeur correspondant à un seuil de signification de 0,05 et à 2 degrés de liberté est égale à 5,99. Or, l'écart que nous avons calculé (50,08) dépasse largement cette *valeur critique*. En fait, nous sommes si loin au-dessus de la valeur critique qu'il est pratiquement *impossible* que les écarts constatés soient imputables à un hasard d'échantillonnage. Cette certitude n'est pas étrangère au fait que nous disposons d'un échantillon très élevé : la loi des grands nombres joue encore en notre faveur. Autrement dit, le Khi carré a d'autant plus de chances d'être élevé que l'échantillon est grand.

Le nombre de degrés de liberté peut s'interpréter de la façon suivante : nous avons sous la main un tableau dans lequel on croise les deux catégories d'une première variable (le sexe) avec les trois catégories d'une seconde variable (le tabagisme). Il va de soi que les totaux rajoutés aux lignes et aux rangées du tableau sont déterminés d'avance et ne relèvent en rien du hasard. Or, une fois que l'on connaît les fréquences associées aux garçons, par exemple, on peut en déduire automatiquement les fréquences des filles. De la même façon, une fois connues les fréquences des deux premières catégories concernant le tabagisme, nous connaissons également la fréquence de la troisième catégorie. En somme, la dernière colonne et la dernière ligne du tableau (totaux exclus) sont prédéterminées par les autres colonnes et lignes. C'est pourquoi on les exclut dans le calcul du nombre de degrés de liberté.

2.3. La relation est-elle forte?

Si nous venons d'éviter une première embûche, il nous faut maintenant évaluer dans quelle mesure la relation observée est forte. Nous pouvons affirmer que les filles sont plus portées à fumer que les garçons. Mais cela est-il suffisant, par exemple, pour nous permettre de prédire facilement, à partir de son sexe, si un élève fume?

Si nous avons croisé le sexe des élèves avec une autre variable, comme le port du soutien-gorge ou de la coquille protectrice dans les arts martiaux, nous aurions sûrement eu des écarts encore plus

tranchés. Notre Khi carré aurait été suffisamment élevé pour que nous puissions déduire à partir de ses habitudes vestimentaires, et sans grand risque de nous tromper, si un *budōka* est une fille ou un garçon.

Un des instruments les plus courants pour évaluer la force de la relation entre deux variables qualitatives est le *V de Cramer*. On comprendra que le Khi carré ne peut tout dire à lui tout seul, puisqu'il dépend en partie de la taille de l'échantillon. Le V de Cramer tient compte de cet aspect. Notez, en observant la formule ci-après, que vous auriez pu facilement inventer vous-même un coefficient correspondant et passer ainsi à la postérité.

$$V \text{ de Cramer} = \sqrt{\frac{\text{Khi}^2}{n \times (K - 1)}}$$

où n représente la taille de l'échantillon et K le nombre minimal de rangées et de colonnes.

On compte 3 rangées et 2 colonnes dans le [tableau 9.2](#). C'est le plus petit de ces deux chiffres que l'on retient pour calculer le V de Cramer.

$$V \text{ de Cramer} = \sqrt{[50,08/5580 \times (2-1)]} = 0,095$$

S'il n'y avait aucune relation entre les deux variables, le V de Cramer serait égal à 0, tout comme le Khi carré. Si la relation était parfaitement tranchée (par exemple si toutes les filles fumaient et si tous les garçons ne fumaient pas), le V de Cramer serait égal à 1. Dans le cas du port du soutien-gorge ou de la coquille protectrice, on aurait peut-être un V de Cramer égal à 0,99. En général, on considère que l'association entre les variables commence à être intéressante à partir de 0,10, forte à partir de 0,40 et robuste à partir de 0,70, mais cela dépend du contexte.

Si vous décidez d'entreprendre une recherche sur le sujet, nous vous prions de nous communiquer les résultats.

Ici, nous avons un coefficient relativement faible (0,095). Il y a certes une différence de comportement entre les garçons et les filles, mais cette différence est trop faible pour qu'on puisse prédire, à partir de son sexe, si un élève fume. Dans les paragraphes qui suivent, nous proposerons deux exemples de conclusions que l'on pourrait tirer à la lecture de ce chiffre : la première est acceptable et la seconde, abusive.

Une conclusion honnête

Il est clair que les écolières fument un peu plus que les garçons. Pour mieux comprendre la situation, nous pourrions nous poser diverses questions : « Pourquoi telle fille fume-t-elle? Pourquoi telle autre ne fume-t-elle pas? Qu'est-ce qui pousse un jeune à fumer? Qu'est-ce qui pousse un garçon à fumer? », etc.

Une conclusion biaisée

« Puisque les filles fument plus que les garçons, je refuse d'embaucher des filles dans mon usine de dynamite. » Cette affirmation serait aussi ridicule que les suivantes : « Puisque les faux-monnayeurs sont plus souvent des étrangers (V de Cramer à l'appui), un commerçant ne devrait pas accepter les billets de ses clients italiens » ou encore « Puisqu'une proportion relativement grande d'Asiatiques ont un QI supérieur à 100, les universités de devraient pas engager de professeurs originaires d'Europe. »

2.4. Mes parents se chicanent, un peu, beaucoup

Les données du tableau 9.3 proviennent d'une enquête effectuée auprès d'un échantillon de 3180 élèves du secondaire du Québec âgés de 11 à 19 ans. L'échantillon initial comptait 6121 élèves choisis au hasard, mais 2916 questionnaires ne furent pas remplis et 25 questionnaires furent rejetés.

Certaines des questions portaient sur la violence verbale et physique au sein de la famille. On demandait notamment à l'élève d'évaluer si ses parents se chicanèrent souvent entre eux, à partir de l'échelle ordinale suivante : 1. Correspond tout à fait à ce que je vis; 2. Correspond un peu à ce que je vis; 3. Ne correspond pas vraiment à ce que je vis; 4. Ne correspond pas du tout à ce que je vis.

TABLEAU 9.3 - Les perceptions des adolescents sur leurs parents
Mes parents se chicanent souvent...

Tableau 9.3a - Fréquences observées

| | Garçons | Filles | Total |
|--------------|-------------|-------------|-------------|
| Tout à fait | 58 | 125 | 183 |
| Un peu | 179 | 269 | 448 |
| Pas vraiment | 395 | 524 | 919 |
| Pas du tout | 738 | 842 | 1580 |
| Total | 1370 | 1760 | 3130 |

Tableau 9.3b - Proportion de chaque case (en %)

| | Garçons | Filles | Total |
|--------------|-------------|-------------|--------------|
| Tout à fait | 1,9 | 4,0 | 5,8 |
| Un peu | 5,7 | 8,6 | 14,3 |
| Pas vraiment | 12,6 | 16,7 | 29,4 |
| Pas du tout | 23,6 | 26,9 | 50,5 |
| Total | 43,8 | 56,2 | 100,0 |

Tableau 9.3c - Fréquences théoriques

| | Garçons | Filles | Total |
|--------------|-------------|-------------|-------------|
| Tout à fait | 80 | 103 | 183 |
| Un peu | 196 | 252 | 448 |
| Pas vraiment | 402 | 517 | 919 |
| Pas du tout | 692 | 888 | 1580 |
| Total | 1370 | 1760 | 3130 |

Tableau 9.3d - Écarts au carré relatifs

| | Garçons | Filles | Total |
|--------------|--------------|-------------|--------------|
| Tout à fait | 6,10 | 4,75 | 10,84 |
| Un peu | 1,49 | 1,16 | 2,65 |
| Pas vraiment | 0,13 | 0,10 | 0,23 |
| Pas du tout | 3,12 | 2,43 | 5,54 |
| Total | 10,83 | 8,43 | 19,27 |

Source : données brutes tirées de l'article de Richard CLOUTIER, « L'image des adolescents rongée par les mythes », Revue québécoise de psychologie, vol. 16, n° 3, 1995.

Note : Données de 1994. Enquête menée au Québec auprès des élèves du secondaire.

Sur les 3130 élèves qui ont été en mesure de répondre à cette question précise, les filles semblent relativement plus nombreuses à estimer que leurs parents se chicanent. Étant donné que les garçons et les filles partagent généralement les mêmes parents, on peut considérer que les réponses reflètent non seulement la réalité, mais également la *perception* de cette réalité. Par ailleurs, il faut noter que la plupart des parents sont bel et bien des adeptes de la coexistence pacifique.

Des calculs similaires à ceux effectués pour le [tableau 9.2](#) nous montrent que le Khi carré est égal à 19,27. Ce tableau contient cependant plus de cases que le précédent : on y compte 2 colonnes et 4 lignes, soit $(2 - 1) \times (4 - 1) = 1 \times 3 = 3$ degrés de liberté. Si l'on prend ici encore un seuil de signification de 0,05, la valeur critique est de 7,82 (revoir la [figure 9.8](#)). Notre Khi carré est donc suffisamment grand pour qu'on ne puisse pas mettre les différences observées entre garçons et filles sur le dos du hasard.

Encore une fois, le V de Cramer est relativement petit : $\sqrt{[19,27/3130 \times (2 - 1)]} = 0,08$. L'association entre le sexe et la perception de la réalité est donc plutôt faible. Même si la relation existe, il serait présomptueux de formuler des généralisations.

EXERCICES 2

1. Des jeunes drogués

- a) Vérifiez dans le [tableau 9.2](#) la proportion, la fréquence théorique et l'écart pour les filles qui fument régulièrement.
- b) Quelle aurait été la valeur critique du Khi carré si nous avons choisi un seuil de signification de 0,01 pour le tableau 9.2? Commentez.

2. Une jeunesse qui s'envole en fumée

Le tableau 9.4 contient des informations similaires à celles du [tableau 9.2](#). Les données sont cependant tirées d'une enquête différente et portent sur une période plus récente.

TABLEAU 9.4 - Le tabagisme chez les élèves du secondaire au Québec quelques années plus tard : « Fumes-tu la cigarette? »

Fréquences observées

| | Garçons | Filles | Total |
|---------------|-------------|-------------|-------------|
| Régulièrement | 174 | 341 | 515 |
| À l'occasion | 153 | 353 | 506 |
| Jamais | 1054 | 1100 | 2154 |
| Total | 1381 | 1794 | 3175 |

Source : données brutes tirées de l'article de Richard CLOUTIER, « L'image des adolescents rongée par les mythes », Revue québécoise de psychologie, vol. 16, n^o 3, 1995.

Note : Données de 1994. Enquête menée au Québec auprès des élèves du secondaire.

- a) Construisez un tableau contenant les proportions pour chaque case.
- b) Construisez un tableau contenant les fréquences théoriques pour chaque case.
- c) Construisez un tableau contenant les écarts au carré relatifs.
- d) Calculez le Khi carré. Comparez à la valeur critique dans le tableau de distribution du Khi carré pour un seuil de signification de votre choix.
- e) Calculez le V de Cramer.
- f) Comparez les résultats obtenus à ceux du [tableau 9.2](#). Commentez.

3. RELATION ENTRE DEUX VARIABLES QUANTITATIVES

Comme nous venons de le voir, le tableau croisé est un outil privilégié pour observer une relation entre deux variables *qualitatives*. Lorsque les deux variables sont *quantitatives*, on doit avoir recours à un autre procédé : la corrélation. La situation suivante expliquera de quoi il en retourne.

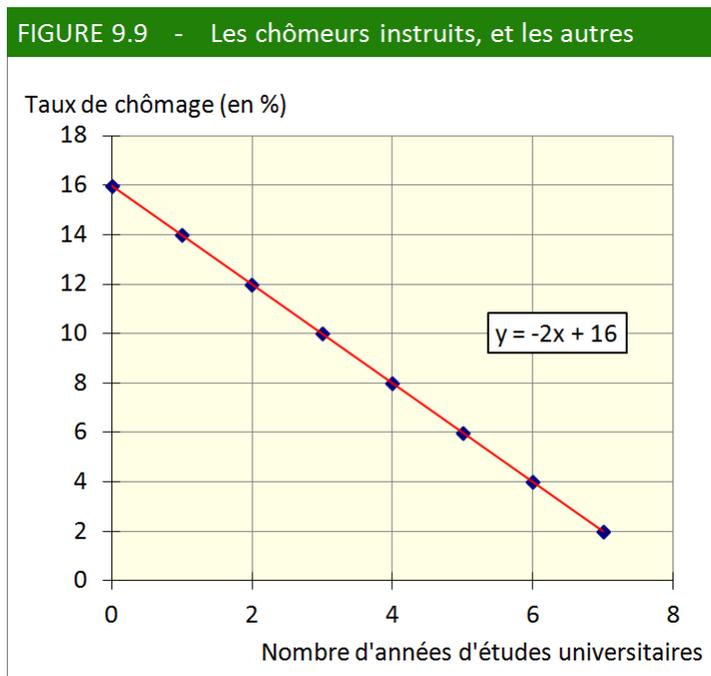
3.1. La corrélation

On dit souvent qu'il y a des chômeurs instruits. Même si cela est vrai, il s'agit néanmoins d'une espèce relativement rare si on la compare à celle des chômeurs « ignorants ». Supposons que le taux de chômage soit en moyenne de 2 % pour une personne qui détient un doctorat (7 ans d'études universitaires), de 6 % pour une personne qui détient une maîtrise (5 ans d'études universitaires) et de 10 % pour une personne qui détient un baccalauréat (3 ans d'études universitaires). Si la tendance se maintient, comme on dit, il est probable que le taux de chômage soit de 16 % pour une personne qui a interrompu ses études juste avant de rentrer à l'université. On pourrait même construire une formule qui permette de prédire le taux de chômage associé à un nombre x d'années d'études universitaires : Puisqu'il semble que chaque année d'étude fasse baisser le chômage de 2 points de pourcentage, la formule du taux de chômage serait donc la suivante :

$$y = 16 - 2x$$

Lorsque des points, représentant la valeur d'une variable par rapport à une autre variable, sont plus ou moins alignés, la droite de régression est celle qui s'éloigne le moins possible de l'ensemble des points.

Dans cette équation, y représente (en points de pourcentage) le taux de chômage, et x , le nombre d'années d'études universitaires. Après 4 années d'études universitaires, le taux de chômage serait, selon notre formule, de $16 - 2 \times 4 = 8$ points de pourcentage. L'équation peut également être représentée sous forme de courbe (ou droite de régression) comme dans la figure 9.9.



Il y a corrélation entre deux variables lorsque ces deux variables se suivent de façon plus ou moins systématique, que ce soit dans le même sens ou en sens inverse.

Dans l'exemple que nous venons de présenter, on peut dire qu'il existe une corrélation parfaite entre les deux variables : le niveau universitaire atteint et le taux de chômage. Évidemment, la réalité humaine n'est pas aussi simple, et de toute façon, on ne peut pas tirer de grandes conclusions d'un échantillon aussi petit. Il existe cependant de nombreuses situations reliées aux sciences humaines dans lesquelles il est possible de tracer une droite de régression mettant en relation deux variables x et y . Il s'agit alors de déterminer la valeur des paramètres qui caractérisent l'équation de cette droite. Ces deux paramètres sont ici les nombres 16 et -2 . On cherchera alors à évaluer dans quelle mesure les faits observés coïncident avec la droite tracée.

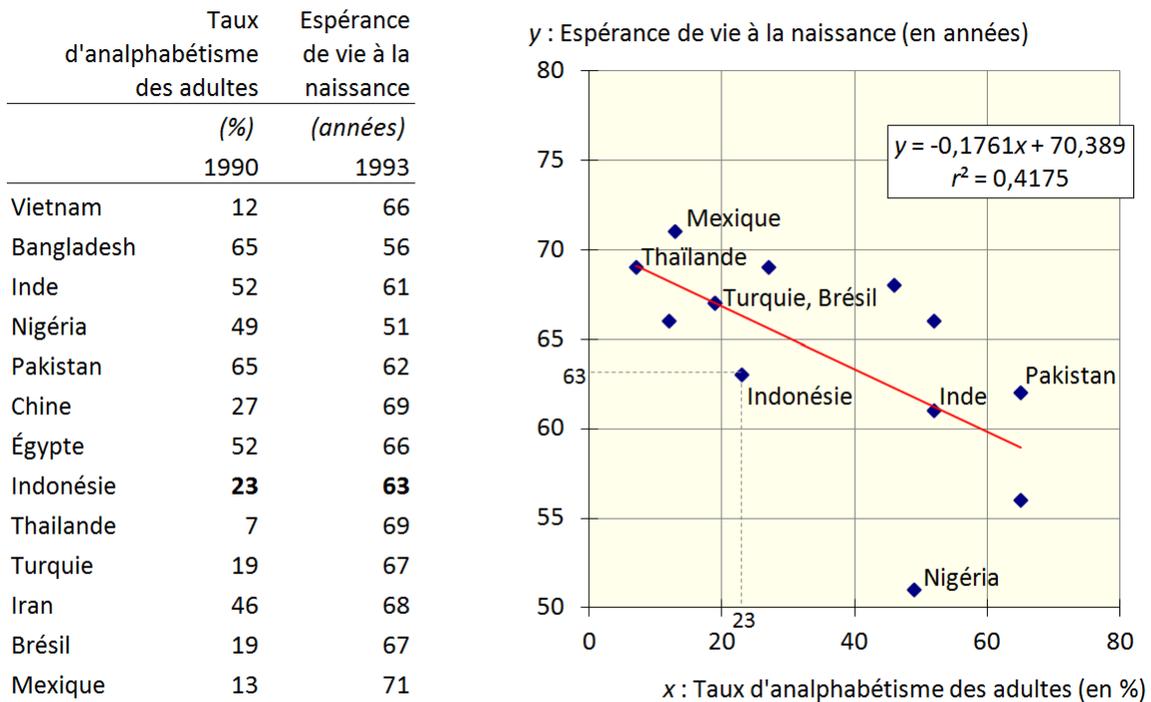
Le coefficient de corrélation mesure la force de la corrélation entre deux variables.

Plus généralement, on écrit l'équation de la droite sous la forme : $y = a + bx$, dans laquelle a est l'ordonnée à l'origine et b la pente. Une fois qu'on a déterminé la valeur de a et de b , on devrait être capable, connaissant x , de trouver y avec un certain degré de fiabilité (mesuré par le coefficient de corrélation). C'est ce que nous allons faire avec des données internationales.

3.2. Qui s'instruit... vit plus longtemps?

Nous avons choisi d'étudier la corrélation éventuelle entre l'éducation et la santé des gens. Plus précisément, nous avons retenu deux variables facilement observables à travers le monde : le taux d'analphabétisme des adultes et l'espérance de vie à la naissance. Pour ne pas encombrer le tableau, nous n'avons sélectionné que les gros pays (peuplés de 50 millions et plus) pour lesquels les données existent. Nous avons exclu de l'échantillon les pays industrialisés pour lesquels les taux officiels d'analphabétismes sont égaux à zéro. Grâce à ces simplifications, peut-être excessives, il nous sera plus facile de montrer comment construire et interpréter la droite de régression. En fin de compte, nous avons retenu 13 pays et nous avons représenté les données sous forme d'un tableau accompagné d'un graphique (voir figure 9.10).

FIGURE 9.10 - Analphabétisme et espérance de vie dans le monde



Source : Banque mondiale, *Rapport sur le développement dans le monde* 1995.

Note : Nous avons sélectionné tous les pays à revenu faible ou intermédiaire dont la population dépassait 50 millions d'habitants et pour lesquels les données étaient disponibles.

Chacun des points du graphique représente un pays. Le point correspondant à l'Indonésie, par exemple, a une abscisse de 23 (par rapport à l'axe horizontal) et une ordonnée de 63 (par rapport à l'axe vertical). On ne voit que 12 points sur le graphique, parce que deux pays, le Brésil et la Turquie, possèdent exactement les mêmes valeurs. Étant donné que le graphique a été tracé à l'aide d'un chiffrier électronique, il n'a pas été difficile, en sélectionnant la bonne option, d'y rajouter la droite de régression qui correspond à l'ensemble des points. Le chiffrier a même eu la bonté de nous fournir l'équation de la droite de régression (sous la forme $y = bx + a$), ainsi qu'un coefficient que nous interpréterons un peu plus loin. Certes, les points du graphique sont loin d'être parfaitement alignés, mais la droite donne quand même une tendance générale. On peut affirmer, d'emblée, qu'il existe une certaine corrélation entre l'analphabétisme et l'espérance de vie, et que ces deux variables évoluent en sens inverse.

3.3. Le coefficient de corrélation

Comme nous l'avons indiqué un peu plus haut, le coefficient de corrélation mesure la force de la relation entre les deux variables. Mais avant d'interpréter ce coefficient, il faut le calculer. Pour ce faire, il existe deux méthodes : la méthode facile (en utilisant les fonctions intégrées d'un chiffrier électronique) et la méthode à papa (en se tapant une série de calculs, simples mais laborieux). Nous avons déjà fait connaissance avec la méthode facile, puisque nous l'avons utilisée pour tracer la droite de régression et obtenir gratuitement l'équation de la droite et un certain coefficient (r^2) qui n'est autre que le carré du coefficient de corrélation (revoir le graphe de la [figure 9.10](#)).

La méthode à papa.

Dans le tableau 9.5, nous indiquons toutes les étapes du calcul du coefficient de corrélation et des paramètres de la droite de régression et nous reproduisons les formules correspondantes. Si vous disposez d'un chiffrier électronique (c'est presque indispensable dès qu'on utilise des chiffres en sciences humaines), vous pouvez vous dispenser de cette étape fastidieuse.

| Tableau 9.5 - Calcul des paramètres de la droite de régression | | | | | |
|--|---------|--|---|--|--|
| | Calculs | | Vérifications avec les fonctions du chiffrier | | |
| b | -0,176 | | -0,176 | Droitereg(plage y; plage x) | |
| a | 70,389 | | 70,389 | Ordonnee.Origine(plage y; plage x) | |
| r | -0,646 | | -0,646 | Coefficient.Corrrelation(plage y; plage x) | |

| | x | y | xy | x ² | y ² |
|-----------------------------|-----------|-----------|-------|----------------|----------------|
| n (taille de l'échantillon) | 13 | | | | |
| Moyenne | 34,5 | 64,3 | | | |
| Écart type | 20,8 | 5,7 | | | |
| Somme | 449 | 836 | 27957 | 20717 | 54148 |
| Vietnam | 12 | 66 | 792 | 144 | 4356 |
| Bangladesh | 65 | 56 | 3640 | 4225 | 3136 |
| Inde | 52 | 61 | 3172 | 2704 | 3721 |
| Nigéria | 49 | 51 | 2499 | 2401 | 2601 |
| Pakistan | 65 | 62 | 4030 | 4225 | 3844 |
| Chine | 27 | 69 | 1863 | 729 | 4761 |
| Égypte | 52 | 66 | 3432 | 2704 | 4356 |
| Indonésie | 23 | 63 | 1449 | 529 | 3969 |
| Thaïlande | 7 | 69 | 483 | 49 | 4761 |
| Turquie | 19 | 67 | 1273 | 361 | 4489 |
| Iran | 46 | 68 | 3128 | 2116 | 4624 |
| Brésil | 19 | 67 | 1273 | 361 | 4489 |
| Mexique | 13 | 71 | 923 | 169 | 5041 |

Pente de la droite : $b = [(n \times \text{Somme des } xy) - (\text{Somme des } x \times \text{Somme des } y)] / [(n \times \text{Somme des } x^2) - (\text{Somme des } x)^2]$

Ordonnée à l'origine : $a = \text{Moyenne des } y - (b \times \text{Moyenne des } x)$

Coefficient de corrélation : $r = b \times (\text{Écart type de } x) / (\text{Écart type de } y)$

Plus le coefficient de corrélation est proche de 0, moins la corrélation est forte.

Le coefficient de corrélation est construit de telle sorte qu'il est égal à +1 ou -1 lorsque les points sont parfaitement alignés. Dans notre exemple, le coefficient de corrélation est égal à -0,646. Le signe négatif indique que les deux variables évoluent en sens inverse. La valeur absolue du coefficient (0,646) semble relativement élevée, mais pour l'interpréter il est nécessaire de tenir compte de la taille de l'échantillon. On pourra alors, moyennant certains calculs supplémentaires, faire une hypothèse sur l'existence d'une association entre les deux variables et la tester avec la table de distribution de Student. Si vous tenez vraiment à savoir comment, il vous faudra consulter un ouvrage spécialisé.

EXERCICES 3

1. Échalote, Bouboule, Brummel et les autres

Vous devez évaluer s'il existe une corrélation entre le *poids* et la *taille* à l'aide d'un échantillon d'au moins 30 individus. (Pour les besoins de la cause, il est acceptable que l'échantillon ne soit pas tiré au hasard.) Tracez le nuage de points et calculez le coefficient de corrélation (à la main ou en utilisant le chiffrier électronique, calcul fourni)

2. Étranges corrélations

Les corrélations présentées ci-après paraissent pour le moins étranges. À vous de leur donner une explication logique et identifiant convenablement les variables impliquées et les relations qui les unissent.

- a) On a observé une corrélation entre homicides et pointures des souliers. Plus précisément, il semble que les homicides soient plus fréquemment commis par des individus possédant des pieds plus grands que la moyenne.
- b) On a déjà constaté, en Californie, une corrélation entre les ventes de bière et le taux de mortalité chez les personnes âgées et les bambins.

4. LES ACCIDENTS DE LA ROUTE

Les jeunes sont-ils susceptibles de provoquer plus d'accidents que leurs aînés? Les femmes conduisent-elles mieux que les hommes? Est-il plus dangereux de rouler en Chine qu'aux États-Unis? Nous essaierons de répondre, partiellement, à toutes ces questions en nous servant de tous les outils vus dans ce chapitre.

4.1. Un bref tour du monde

Examiner quelques données brutes pour découvrir le schéma de variables.

Commençons par déblayer le terrain en examinant quelques données brutes publiées par le gouvernement japonais dans les années 1990. Nous reproduisons dans le tableau 9.6 le nombre d'accidents de la route et le nombre de victimes de ces accidents pour six pays. Si les États-Unis se classent premiers, devant le Japon, au chapitre des accidents et du nombre de blessés, c'est la Chine qui détient le triste record du nombre de tués.

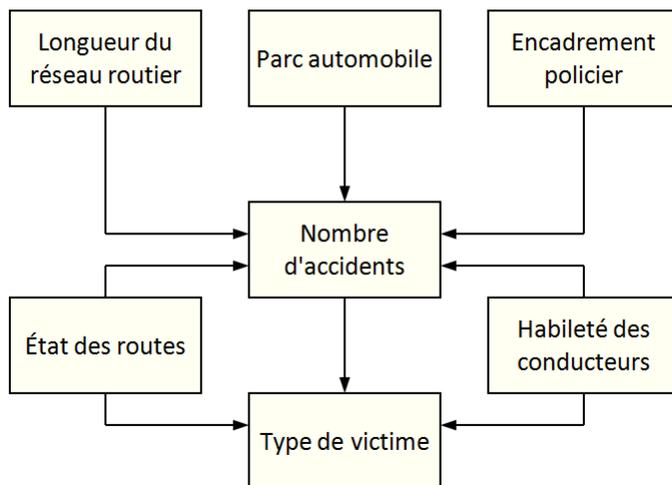
| | Accidents | Blessés | Tués | (année) |
|------------|-----------|-----------|--------|---------|
| Chine | 228 326 | 144 021 | 58 723 | 1992 |
| États-Unis | 2 540 946 | 3 600 307 | 44 529 | 1990 |
| Japon | 662 388 | 810 245 | 11 105 | 1991 |
| France | 148 890 | 205 968 | 9 617 | 1991 |
| Mexique | 18 337 | 43 000 | 5 700 | 1991 |
| Canada | 182 294 | 261 139 | 3 957 | 1990 |
| Québec | | | 2 209 | 1973 |
| Québec | 304 160 | 45 155 | 845 | 1995 |
| Québec | 104 070 | 39 105 | 436 | 2012 |

Source : Japan international statistical survey, 1994. SAAQ, Bilan routier 2012.

Le tableau 9.6 contient également quelques données concernant le Québec et couvrant quatre décennies. On y constate que le nombre d'accidents et de victimes tend à diminuer considérablement avec le temps, malgré l'augmentation du nombre d'automobiles sur les routes.

Ce premier contact chiffré avec la situation étudiée nous permet de prendre conscience de la complexité du problème. Il est clair que l'habileté des conducteurs ou l'état des routes ne sont pas les seules variables qui peuvent exercer une influence sur le nombre d'accidents. Aux États-Unis, il y a plus de kilomètres de routes qu'au Japon (pays plus petit) et plus de véhicules qu'au Canada (pays moins peuplé) et qu'au Mexique (pays moins riche). La figure 9.11 illustre une façon de relier ces différentes variables ainsi que d'autres.

FIGURE 9.11 - Les accidents de la route



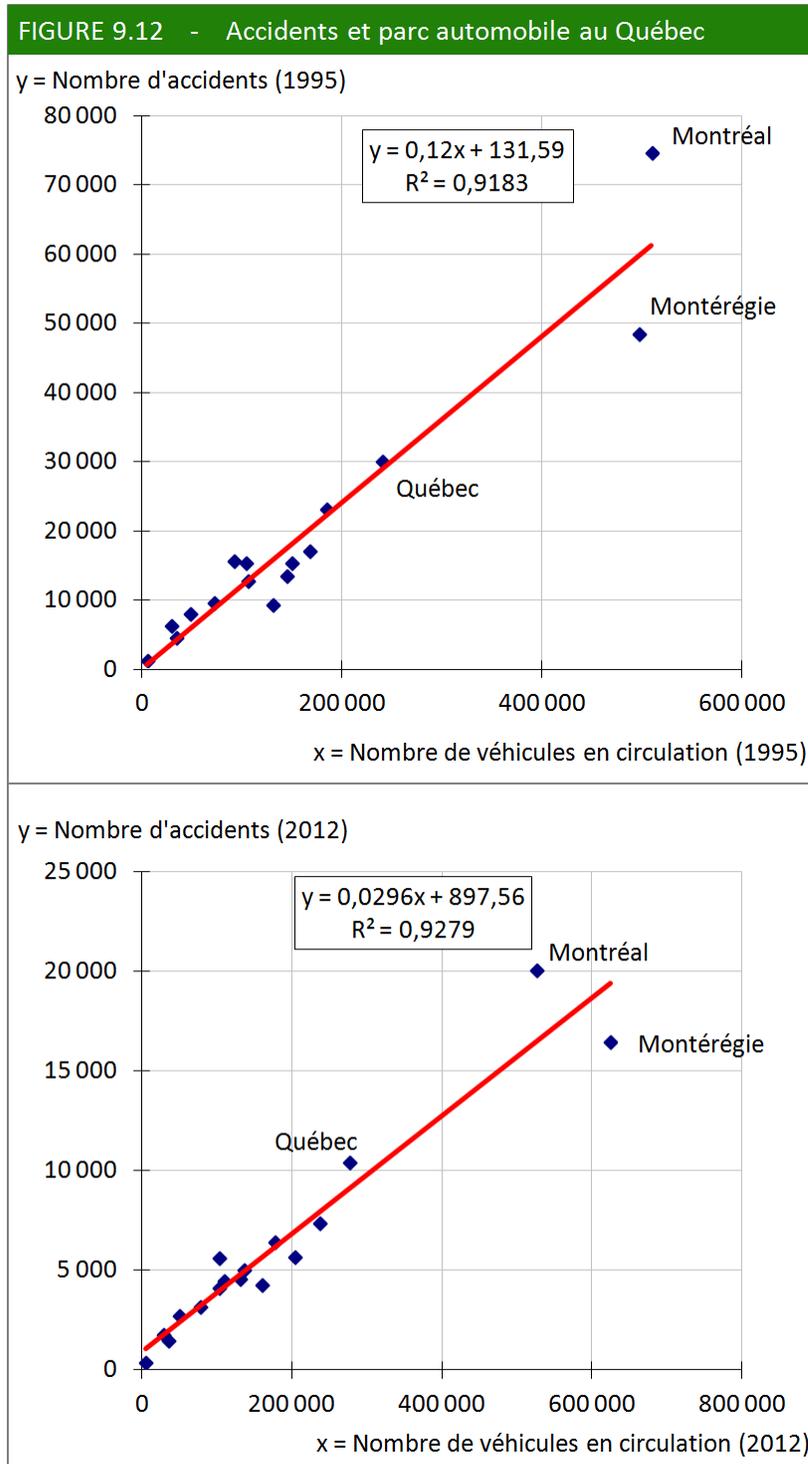
Il est plus difficile de comparer des situations hétérogènes.

Dans le [tableau 9.6](#), avez-vous remarqué la situation anormale de la Chine qui compte 25 fois moins de blessés mais plus de tués que les États-Unis (dans ce dernier cas, s’agit-il de piétons?). Cela pourrait bien être attribuable en bonne partie à la façon dont chaque pays définit chaque variable : si tout le monde s’entend sur ce qu’est un tué, il en va autrement pour les notions d’accident et de blessé. Nous éviterons ce genre d’écueil en restreignant maintenant notre étude au Québec.

4.2. La situation dans les régions du Québec

Dans la figure 9.12, nous avons choisi de relier la variable *nombre de véhicules* à la variable *nombre d’accidents* en examinant la situation dans les diverses régions du Québec. La figure est accompagnée d’une droite de régression, entièrement construite avec un chiffrier électronique et sans aucun calcul. La relation entre les deux variables saute aux yeux. Le coefficient de corrélation (r) est très proche de 1, ce qui signifie que la corrélation est très forte. Il reste néanmoins que le nombre de véhicules n’explique pas entièrement le nombre d’accidents. Il vaudrait la peine de chercher à isoler les autres variables.

Il existe une explication relativement évidente à la situation de Montréal et de la Montérégie, dont les points s’écartent étrangement — et symétriquement — de la droite de régression. De nombreux conducteurs de la Montérégie se rendent chaque jour à Montréal. En même temps que les automobilistes, les risques d’accident se déplacent alors d’une région à l’autre. Le même phénomène entre les villes-centres et leur grande banlieue se remarque, à moins grande échelle, dans la région de Québec en 2012.



(Données de la figure 9.12 - Accidents et parc automobile au Québec)

| | Nombre de véhicules de promenade en circulation | | Nombre total d'accidents | | |
|-------------------------------|---|------------------|--------------------------|----------------|-------------------------------|
| | 1995 | 2012 | 1995 | 2012 | |
| Gaspésie-Îles de la Madeleine | 33 875 | 36 424 | 4 454 | 1 494 | Gaspésie-Îles de la Madeleine |
| Bas-Saint-Laurent | 72 234 | 78 641 | 9 542 | 3 195 | Bas-Saint-Laurent |
| Saguenay-Lac-Saint-Jean | 92 473 | 103 640 | 15 608 | 5 618 | Saguenay-Lac-Saint-Jean |
| Québec | 240 644 | 277 558 | 29 957 | 10 413 | Capitale-Nationale |
| Chaudière-Appalaches | 149 742 | 178 067 | 15 343 | 6 421 | Chaudière-Appalaches |
| Mauricie-Bois-Francs | 184 809 | 110 631 | 23 091 | 4 451 | Mauricie |
| | | 103 647 | | 4 112 | Centre-du-Québec |
| Estrie | 106 090 | 131 180 | 12 759 | 4 561 | Estrie |
| Montérégie | 497 115 | 624 682 | 48 351 | 16 451 | Montérégie |
| Montréal | 509 606 | 526 548 | 74 647 | 20 090 | Montréal |
| Laval | 130 761 | 160 481 | 9 250 | 4 277 | Laval |
| Lanaudières | 145 225 | 205 169 | 13 473 | 5 658 | Lanaudières |
| Laurentides | 167 435 | 237 580 | 17 088 | 7 371 | Laurentides |
| Outaouais | 104 441 | 137 078 | 15 313 | 5 049 | Outaouais |
| Abitibi-Témiscamingue | 48 260 | 50 272 | 7 886 | 2 725 | Abitibi-Témiscamingue |
| Côte-Nord | 29 082 | 29 339 | 6 162 | 1 791 | Côte-Nord |
| Nord du Québec | 4 995 | 4 912 | 1 236 | 393 | Nord du Québec |
| Non précisé | 928 | 2 641 | | | Non précisé |
| Total | 2 517 715 | 2 998 490 | 304 160 | 104 070 | Total |

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 1995, Bilan statistique 2012.

Dans les *Cahiers de géographie du Québec* d'avril 1996, des chercheurs ont utilisé des variables plus raffinées pour étudier la situation : la densité d'accidents (nombre d'accidents au km²), le taux de motorisation (nombre de véhicules de promenade par habitant), la densité de la population (nombre d'habitants par km²) et l'encadrement policier (nombre de policiers pour 1000 habitants). On remarque que ces quatre variables sont en réalité des rapports, ce qui permet de comparer des régions dont les caractéristiques sont différentes. Les chercheurs ont alors calculé la valeur de ces variables pour chacune des municipalités régionales de comté du Québec et ont ensuite cherché à établir des corrélations.

Cette étude montre qu'au Québec, la densité d'accidents est directement reliée à la densité de la population. L'équation de régression qui relie ces deux variables est la suivante :

$$\text{Densité d'accidents} = 0,0084 + 0,0053 \text{ Habitants/km}^2$$

Selon cette équation, une région qui compterait 10 habitants/km² aurait une densité annuelle d'accidents de $0,0084 + (10 \times 0,0053) = 0,0084 + 0,053 = 0,0614$ accident/km². Le coefficient de corrélation est de 0,99, ce qui indique que les deux variables se suivent de très près. Lorsque l'on croise les variables *densité d'accident* et *encadrement policier*, la corrélation est beaucoup plus faible

($r = -0,37$) et les variables évoluent en sens inverse. Même si la présence policière semble calmer les chauffards, son influence semble relativement faible par rapport à d'autres variables. Notons enfin que des calculs similaires effectués dans les 52 districts de gendarmerie de la Belgique donnent des résultats très semblables. Ce genre d'étude peut être particulièrement utile lorsqu'il s'agit de mettre en place une politique de sécurité routière.

4.3. Qui conduit le mieux?

Nous nous demandons un peu plus haut si le sexe ou l'âge peuvent exercer une influence sur le nombre d'accidents. Le tableau 9.7 montre qu'il y a *relativement* moins de femmes que d'hommes qui se trouvent impliqués dans des accidents au Québec. Mais ici encore, il faut être prudent avant de conclure que les femmes conduisent mieux (ou se soûlent moins) que les hommes. On devrait se poser les questions suivantes : lequel des deux sexes fait le plus de kilomètres dans l'année? Lequel roule le plus la nuit? Lequel fréquente le plus souvent des routes de campagne? Etc.?

| Tableau 9.7 - Les accidents selon le sexe du conducteur au Québec | | | |
|--|-----------------|-----------------------|------------------|
| (1995) | Accident | Pas d'accident | Total |
| Femmes | 77 502 | 1 897 904 | 1 975 406 |
| Hommes | 180 136 | 2 162 022 | 2 342 158 |
| Total | 257 638 | 4 059 926 | 4 317 564 |
| Khi ² | 27 110 | | |
| V de Cramer | 0,079 | | |

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 1995.

| (2012) | Accident | Pas d'accident | Total |
|------------------|-----------------|-----------------------|------------------|
| Femmes | 45 351 | 2 441 185 | 2 486 536 |
| Hommes | 72 003 | 2 636 221 | 2 708 224 |
| Total | 117 354 | 5 077 406 | 5 194 760 |
| Khi ² | 4 092 | | |
| V de Cramer | 0,028 | | |

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 2012.

Dans le tableau 9.7, on constate à nouveau que le nombre d'accidents a tendance à baisser fortement au Québec sur le long terme, encore plus pour les hommes que pour les femmes. Cela dit, la très grande majorité des conducteurs n'ont pas eu d'accident au cours de l'année considérée.

Dans le tableau 9.8, on constate que les jeunes conducteurs ont *relativement* plus d'accidents que leurs aînés. Mais là encore, le phénomène peut être influencé par plusieurs variables : l'expérience du conducteur, son attitude, l'état des routes qu'il fréquente particulièrement et l'âge du véhicule.

Tableau 9.8 - Les accidents selon l'âge du conducteur au Québec

| (1995) | Accident | Pas d'accident | Total |
|------------------|-----------------|-----------------------|------------------|
| 16-19 ans | 23 346 | 150 221 | 173 567 |
| 20-24 ans | 32 903 | 312 172 | 345 075 |
| 25-35 ans | 64 626 | 903 727 | 968 353 |
| Total | 120 875 | 1 366 120 | 1 486 995 |
| Khi ² | 10 241 | | |
| V de Cramer | 0,083 | | |

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 1995.

| (2012) | Accident | Pas d'accident | Total |
|------------------|-----------------|-----------------------|------------------|
| 16-19 ans | 12 207 | 140 845 | 153 052 |
| 20-24 ans | 18 826 | 344 110 | 362 936 |
| 25-34 ans | 27 067 | 810 923 | 837 990 |
| Total | 58 100 | 1 295 878 | 1 353 978 |
| Khi ² | 8 066 | | |
| V de Cramer | 0,077 | | |

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 2012.

Pour les deux tableaux précédents, nous avons calculé le Khi carré et le V de Cramer. Étant donné que l'échantillon est très grand (il s'agit en réalité de la population au complet), le Khi carré est très élevé. Nos données sont, par définition, représentatives, aussi il est inutile de faire un test d'hypothèse. Le Khi carré nous sert uniquement à calculer la force de l'association grâce au V de Cramer. On constate, dans le tableau 9.8 que le V de Cramer est de 0,083 en 1995. Cela signifie que si l'âge explique une partie du problème, il est très loin de l'expliquer dans son entier. Bien que faible, le coefficient de Cramer s'avère intéressant si on cherche à comparer la situation de 1995 à celle de 1991 ou de 2012, par exemple. Le V de Cramer était alors de 0,077, ce qui prouve qu'il existe une certaine stabilité dans la relation.

4.4. Des calculs plus simples en disent parfois plus long

Pour tirer les choses au clair, mettons de côté les coefficients compliqués et servons-nous de simples rapports. Selon une enquête concernant la région de l'Outaouais (1995), les hommes au volant sont impliqués 3,5 fois plus souvent que les femmes dans des accidents mortels. Ce rapport baisse à 1,8 pour les accidents avec blessures légères. Par ailleurs, les conducteurs parcourent deux fois plus de distance que les conductrices. Cela démontre qu'à distance égale parcourue, les conducteurs masculins ont plus d'accidents que les femmes, du moins en ce qui concerne les accidents graves.

Pour mieux cerner la relation entre l'âge et la fréquence des accidents, nous avons calculé le nombre d'accidents par rapport au nombre de titulaires de permis de conduire pour différents groupes d'âge. Les chiffres, que nous reproduisons dans le tableau 9.9, parlent d'eux-mêmes (troisième colonne du tableau).

Tableau 9.9 - Rapport entre le nombre d'accidents et le nombre de titulaires de permis selon l'âge

| (1995) | Nombre de conducteurs impliqués dans un accident | Nombre de détenteurs de permis de conduire | Rapport Accidents/ Détenteurs de permis |
|----------------|---|---|--|
| 16-19 ans | 23 346 | 173 567 | 0,135 |
| 20-24 ans | 32 903 | 345 075 | 0,095 |
| 25-35 ans | 64 626 | 968 353 | 0,067 |
| 35-44 ans | 58 534 | 1 088 868 | 0,054 |
| 45-54 ans | 40 853 | 830 264 | 0,049 |
| 55-64 ans | 21 591 | 499 919 | 0,043 |
| 65 ans et plus | 17 750 | 403 351 | 0,044 |

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 1995.

| (2012) | Nombre de conducteurs impliqués dans un accident | Nombre de détenteurs de permis de conduire | Rapport Accidents/ Détenteurs de permis |
|----------------|---|---|--|
| 16-19 ans | 12 207 | 153 052 | 0,080 |
| 20-24 ans | 18 826 | 362 936 | 0,052 |
| 25-34 ans | 27 067 | 837 990 | 0,032 |
| 35-44 ans | 23 709 | 901 271 | 0,026 |
| 45-54 ans | 24 532 | 1 108 089 | 0,022 |
| 55-64 ans | 17 856 | 960 971 | 0,019 |
| 65 ans et plus | 15 948 | 863 530 | 0,018 |

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 2012.

EXERCICES 4

1. Mon camion c'est ma maison

Commentez le tableau 9.10.

| Tableau 9.10 - Types de véhicules dans la région métropolitaine | | | |
|---|------------------|----------------|------------------|
| (1995) | Automobiles | Camions légers | Total |
| Montréal | 497 115 | 107 671 | 604 786 |
| Montréal | 509 606 | 71 287 | 580 893 |
| Total | 1 006 721 | 178 958 | 1 185 679 |

Source : Société de l'Assurance automobile du Québec, Bilan routier 1995.

Khi²: 7074

V de Cramer: 0,077

| (2012) | Automobiles | Camions légers | Total |
|--------------|------------------|----------------|------------------|
| Montréal | 624 682 | 255 111 | 879 793 |
| Montréal | 532 392 | 184 394 | 716 786 |
| Total | 1 157 074 | 439 505 | 1 596 579 |

Source : Société de l'Assurance automobile du Québec, Bilan routier 2012.

2. Jeune et fou

a) Complétez le tableau 9.11.

Tableau 9.11 - Rapport entre le nombre d'accident et le nombre de titulaires de permis selon l'âge

| (1991) | Nombre de conducteurs impliqués dans un accident | Nombre de détenteurs de permis de conduire | Rapport Accidents/ Détenteurs de permis |
|----------------|--|--|---|
| 16-19 ans | 25 246 | 185 762 | |
| 20-24 ans | 38 884 | 371 183 | |
| 25-34 ans | 78 367 | 1 073 556 | |
| 35-44 ans | 59 296 | 1 001 300 | |
| 45-54 ans | 36 738 | 681 342 | |
| 55-64 ans | 21 272 | 456 939 | |
| 65 ans et plus | 15 052 | 314 768 | |

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 1995.

b) Tracez sur un graphe la courbe du rapport Accidents/Détenteurs de permis en fonction de l'âge. Tracez une deuxième courbe sur le même graphe en utilisant les données de 1995 (voir [tableau 9.9](#)).

c) Comparez la situation de 1991 à celle de 1995 et commentez.

3. Recherche

Mettez à jour les tableaux 9.7, 9.8 et 9.9, et commentez l'évolution des données.

EXERCICES SUPPLÉMENTAIRES

1. Des relations à vérifier

La pauvreté est en hausse aux États-Unis dans les années 1990. Une idée très répandue dans les milieux conservateurs veut que l'augmentation des dépenses de l'aide sociale soit causée par la prolifération des mères adolescentes et des mères célibataires. Dans d'autres milieux, soi-disant progressistes, on affirme souvent que l'instruction n'est plus un moyen efficace pour éviter la pauvreté. Avant de commenter ces affirmations, répondez aux questions *a* et *b* ci-après.

a) Représentez graphiquement les données suivantes : le salaire mensuel moyen est de 508 \$ pour ceux qui n'ont pas terminé leurs études secondaires, de 1080 \$ pour ceux qui détiennent un diplôme du secondaire, de 1303 \$ pour ceux qui ont entamé des études postsecondaires et de 2339 \$ pour ceux qui ont terminé des études universitaires.

b) Représentez graphiquement les données suivantes : l'âge médian des mères qui bénéficient de l'aide sociale est de 27,4 ans, 24,5 % d'entre elles ont moins de 21 ans et 7,1 % d'entre elles ont moins de 18 ans.

(Source des données : Scientific American, octobre 1996, données de 1993 et 1995.)

2. Le suicide selon Durkheim

Dans un ouvrage classique intitulé *Le suicide*, Émile Durkheim constate que, si chaque suicide est un phénomène individuel, le taux de suicide d'une population s'avère facilement prévisible à court terme. Entre 1871 et 1875, le taux de suicide (nombre annuel de suicides pour 100 000 habitants) était de 25,5 au Danemark et de 3,5 en Italie. En Suisse, en 1876, il variait de 8,3 à 8,7 chez les catholiques et de 29,3 à 45,6 chez les protestants.

En France, entre 1835 et 1843, ce taux était de 30,6 en été, de 28,3 au printemps, de 21,0 en automne et de 20,1 en hiver. Entre 1848 et 1857, il s'élevait à 4,6 pour les gens de 16 à 21 ans, à 9,8 pour ceux de 21 à 30 ans, à 11,5 pour ceux de 31 à 40 ans et à 16,4 pour ceux de 41 à 50 ans. Entre 1889 et 1891, ce taux de suicide chez les hommes mariés âgés de 26 à 35 ans était de 10,6 contre 25,7 chez les hommes célibataires. Chez les femmes mariées du même groupe d'âge, ces taux étaient respectivement de 2,8 et de 6,1 chez les célibataires. Durkheim note d'autre part que le taux est plus élevé dans les grandes villes que dans les petites, plus élevé le jour que la nuit, plus élevé au début de la semaine qu'à la fin.

Par ailleurs, des données relevées au XX^e siècle indiquent que le taux de suicide était au plus bas en 1917 (14,0) et en 1943 (11,3) alors qu'il atteignait un sommet en 1912 (22,9) et en 1934 (21,7).

Au Québec, le taux de suicide passait de 4,2 en 1950-54 à 13,8 en 1975-79. En 1995, le taux de suicide était de 58 en Hongrie, de 30,0 en France, de 34,0 en Suisse, de 12,0 en Espagne et de 20,4 aux États-Unis. Du côté des personnes de 75 ans et plus, le taux de suicide se révèle 12 fois plus élevé pour les hommes que pour les femmes pour atteindre 186,2 en Hongrie et 114,0 en France.

Voici pour terminer quelques comparaisons entre pays pour la période 2001-2010. Le taux de suicide du Canada était de 17,3 contre 17,7 aux États-Unis. Celui du Japon était de 36,2 contre 13,0 en Inde. Celui de Cuba était de 19,0 contre 3,9 en République dominicaine, celui de la Russie était de 53,9 contre 37,8 en Ukraine.

(Sources : Émile Durkheim, *Le suicide*, 1897, réédition 1960 : Presses universitaires de France. Organisation mondiale de la Santé. Marie-France Charron, *Le suicide au Québec*, 1983. PNUD, *Rapport sur le développement humain 2014*.)

- a) Construisez un schéma contenant toutes les variables qui influent sur le taux de suicide. Indiquez par des flèches le sens des relations. Lorsque cela est pertinent, accompagnez la flèche d'un signe positif ou négatif pour indiquer si les variables évoluent dans le même sens ou en sens inverse.
- b) Recherche. Obtenez d'autres données dans le but de confirmer, de rectifier ou de compléter le schéma de variables que vous venez de tracer.

3. Ex-maltraité, future brute?

Joan Kaufman et Edward Zigler, de l'université de Yale, ont essayé d'établir si la violence parentale envers les enfants se transmettait de génération en génération. Pour ce faire, ils ont utilisé les résultats d'une enquête menée auprès de 282 parents d'enfants admis dans un service de soins intensifs pour enfants. Ils constatèrent que 49 de ces parents avaient été maltraités durant leur enfance et que 10 de leurs enfants furent maltraités dans l'année qui suivit leur visite à l'hôpital. Parmi ces 10 enfants, 9 étaient issus de parents qui avaient été maltraités eux-mêmes. (Source : *Sciences humaines*, no 65, octobre 1996.)

- a) Construisez, à partir des données ci-dessus, un tableau croisé de deux colonnes et de deux lignes.
- b) Faites un test pour vérifier l'hypothèse qu'il existe une relation entre la façon dont les parents ont été traités dans leur enfance et la façon dont ils traitent leurs enfants.
- c) Commentez l'affirmation suivante : « 90 % des enfants maltraités sont issus de parents eux-mêmes maltraités dans leur enfance, mais seulement 18 % des ex-enfants maltraités deviennent des parents maltraitants ».

4. Deux variables qui se suivent de près

a) Représentez les points du tableau 9.12 sur un graphe.

Tableau 9.12 - Accidents et parc automobile au Québec

| (1991) | Nombre de véhicules de promenade en circulation | Nombre total d'accidents |
|-------------------------------|--|-----------------------------|
| Gaspésie-Îles de la Madeleine | 31 865 | 4 276 |
| Bas-Saint-Laurent | 69 853 | 9 022 |
| Saguenay-Lac-Saint-Jean | 91 751 | 15 647 |
| Québec | 237 672 | 31 446 |
| Chaudière-Appalaches | 144 177 | 14 377 |
| Mauricie-Bois-Francs | 182 066 | 22 567 |
| Estrie | 102 461 | 12 767 |
| Montérégie | 488 436 | 48 390 |
| Montréal | 534 938 | 77 850 |
| Laval | 130 294 | 8 938 |
| Lanaudières | 136 142 | 13 401 |
| Laurentides | 155 532 | 16 661 |
| Outaouais | 100 622 | 16 438 |
| Abitibi-Témiscamingue | 46 797 | 7 555 |
| Côte-Nord | 28 348 | 6 110 |
| Nord du Québec | 5 282 | 1 274 |
| Non précisé | 902 | |
| Total | 2 487 138 | 306 719 |

Source des données brutes: Société de l'Assurance automobile du Québec, bilan 1995.

b) Commentez le graphe.

c) Si vous avez un chiffrier électronique, essayez de tracer une droite de régression sur le même modèle que la [figure 9.12](#).

5. Un sport dangereux?

Quelque 223 hockeyeurs qui ont été admis à l'urgence de l'hôpital de l'Enfant Jésus de Québec, entre le 1^{er} octobre 1991 et le 30 avril 1992. Le tableau 9.13 indique comment se répartissaient les blessures selon le type de jeu pratiqué et la partie du corps touchée.

Tableau 9.13 - Les blessures au hockey

| Type de jeu | Type de blessure | | | | Total |
|---|------------------|------------------|------------------|-----------|------------|
| | Tête | Membre supérieur | Membre inférieur | Tronc | |
| Hockey organisé (ligue, club) avec patins | 20 | 42 | 48 | 20 | 130 |
| Hockey non organisé avec patins | 11 | 10 | 8 | 6 | 35 |
| Hockey organisé (ligue, club) sans patins | 11 | 5 | 14 | 3 | 33 |
| Hockey non organisé sans patins | 11 | 5 | 6 | 3 | 25 |
| Total | 53 | 62 | 76 | 32 | 223 |

Source : Revue canadienne de Santé publique, volume 87, n^o 4, juillet-août 1996.

Essayez d'établir s'il existe une relation entre le type de hockey pratiqué et le type de blessure reçue. Si nécessaire, réduisez le tableau à deux lignes, en regroupant certaines catégories.

6. Recherche : un tableau croisé

Faites une mini enquête qui vous permettra de construire un tableau croisé. Choisissez deux variables qualitatives comportant chacune de deux à quatre catégories. Constituez un échantillon (aléatoire dans la mesure du possible) et compilez vos résultats dans un tableau croisé. Si nécessaire, augmentez la taille de votre échantillon pour éviter que certaines cases du tableau ne contiennent des fréquences inférieures à 5. Calculez le Khi carré et le V de Cramer et commentez les résultats.

7. Recherche : l'héritabilité du poids

Dans cette mini enquête, vous chercherez à mesurer la corrélation entre le poids (ou la taille) des pères et de leurs fils, ou des mères et de leurs filles.

Prenez une balance (ou un mètre) et promenez-vous dans votre quartier. Pesez les paires (parent-enfant) tout en ménageant leur susceptibilité. Notez leur âge et vérifiez si vous êtes en présence d'un parent biologique ou d'un parent adoptif. Pour éliminer l'influence de la variable *âge*, prenez uniquement des enfants adultes ou, à la rigueur, des enfants d'un âge précis.

Tracez un nuage de points. Tracez la courbe de tendance. Calculez le coefficient de corrélation.

Que proposez-vous pour améliorer l'échantillon choisi?

DOSSIER 9 UN TABOU UNIVERSEL

L'inceste, ou union de deux êtres du même sang, est un des derniers tabous de notre société « post-moderne ». C'est aussi un tabou universel, puisqu'on le retrouve pratiquement dans toutes les sociétés depuis la nuit des temps. Dans ce domaine, comme c'est souvent le cas en sciences humaines, les partisans de *l'inné* s'opposent aux partisans de *l'acquis*. Pour les premiers, l'horreur de l'inceste est inscrite dans les gènes des humains, comme dans celui d'autres animaux. Pour les seconds, au contraire, l'obligation du mariage à l'extérieur du clan vient du besoin de maintenir la paix (en créant des liens familiaux avec des ennemis potentiels) et de préserver la santé de l'espèce. Quoi qu'il en soit, la nature, dans sa bienveillance, ou la société, dans sa sagesse, auraient, en créant ce tabou, évité à la race humaine de dégénérer. Car il est largement accepté que la consanguinité est source de déficiences physiques ou mentales.

L'étude de Seemenova, effectuée en ex-Tchécoslovaquie auprès de 256 enfants tend à confirmer cette croyance, du moins à première vue. Mais si on prend la peine d'approfondir un peu la question, on se rend compte que la relation de cause à effet entre consanguinité et dégénérescence est loin d'être évidente.

Selon l'étude de Seemenova, la *proportion* d'enfants ayant des troubles physiques ou mentaux est considérablement plus élevée chez les enfants nés d'un inceste (32/161) que chez les autres enfants (2/95). La même remarque s'applique à la mortalité infantile (voir le tableau D9.1).

Tableau D9.1 - Proximité génétique et déficiences physiques ou mentales

| Échantillon | Enfants nés de l'inceste | Groupe témoin |
|----------------------|--------------------------|---------------|
| Femme avec son père | 88 | |
| Femme avec son frère | 72 | |
| Femme avec son fils | 1 | |
| Total | 161 | 95 |

| Pathologie | Nombre d'enfants ayant des troubles physiques ou mentaux | Nombre d'enfants morts dans leur première année |
|--------------------------|--|---|
| Parents incestueux | 32 | 15 |
| Parents du groupe témoin | 2 | 5 |

Source: Recherche effectuée en Tchécoslovaquie par Seemenova, dans in Jacques-Dominique de LANNOY et Pierre FEYERIESEN, *L'inceste*, PUF, 1992.

Il y a donc une relation apparente entre la consanguinité des parents et les déficiences des enfants. Mais quelles sont les chances que les différences observées entre les deux groupes d'enfants soient dues au hasard de l'échantillonnage? C'est ce que nous allons vérifier à l'aide du calcul du Khi carré.

Calcul des fréquences théoriques

Compte tenu de l'échantillon choisi, le tableau D9.2 montre que les différences entre les fréquences observées et les fréquences théoriques sont relativement faibles (voir tableaux D9.2a et D9.2b). Nous

avons calculé les fréquences théoriques de la façon suivante : s'il n'y avait aucune association entre les variables, 20/256 des 161 enfants nés de parents incestueux devraient être morts dans leur première année et 236/256 de ces 161 enfants devraient avoir survécu. Le chiffre figurant dans la case supérieure gauche du tableau D9.2b est donc de $20/256 \times 161 = 12,6$.

TABLEAU D9.2 - Consanguinité et déficiences

Tableau D9.2a - Fréquences observées

| | Enfants morts dans leur première année | Autres enfants | Total |
|--------------------|--|----------------|------------|
| Parents incestueux | 15 | 146 | 161 |
| Autres parents | 5 | 90 | 95 |
| Total | 20 | 236 | 256 |

Tableau D9.2b - Fréquences théoriques

| | Enfants morts dans leur première année | Autres enfants | Total |
|--------------------|--|----------------|--------------|
| Parents incestueux | 12,6 | 148,4 | 161,0 |
| Autres parents | 7,4 | 87,6 | 95,0 |
| Total | 20,0 | 236,0 | 256,0 |

Tableau D9.2c - Écarts au carré relatifs

| | Enfants morts dans leur première année | Autres enfants | Total |
|--------------------|--|----------------|-------------|
| Parents incestueux | 0,47 | 0,04 | 0,51 |
| Autres parents | 0,79 | 0,07 | 0,86 |
| Total | 1,26 | 0,11 | 1,36 |

Tableau D9.2d - Proportion observées (en %)

| | Enfants morts dans leur première année | Autres enfants | Total |
|--------------------|--|----------------|--------------|
| Parents incestueux | 5,9 | 57,0 | 62,9 |
| Autres parents | 2,0 | 35,2 | 37,1 |
| Total | 7,8 | 92,2 | 100,0 |

Tableau D9.2e - Proportion théoriques (en %)

| | Enfants morts dans leur première année | Autres enfants | Total |
|--------------------|--|----------------|--------------|
| Parents incestueux | 4,9 | 58,0 | 62,9 |
| Autres parents | 2,9 | 34,2 | 37,1 |
| Total | 7,8 | 92,2 | 100,0 |

| | |
|-------|---|
| 1 | Degrés de liberté |
| 0,05 | Seuil de signification |
| 3,841 | Valeur critique du Khi^2 |
| 0,243 | Probabilité de dépasser le Khi^2 calculé |

Note : Les chiffres affichés sont arrondis, mais les calculs sont effectués avec un degré de précision maximum.

Calcul de l'écart global : le Khi carré

Nous avons ensuite calculé les écarts au carré relatifs entre les fréquences théoriques et observées. Pour la case supérieure gauche, on obtient $(12,6 - 15)^2/12,6 = 2,42/12,6 = 5,76/12,6 = 0,47$. Pour l'ensemble du tableau, ces écarts atteignent un total de 1,36 : il s'agit du Khi carré calculé.

Interprétation du Khi carré calculé

S'il n'existait aucune association entre les variables étudiées (Consanguinité et Déficiences), les fréquences observées devraient normalement correspondre aux fréquences théoriques. Mais comme nous travaillons avec un échantillon, il se peut que celui-ci s'écarte plus ou moins de sa distribution théorique par le simple jeu du hasard. Si l'écart global mesuré (ici le Khi carré) est trop faible, on peut alors juger qu'il n'est pas significatif. Par contre, si cet écart est suffisamment fort, on ne peut plus accuser le hasard : la présence d'un écart fort montre qu'il y a une association entre les variables. Il nous reste maintenant à chiffrer précisément ce que nous entendons par écart *faible* et *fort*.

En prenant le risque de nous tromper 5 fois sur 100, nous pourrions affirmer qu'il y a une association entre les variables si le Khi carré calculé dépasse 3,84.

Lorsque l'on consulte une table de distribution du Khi carré, on constate que la *valeur critique* du Khi carré pour un *seuil de signification* de 0,05 et avec 1 *degré de liberté* est de 3,84*. Voici ce que ce jargon signifie : parmi tous les échantillons que nous aurions pu recueillir, seulement 5 % (ou 0,05) d'entre eux ont un écart global (ou Khi carré calculé) supérieur à 3,84.

Rappelons que le nombre de degrés de liberté correspond au nombre de cases à l'intérieur du tableau à l'exclusion des cases de la dernière colonne et de la dernière ligne. Puisque notre tableau comprend ici 2 lignes et 2 colonnes, il possède 1 degré de liberté.

Est-ce que l'écart que nous avons observé correspond à nos attentes?

L'écart global de notre tableau (le Khi carré calculé, qui valait 1,36) est plus faible que le minimum attendu (la valeur trouvée dans la table de Khi carré, qui était de 3,84). La possibilité que l'écart observé soit le fruit du hasard est trop forte (ou n'est pas assez faible) pour que nous puissions affirmer qu'il existe une association significative entre les variables *Consanguinité* et *Mortalité infantile*.

Il existe une association significative entre *consanguinité* et *troubles physiques et mentaux*.

Si on répète le même exercice pour tester l'hypothèse d'une association entre *consanguinité* et *troubles physiques et mentaux*, on constate que la valeur du Khi carré calculé est nettement plus significative : elle est de 16,38. La probabilité de tomber sur un tel échantillon uniquement par le fait du hasard est à toutes fins pratiques nulle (cette probabilité est de 0,005 % plus précisément). Nous avons obtenu ces derniers résultats en quelques secondes : le tableau D9.2 ayant été créé avec un chiffrier électronique, il nous a suffi de modifier les 4 fréquences observées pour obtenir instantanément le nouveau Khi carré calculé. Dans le même ordre d'idée, la valeur critique du Khi carré a été obtenue par une simple fonction du chiffrier : il est rare aujourd'hui que des chercheurs utilisent encore des outils aussi primitifs que des tables de distribution.

Maintenant que nous savons que l'écart obtenu n'est pas dû au hasard, il reste à en évaluer l'ampleur en tenant compte de la taille de l'échantillon. Pour cela, nous calculons le V de Cramer, selon la [formule](#) donnée dans le chapitre.

$$V \text{ de Cramer} = \sqrt{\frac{K\chi^2}{n \times (K - 1)}}$$

$$V \text{ de Cramer} = \sqrt{[16,37 / (256 \times 1)]} = \sqrt{[0,064]} = 0,253$$

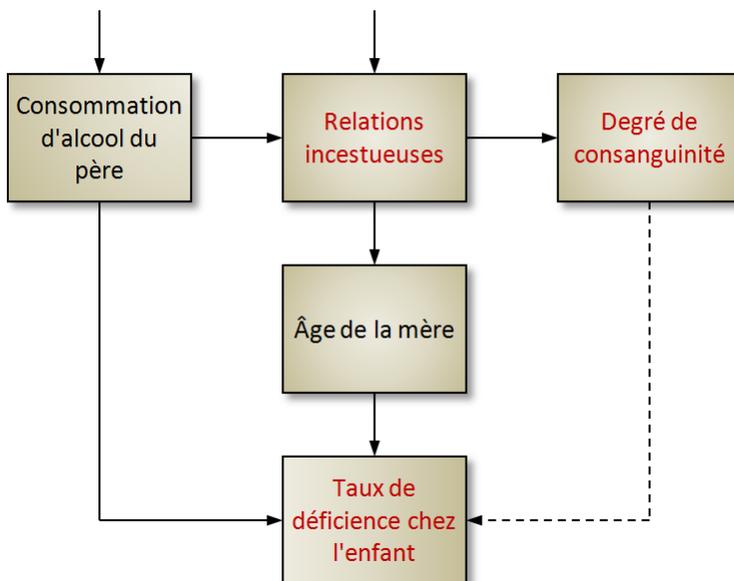
En fin de compte, l'association entre *consanguinité* et *troubles physiques et mentaux* est bien réelle, même si elle n'est pas extrêmement forte. Cela est d'autant plus significatif que nous n'avons pas été en mesure de prouver l'existence d'une association entre *consanguinité* et *mortalité infantile*.

L'association entre *consanguinité des parents* et *déficiences des enfants* semble donc démontrée et l'affaire aurait pu être classée si on n'avait pas eu la curiosité de mieux connaître les parents incestueux. On constata alors deux faits troublants : comparativement aux parents du groupe témoin, les pères incestueux étaient plus facilement alcooliques, et les mères incestueuses accouchaient à un âge significativement plus précoce (voir le tableau D9.3). Or, on sait que ces deux facteurs augmentent les risques de déficiences (quand les parents boivent, les enfants trinquent). Il pourrait alors n'y avoir qu'une simple corrélation entre *inceste* et *déficience*, ces deux variables dépendant (en partie) de variables communes (Consommation d'alcool du père, Âge précoce de la mère). Le schéma de la figure D9.1 tente de mettre en évidence plusieurs aspects de cette corrélation.

Tableau D9.3 - Autres facteurs pouvant influencer les variables

| Autres facteurs | Alcooliques chroniques | Âge de la mère à l'accouchement |
|--------------------|------------------------|---------------------------------|
| Parents incestueux | 13 | Plus précoce |
| Autres parents | 2 | Moins précoce |

FIGURE D9.1 - Schéma de variables



Si l'on tient compte des facteurs socio-économiques liés à la situation, on s'aperçoit que la relation entre consanguinité et déficience est loin d'être prouvée. Les cas d'inceste provenaient souvent de

familles défavorisées en ce qui concerne les conditions de logement et d'alimentation, l'état de santé de la mère et la connaissance des soins à apporter aux enfants. Or, tous ces facteurs peuvent avoir une influence directe sur la santé des enfants aussi bien que sur le comportement des parents.

L'inceste est un sujet difficile à traiter. Tout d'abord, l'inceste suscite un dégoût « naturel » chez la plupart des gens. D'autre part, il est souvent associé à l'abus sexuel d'un adulte envers un enfant. Enfin, un certain nombre de préjugés pseudo-scientifiques circulent à son sujet. Mais comme tout sujet relié à l'être humain, il mérite d'être étudié avec sérieux et rigueur.

QUESTIONS

1. Les pères incestueux sont-ils des alcooliques?

Pour répondre aux questions suivantes, utilisez les données des tableaux D9.3 et D9.4.

- Quelle est la proportion d'alcooliques chroniques dans chacune des catégories de pères (pères incestueux et pères du groupe témoin)?
- Est-il possible que les différences observées entre ces catégories soient uniquement attribuables au hasard? Comment peut-on évaluer la probabilité d'une telle éventualité?
- Complétez le tableau D9.4.
- Calculez le Khi carré et interprétez.

TABLEAU D9.4 - Inceste et alcoolisme

Tableau D9.4a - Fréquences observées

| | Alcoolisme chronique | Non alcoolisme chronique | Total |
|------------------|----------------------|--------------------------|-------|
| Pères incestueux | 13 | 148 | |
| Autres pères | 2 | 93 | |
| Total | | | |

Tableau D9.4d - Proportion observées (en %)

| | Alcoolisme chronique | Non alcoolisme chronique | Total |
|------------------|----------------------|--------------------------|---------------|
| Pères incestueux | | | |
| Autres pères | | | 37,1% |
| Total | | | 100,0% |

Tableau D9.4b - Fréquences théoriques

| | Alcoolisme chronique | Non alcoolisme chronique | Total |
|------------------|----------------------|--------------------------|-------|
| Pères incestueux | 9 | | |
| Autres pères | | | |
| Total | | | |

Tableau D9.4e - Proportion théoriques (en %)

| | Alcoolisme chronique | Non alcoolisme chronique | Total |
|------------------|----------------------|--------------------------|---------------|
| Pères incestueux | 3,7% | | |
| Autres pères | | | 37,1% |
| Total | | | 100,0% |

Tableau D9.4c - Écarts au carré relatifs

| | Alcoolisme chronique | Non alcoolisme chronique | Total |
|------------------|----------------------|--------------------------|-------|
| Pères incestueux | 1,35 | | |
| Autres pères | | | |
| Total | 3,63 | | |

| | |
|------|---|
| | Degrés de liberté |
| 0,05 | Seuil de signification |
| | Valeur critique du Khi ² |
| | Probabilité de dépasser le Khi ² calculé |

2. Schéma de variables

Complétez le schéma proposé à la figure 1 en ajoutant quelques cases (variables) et quelques flèches (relations).