



## La pratique des sport au Canada

## Chapitre 16

### 16.0 Méthode

#### 16.1 L'Enquête sociale générale

Statistique Canada, au moyen d'un supplément à l'Enquête sociale générale concernant les sports, a sondé environ 10 000 Canadiens âgés de 15 ans et plus sur la nature et l'ampleur de leur participation au sport au cours des 12 mois précédents. L'enquête a été menée en 1992 et en 1998. Ces deux suppléments sur les sports représentent la principale source de données de la présente étude, plus particulièrement les cycles 7 et 12 intitulés Emploi du temps de l'ESG.

Les responsables du Programme de la statistique culturelle de Statistique Canada ont élaboré les questions sur le sport en tenant compte des besoins en information des partenaires fédéraux et provinciaux/territoriaux concernés. Un groupe d'utilisateurs des données aux échelons fédéral et provincial ont commandité des questions dans le but d'obtenir un aperçu plus complet de l'étendue de la participation au sport au Canada.

On déterminait cette participation par la question suivante :

*« Avez-vous pratiqué des sports de façon régulière au cours des 12 derniers mois? »*

« Régulièrement » signifiait que le répondant avait pratiqué un sport au moins une fois par semaine pendant la saison ou pendant une certaine période de l'année. Par exemple, bien que le jeu de quilles ne représente pas un sport saisonnier, le répondant pouvait l'inclure parmi les activités sportives s'il s'y était adonné de façon régulière durant une période de l'année.

Les types de sports visés par cette enquête avaient été déterminés à l'aide d'une liste fournie par Sport Canada (voir annexe), qui définissait le sport comme une activité qui :

- met aux prises deux participants ou plus dans le but d'évaluer leur performance personnelle dans un cadre de compétition;
- comprend des règles et des méthodes officielles;
- repose sur des tactiques et des stratégies;
- fait appel à des habiletés neuromusculaires spécialisées qui peuvent être enseignées et apprises;
- fait travailler de façon intensive des groupes de muscles importants;

- présente un degré élevé de difficultés, de risques et d'efforts dans la reproduction de mouvements ou de figures;
- comprend la formation d'entraîneurs compétents aux fins de la compétition;
- comprend essentiellement une interaction physique entre les participants et le milieu;
- exclut toute activité pour laquelle l'utilisation d'un véhicule motorisé est le principal déterminant des résultats de la compétition. Lorsque des véhicules mécanisés ou d'autres moyens de transport sont utilisés, l'activité doit comporter des efforts physiques importants en vue d'actionner le véhicule ou le moyen de transport.

Nous nous sommes appuyés sur ces lignes directrices générales de définition du sport pour exclure plusieurs activités physiques ou de loisirs telles que la gymnastique aérobique, la dansexercice, l'aquaforme, la bicyclette récréative ou comme moyen de transport, la course à pied et la marche, etc.

La période de référence des deux cycles d'enquête (1992 et 1998) était la participation au cours des 12 derniers mois. On admet qu'une période de référence d'un an comporte le problème du rappel. De plus, il faut tenir compte des risques d'exagération puisque les répondants peuvent hésiter à admettre, particulièrement en cette époque où la santé est mise en évidence, qu'ils n'ont pas été physiquement actifs au cours de la dernière année. On ne peut mesurer, à partir du questionnaire, si les répondants en ont trop mis ou pas assez quant à leur participation au sport.

L'ESG, qui remonte à 1985, se déroule par téléphone dans les dix provinces. Elle consiste en la collecte continue de données, ce qui permet ensuite d'analyser les tendances. Voici la liste des sujets de collecte des données de l'ESG :

| Sujets des cycles de l'ESG  | Date de l'ESG (n° de cycle) |                      |                      |
|---|-----------------------------|----------------------|----------------------|
|   | 1 <sup>re</sup> série       | 2 <sup>e</sup> série | 3 <sup>e</sup> série |
| La santé  | 1985 (1)                    | 1991 (6)             |                      |
| L'emploi du temps   | 1986 (2)                    | 1992 (7)             | 1998 (12)            |
| La victimisation  | 1988 (3)                    | 1993 (8)             | 1999 (13)            |
| Les études, le travail et la retraite   | 1989 (4)                    | 1994 (9)             |                      |
| La famille et les amis  | 1990 (5)                    | 1995 (10)            |                      |
| L'aide du milieu  | 1985 (1)                    | 1990 (5)             | 1996 (11)            |
| L'accessibilité de la technologie de l'information et de la communication et son utilisation par les citoyens | 2000 (14)                   |                      |                      |

Chaque enquête contient un sujet de base de même qu'un ensemble normalisé de questions socio-démographiques. La population cible de l'ESG est celle de toutes les personnes de 15 ans et plus résidant au Canada. Les résidents des territoires et les pensionnaires à temps plein d'un établissement institutionnel sont exclus de l'échantillon. La méthode de collecte des données est l'interview téléphonique assistée par ordinateur (ITAO). Les ménages ne possédant pas de téléphone ont donc été exclus de l'échantillon. Les personnes vivant dans des ménages dépourvus de téléphone représentent moins de 2 % de la population cible. Toutefois, les estimations de l'enquête ont été ajustées (c.-à-d. pondérées) pour tenir compte de ces personnes. La collecte de données par téléphone est intéressante, car elle ne coûte pas cher et offre une grande souplesse quant au plan d'échantillonnage. Néanmoins, les entrevues par téléphone comportent quelques désavantages : les foyers non visés, bien que leur nombre soit faible, sont concentrés dans les groupes de population ayant un faible niveau de scolarité ou de revenu; les taux de réponse tendent à être plus bas que les entrevues en personne, et il y a des limites à la quantité et au type de données qu'on peut recueillir par téléphone.

La taille de l'échantillon, jusqu'en 1998, était d'environ 10 000 personnes. Elle a été augmentée à 25 000 personnes en 1999, rendant possible la collecte de résultats plus détaillés à l'échelon national et provincial.

Le plan d'échantillonnage ne permet pas d'établir des estimations concernant les Autochtones.

#### Taux de réponse et de non-réponse à l'ESC ( % )

| Cycle | Sujet   | Année d'enq. | Taille de l'éch. | Taux de rép. |
|-------|---|--------------|------------------|--------------|
| 1     | Santé   | 1985         | 11 200           | 83,4         |
| 2     | Emploi du temps   | 1986         | 16 390           | 78,9         |
| 3     | Risques personnels  | 1988         | 9 870            | 82,4         |
| 4     | Éducation et travail  | 1989         | 9 338            | 80,7         |
| 5     | Famille et amis   | 1990         | 13 495           | 75,8         |
| 6     | Santé   | 1991         | 11 924           | 80,2         |
| 7     | Emploi du temps   | 1992         | 9 815            | 76,8         |
| 8     | Risques personnels  | 1993         | 10 385           | 81,6         |
| 9     | Études, travail et retraite   |              | 1994<br>81,2     | 11 876       |
| 10    | Famille   | 1995         | 10 749           | 81,4         |
| 11    | Aide du milieu  | 1996         | 12 756           | 85,3         |
| 12    | Emploi du temps   | 1998         | 10 749           | 77,6         |
| 13    | Victimisation   | 1999         | 25 000 (e)       | non disp.    |
| 14    | Accessibilité de la technologie de l'information et de la communication et son utilisation par les citoyens | 2000         | 25 000 (e)       | non disp.    |

Les données du cycle 12 de l'ESG ont été recueillies entre février 1998 et janvier 1999. L'échantillon a été réparti également sur les 12 mois afin de tenir compte des variations saisonnières dans les données.

### **Ponderation**

Lorsqu'on utilise un échantillon aléatoire, comme ce fut le cas pour l'ESG, le principe qui sous-tend l'estimation est que chacune des personnes sélectionnées représente (outre elle-même) plusieurs autres personnes n'entrant pas dans l'échantillon. Par exemple, dans le cas d'un échantillon aléatoire simple de 2 % de la population, chaque personne de l'échantillon en représente 50 dans la population.

Au moment d'analyser les données du cycle 12 de l'ESG, il faut utiliser comme facteur de pondération soit WGHTFIN (fichier principal), soit WGHTPEI (fichier épisodique de l'emploi du temps).

**Les utilisateurs ne doivent pas diffuser des tableaux non pondérés ou effectuer des analyses fondées sur des résultats d'enquête non pondérés.** Les taux d'échantillonnage et les taux de non-réponse varient considérablement d'une province à l'autre et selon diverses caractéristiques démographiques. Par exemple, nous savons que les non-répondants sont probablement des hommes et que ceux-ci sont généralement jeunes. Une proportion de 3,3 % des répondants étaient des hommes âgés entre 15 et 19 ans, alors que la population globale en compte une proportion de 4,3 %. Par conséquent, il est clair que l'échantillon non pondéré ne peut être considéré comme représentatif de la population cible de l'enquête.

Il faut utiliser les facteurs de pondération de l'enquête au moment de produire des estimations ou de réaliser des analyses afin de tenir compte autant que possible de la sous-représentation ou surreprésentation géographique ainsi que de la sous-représentation ou surreprésentation des groupes répartis selon l'âge et le sexe, des mois de l'année, ou des journées de la semaine dans le fichier non pondéré.

Dans un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations tirées de l'enquête. Toutefois, les erreurs qui se produisent systématiquement contribueront à biaiser les estimations de l'enquête. Des erreurs non liées à l'échantillonnage peuvent se produire à presque toutes les phases de l'enquête. Il se peut que les intervieweurs comprennent mal les instructions, que les répondants se trompent en répondant aux questions, que des réponses soient mal entrées dans l'ordinateur et que des erreurs se produisent au moment du traitement ou de la mise en tableaux des données. Il s'agit d'exemples d'erreurs non dues à l'échantillonnage.

Nous avons consacré beaucoup de temps et d'efforts à réduire le nombre d'erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chaque étape du cycle de collecte et de traitement des données afin de contrôler la qualité des données. Entre autres mesures utilisées, mentionnons les suivantes : recours à des intervieweurs très qualifiés, à qui on a donné une formation poussée en ce qui touche les procédures d'enquête et le questionnaire; observation des intervieweurs pour repérer les problèmes liés à la conception du questionnaire ou à la mauvaise interprétation des instructions; procédures pour réduire au minimum les erreurs de saisie des données; vérifications de la qualité du codage et des corrections pour s'assurer de la logique du traitement.

L'incidence de la non-réponse sur les résultats d'enquête est une source importante d'erreurs non dues à l'échantillonnage. La non-réponse peut aller de la non-réponse totale à la non-réponse partielle (l'absence de réponse à seulement une ou plusieurs questions). Il y a non-réponse totale lorsque l'intervieweur n'a pu communiquer avec le répondant, qu'aucun membre du ménage n'a pu fournir l'information, ou que le répondant a refusé de participer à l'enquête. On a traité la non-réponse totale en ajustant le coefficient de pondération des ménages qui ont répondu à l'enquête.

### **Coefficient de variation (c.v.)**

Comme il est inévitable que les estimations d'une enquête menée à partir d'un échantillon fassent l'objet d'une erreur d'échantillonnage, la pratique habituelle en statistique est de fournir à l'utilisateur un indice de l'importance de cette erreur.

En raison de la grande diversité des estimations que l'on peut tirer d'une enquête, l'écart-type d'une estimation est habituellement exprimé en fonction de l'estimation à laquelle elle se rapporte. La mesure qui en découle, appelée coefficient de variation (c.v.) d'une estimation, s'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même et s'exprime en pourcentage de l'estimation. **Avant de diffuser ou de publier des estimations faites à partir du fichier de microdonnées, les utilisateurs devraient d'abord s'appuyer sur les lignes directrices suivantes.**

| <b>Coefficient de variation</b> |   | <b>Instruction générale</b>        |
|---------------------------------|---|------------------------------------|
| De 0,0 à 16,5 %                 | Variabilité d'échantillonnage modérée     | Communicable                       |
| De 16,6 % à 33,3 %              | Variabilité d'échantillonnage élevée      | Communicable<br>avec mise en garde |
| 33,4 % ou plus                  | Variabilité d'échantillonnage très élevée | <b>Non communicable</b>            |

Nota : Il faudrait également appliquer aux estimations arrondies la politique relative à la variabilité d'échantillonnage.

Pour que les estimations produites à partir des fichiers de microdonnées de l'ESG correspondent à celles produites par Statistique Canada, nous exhortons les utilisateurs à arrondir les estimations au millier le plus près au moyen de la technique d'arrondissement habituelle. Il pourrait être trompeur de diffuser des estimations non arrondies étant donné qu'elles supposent une précision plus grande qu'en réalité. Si on utilise une technique d'arrondissement différente en raison de limites techniques ou autres, donnant ainsi des estimations différentes de celles de Statistique Canada, les utilisateurs sont encouragés à indiquer la raison de ces différences dans le document publié.

### **Test t**

On a procédé à un test t pour les taux de participation au sport afin de déterminer si la différence dans les taux était significative ou non sur le plan statistique.

Une couple d'hypothèses sont énoncées : ainsi, supposons que la répartition est normale. Un test bilatéral sera utilisé à un niveau de confiance de 95 %. La valeur critique est de 1,96.

Hypothèse : Si les taux de participation de 1992 et de 1998 n'ont presque pas changé, c'est-à-dire si la différence se situe entre 1,96 et -1,96, on peut en conclure que ce changement n'est pas significatif sur le plan statistique. Mais si le résultat du test t est supérieur à + ou - 1,96, les changements dans les taux de participation entre 1992 et 1998 sont alors importants sur le plan statistique.

Le résultat du test t, à savoir -13,2, se situe bien au-delà de la valeur critique de + ou - 1,96, de sorte que la différence dans les taux de participation entre 1992 et 1998 est statistiquement importante, à un niveau de confiance de 95 %.