

10

**RAPPORT SUR LA QUALITÉ
DES DONNÉES**

Introduction

Le coefficient de variation d'une statistique donnée est largement tributaire de la taille de l'échantillon global de l'enquête et de l'importance de la sous-population par rapport à la population canadienne totale. Il dépend également du taux de non-réponse et du plan de sondage adopté. Le rapport présente, à titre d'information générale, des tableaux affichant les tailles d'échantillon selon la province de même que les taux de réponse mensuels selon la province.

Ce chapitre précise la façon d'obtenir le coefficient de variation approximatif d'une statistique donnée à partir des tableaux des degrés approximatifs de variabilité de l'échantillonnage pour l'Enquête sur les voyages des Canadiens.

***Taille de
l'échantillon***

Le tableau suivant indique le nombre de membres des ménages qui, dans les groupes de renouvellement échantillonnés de l'EPA, étaient admissibles à l'enquête supplémentaire, soit l'Enquête sur les voyages des Canadiens.

TABLEAU 6. *Taille des échantillons mensuels selon la province, 1996*

	Jan.	Fév.	Mar.	Avr.	Mai	Juin
T.-N.	577	590	585	879	898	888
Î.-P.-É.	462	416	440	451	439	460
N.-É.	1110	1046	1072	1095	1100	1117
N.-B.	997	1004	990	1018	1010	1032
Qc	3217	3278	3328	3326	3313	3282
Ont.	4861	5019	4960	5071	5174	5131
Man.	1121	1185	1149	1219	1158	1171
Sask.	1055	1059	1054	1022	1079	1016
Alb.	1222	1218	1190	1241	1292	1233
C.-B.	1506	1441	1494	1528	1530	1499
Canada	16128	16256	16262	16850	16993	16829

TABLEAU 6.
*(suite)***Taille des échantillons mensuels selon la province, 1996**

	Juil.	Août	Sept.	Oct.	Nov.	Déc.
T.-N.	887	924	930	949	937	918
Î.-P.-É.	455	448	463	425	442	468
N.-É.	1124	1148	1132	1106	1153	1134
N.-B.	1042	1046	1076	1079	1051	1061
Qc	3391	3475	3588	3466	3433	3502
Ont.	5209	5156	5213	5145	5109	5109
Man.	1242	1200	1198	1256	1194	1212
Sask.	1026	1093	1087	1025	1100	1088
Alb.	1282	1328	1240	1309	1279	1220
C.-B.	1595	1607	1552	1612	1587	1536
Canada	17253	17425	17479	17372	17285	17248

Taux de réponse

Le tableau suivant présente les taux de réponse de l'Enquête sur les voyages des Canadiens de 1996. Les taux de réponse indiqués correspondent à la proportion des personnes admissibles à l'Enquête sur les voyages des Canadiens qui ont fourni des renseignements. Ces taux de réponse ne sont pas cumulatifs, c'est-à-dire qu'ils ne tiennent pas compte des personnes qui auraient été admissibles à l'EVC mais qui n'ont pas pris part à l'EPA. En effet, on ne demande pas aux personnes qui ne participent pas à l'EPA si elles veulent répondre aux questions de

l'EVC. On ne peut donc pas les considérer comme des non-répondants à l'EVC.

TABLEAU 7. *Taux de réponse mensuels selon la province, en pourcentage, 1996*

	Jan.	Fév.	Mar.	Avr.	Mai	Juin
T.-N.	94.1	95.4	92.0	92.0	91.1	88.0
Î.-P.-É.	94.4	93.3	94.0	92.2	94.3	92.0
N.-É.	95.0	95.0	92.0	93.0	92.0	92.0
N.-B.	94.0	92.4	92.0	94.0	91.0	89.1
Qc	95.0	94.0	94.1	93.0	93.0	92.0
Ont.	93.0	91.0	91.2	90.3	90.3	88.4
Man.	91.0	91.0	90.3	93.0	91.1	88.0
Sask.	91.0	89.2	89.3	89.0	87.0	85.3
Alb.	94.0	93.0	93.0	92.3	90.0	89.0
C.-B.	93.1	91.0	89.0	91.0	90.3	87.0
Canada	93.4	92.1	92.0	92.0	91.0	89.0

TABLEAU 7. *Taux de réponse mensuels selon la province, en pourcentage, 1996*
(suite)

	Juil.	Août	Sept.	Oct.	Nov.	Déc.
T.-N.	90.0	93.0	91.2	94.3	93.0	94.2
Î.-P.-É.	92.3	94.0	96.0	94.4	96.2	96.4
N.-É.	92.0	92.2	92.3	93.3	93.1	94.0
N.-B.	91.0	90.0	92.0	92.0	89.2	91.0
Qc	93.0	94.0	94.3	93.4	94.0	94.0
Ont.	89.4	90.0	90.4	91.1	91.0	90.0
Man.	88.2	88.2	92.0	92.1	91.2	90.0
Sask.	88.0	87.4	89.1	90.2	88.4	87.2
Alb.	90.4	88.0	92.2	93.0	89.0	90.2
C.-B.	89.2	90.0	91.0	91.4	90.3	90.4
Canada	90.3	91.0	92.0	92.2	91.4	91.1

Effet du plan de sondage

Le tableau suivant présente l'effet du plan de sondage, la taille de l'échantillon et les chiffres de population selon la province, qui ont servi à produire les tableaux des degrés approximatifs de variabilité de l'échantillonnage visant le poids de la personne. Il convient de noter qu'en dépit du fait que l'EVC se caractérise par des échantillons et des populations de taille différente chaque mois, l'effet du plan de sondage demeure constant d'un mois à l'autre. C'est pourquoi l'effet du plan de sondage, la taille

de l'échantillon et l'effectif de la population ne sont présentés que pour un seul mois.

TABLEAU 8. *Effets du plan de sondage, mars 1996*

Province	Effet du plan de sondage	Taille de l'échantillon	Population
T.-N.	2.0	536	453272
Î.-P.-É.	2.0	412	105966
N.-É.	2.0	986	734177
N.-B.	2.0	906	599604
Qc	2.0	3133	5858168
Ont.	2.0	4523	8820143
Man.	2.5	1037	853597
Sask.	2.0	941	751563
Alb.	2.5	1103	2097529
C.-B.	2.5	1330	3008863
Provinces atlantiques	2.0	2840	1893019
Man. et Sask.	2.5	1978	1605160
Alb. et C.-B.	2.5	2433	5106392
Canada	3.0	14907	23282882

**Seuils
relatifs à la
diffusion
pour l'EVC**

L'effectif minimum des groupes visés par l'estimation (si l'on tient compte des poids de la personne) est indiqué dans le tableau suivant pour le niveau provincial, régional et pour le Canada. Les estimations qui portent sur des effectifs inférieurs à ceux de l'effectif minimum indiqué dans la colonne Inacceptable ne peuvent être diffusées en aucun cas. Il convient de noter qu'un seul tableau affichant les seuils relatifs à la diffusion est présenté ci-dessous. Ce tableau donne les seuils relatifs à la diffusion pour mars 1996. Des tableaux supplémentaires pourraient être disponibles pour d'autres mois, soit pour ceux au cours desquels on a utilisé plus de deux groupes de renouvellement de l'EPA dans l'échantillon de certaines provinces.

TABLEAU 9. Spécimen de tableau des seuils relatifs à la diffusion, mars 1996

Province	Acceptable	Médiocre	Confidentiel	Inacceptable
T.-N.	54,500 +	25,500-54,499	15,000-25,499	moins de 15,000
Î.-P.-É.	16,000 +	7,500-15,999	4,500-7,499	moins de 4,500
N.-É.	51,000 +	23,000-50,999	13,000-22,999	moins de 13,000
N.-B.	45,000 +	20,500-44,999	11,500-20,499	moins de 11,500
Qc	134,000 +	59,000-133,999	33,500-58,999	moins de 33,500
Ont.	141,000 +	62,000-140,999	35,000-61,999	moins de 35,000
Man.	69,500 +	31,500-69,499	18,000-31,499	moins de 18,000
Sask.	54,500 +	24,500-54,499	14,000-24,499	moins de 14,000
Alb.	161,000 +	73,500-160,999	42,000-73,499	moins de 42,000
C.-B.	194,500 +	88,000-194,499	50,000-87,999	moins de 50,000
Provinces atlantiques	47,500 +	21,000-47,499	12,000-20,999	moins de 12,000
Man. et Sask.	71,000 +	32,000-70,999	18,000-31,999	moins de 18,000
Alb. et C.-B.	185,500 +	82,500-185,499	47,000-82,499	moins de 47,000
Canada	171,000 +	74,500-170,999	42,000-74,999	moins de 42,000

**Obtention
des CV
approximatif
à partir des
tableaux**

Les coefficients de variation approximatifs (CV) sont présentés dans les tableaux des degrés approximatifs de variabilité de l'échantillonnage à la fin de ce chapitre et sur le CD-ROM. Avant d'appliquer le critère du coefficient de variation, il importe de suivre les lignes directrices fondées sur la taille des échantillons décrites au [chapitre 9](#) (Lignes directrices pour la diffusion).

Les règles et les exemples qui suivent devraient permettre à l'utilisateur de déterminer, à partir des tableaux de variabilité de l'échantillonnage, les coefficients de variation approximatifs d'estimations du nombre de personnes qui, dans la population observée, possèdent une caractéristique donnée. Les exemples concrets visent à aider l'utilisateur à appliquer ces règles. Ils sont fondés sur des variables nécessitant l'application du poids de la personne en vue de la production des estimations.

Règle 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Sur le tableau de variabilité de l'échantillonnage correspondant à la région géographique appropriée, repérez le nombre estimé dans la colonne à l'extrême gauche de la table (intitulée «Estimation») et suivez les astérisques (le cas échéant) jusqu'au premier nombre. Ce nombre constitue le coefficient de variation approximatif.

Exemple illustrant la règle 1

Supposons qu'un utilisateur estime à 6 032 234 le nombre de personnes ayant fait au moins un voyage en mars 1996. Comment détermine-t-il le coefficient de variation de cette estimation?

- ▶ Reportez-vous à la table de CV pour le CANADA.

- ▶ Comme l'estimation de l'agrégat (6 032 234) n'apparaît pas dans la colonne de gauche (intitulée «Numérateur du pourcentage»), il faut utiliser le chiffre qui s'en rapproche le plus, soit 6 000 000.

- ▶ Le coefficient de variation de l'estimation de l'agrégat correspond au premier chiffre suivant les astérisques dans cette rangée, soit 2,3 %.

- ▶ Par conséquent, le coefficient de variation approximatif de cette estimation est égal à 2,3 %. Le résultat obtenu, soit que 6 032 234 personnes ont fait au moins un voyage en mars 1996, peut donc être publié sans restriction.

Règle 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Le coefficient de variation de l'estimation d'une proportion ou d'un pourcentage dépend à la fois de la taille de la proportion ou du pourcentage et de la taille de la population totale à laquelle cette proportion ou ce pourcentage se rapporte. Les estimations de proportions ou de pourcentages sont

relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage lorsque la proportion ou le pourcentage se rapporte à un sous-groupe de la population. Par exemple, la proportion des «personnes âgées de 15 ans et plus ayant fait au moins un voyage au cours du mois de référence» est plus fiable que l'estimation du nombre de «personnes âgées de 15 ans et plus ayant fait au moins un voyage au cours du mois de référence». (Il convient de noter que, dans les tableaux, la valeur des CV décroît de la gauche vers la droite).

Lorsque la proportion ou le pourcentage se rapporte à la population totale de la région géographique visée par le tableau, le CV de cette proportion ou de ce pourcentage est égal à celui du numérateur de la proportion ou du pourcentage. Dans ce cas, on peut utiliser la règle 1.

Lorsque la proportion ou le pourcentage se rapporte à un sous-ensemble de la population totale (p. ex., aux personnes d'un sexe ou d'un groupe d'âge donné), on doit se reporter à la proportion ou au pourcentage (rangée au haut du tableau) et au numérateur de cette proportion ou de ce pourcentage (colonne de gauche du tableau). Le coefficient de variation se trouve à l'intersection de la rangée et de la colonne appropriées.

Exemple illustrant la règle 2

Supposons qu'un utilisateur estime à $2\,951\,511 / 6\,032\,234 = 49\%$ le pourcentage des personnes qui, ayant voyagé en mars, ont effectué au moins un

voyage du même jour. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- ▶ Reportez-vous à la table visant le CANADA.
- ▶ Étant donné que l'estimation est un pourcentage se rapportant à un sous-ensemble de la population totale (soit les voyageurs qui ont fait au moins un voyage du même jour en mars), il faut utiliser à la fois le pourcentage (49 %) et le numérateur de ce pourcentage (2 951 511) pour déterminer le coefficient de variation.
- ▶ Le numérateur, 2 951 511, n'apparaissant pas dans la colonne de gauche (intitulée «Numérateur du pourcentage»), il convient d'utiliser le chiffre qui s'en rapproche le plus, soit 3 000 000. De la même façon, l'estimation du pourcentage ne figure dans aucun des titres de colonne, et il faut donc se reporter au pourcentage qui s'en rapproche le plus, soit 50,0 %.
- ▶ Le chiffre indiqué à l'intersection de la rangée et de la colonne consultées, soit 2,8 %, correspond au coefficient de variation recherché.
- ▶ Par conséquent, le coefficient de variation approximatif de cette estimation est 2,8 %. Le résultat obtenu, soit que 49 % des personnes ayant voyagé en mars ont fait au moins un voyage du même jour, peut donc être publié sans restriction.

Règle 3 : Estimations des différences entre des agrégats ou des pourcentages

L'erreur-type de la différence entre deux estimations correspond à peu près à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. Par conséquent, l'erreur-type de la différence ($d = X_1 - X_2$) se calcule ainsi :

$$\sigma_d = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où X_1 représente l'estimation 1, X_2 , l'estimation 2, et α_1 et α_2 sont les coefficients de variation de X_1 et de X_2 , respectivement. Le coefficient de variation de d est donné par σ_d/d . Cette formule donne des résultats précis lorsqu'on l'applique à la différence entre des caractéristiques distinctes et non corrélées; elle donne des résultats approximatifs dans les autres cas.

Exemple illustrant la règle 3

Supposons qu'un utilisateur estime à $2\,951\,511 / 6\,032\,234 = 49\%$ le pourcentage de personnes qui, ayant voyagé en mars, ont fait au moins un voyage du même jour et à $3\,998\,785 / 6\,032\,234 = 66,3\%$ le pourcentage de personnes qui, ayant voyagé en mars, ont fait au moins un voyage d'une nuit ou plus. (Notons qu'une même personne peut avoir fait un voyage du même jour et un voyage d'une nuit ou plus au cours du même mois, ce qui se traduirait par un chevauchement des deux estimations). Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- ▶ L'utilisation de la table des CV pour le CANADA selon la méthode décrite à l'exemple 2 permet d'établir à 2,8 % le CV de l'estimation des voyageurs ayant fait au moins un voyage du même jour et à 2,1 % le CV des voyageurs ayant fait au moins un voyage d'une nuit ou plus.
- ▶ Conformément à la règle 3, on calcule l'erreur-type d'une différence exprimée sous la forme ($\hat{d} = \hat{X}_1 - \hat{X}_2$) comme suit :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 représente l'estimation 1, \hat{X}_2 , l'estimation 2, et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 , respectivement.

L'erreur-type de la différence $\hat{d} = (.663 - .490) = .173$ est donc égale à :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(.49)(.028)]^2 + [(.660)(.021)]^2} \\ &= \sqrt{(.00001882384) + (.0001920996)} \\ &= .0195022\end{aligned}$$

Le coefficient de variation de \hat{d} correspond à $\sigma_{\hat{d}}/\hat{d} = .019/.173 = 0.11$.

- ▶ Par conséquent, le coefficient de variation approximatif de la différence entre les estimations est égal à 11 %. Cette estimation peut donc être publiée sans restriction.

Règle 4 : Estimations de rapports

Si le numérateur est un sous-ensemble du dénominateur, il faut exprimer le rapport en pourcentage et appliquer la règle 2. Cette situation se produirait, par exemple, si le dénominateur correspondait au nombre de «personnes ayant fait au moins un voyage au cours du mois de référence» et si le numérateur correspondait au nombre de «personnes ayant fait au moins un voyage d'affaires au cours du mois de référence».

Si le numérateur n'est pas un sous-ensemble du dénominateur (par exemple, s'il s'agit du rapport entre le nombre de «personnes ayant fait au moins un voyage d'affaires au cours du mois de référence» et le nombre de «personnes ayant fait au moins un voyage d'agrément au cours du mois de référence»), l'écart-type du rapport entre les estimations correspond à peu près à la racine carrée de la somme des carrés de chaque coefficient de variation considéré séparément multipliée par R . L'erreur-type d'un rapport exprimée sous la forme ($R = X_1 / X_2$) se calcule donc ainsi :

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 représentent les coefficients de variation de X_1 et de X_2 respectivement.

Le coefficient de variation de R est donné par σ_R/R . La formule tendra à surestimer l'erreur s'il existe une corrélation positive entre X_1 et X_2 et à sous-

estimer l'erreur s'il existe une corrélation négative entre X_1 et X_2 .

Exemple illustrant la règle 4

Supposons qu'un utilisateur estime à 3 998 785 le nombre de personnes qui, ayant voyagé en mars, ont fait au moins un voyage d'une nuit ou plus et à 2 951 511 le nombre de personnes qui, ayant voyagé en mars, ont fait au moins un voyage du même jour. L'utilisateur cherche à comparer l'estimation du nombre de voyageurs ayant fait des voyages d'une nuit ou plus à celle du nombre de voyageurs ayant fait des voyages du même jour et veut exprimer cette comparaison sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- ▶ Tout d'abord, cette estimation est une estimation par ratio dont le numérateur ($= X_1$) correspond au nombre de personnes qui, ayant voyagé en mars, ont fait au moins un voyage d'une nuit ou plus. Le dénominateur de l'estimation ($= X_2$) correspond au nombre de personnes qui, ayant voyagé en mars, ont fait au moins un voyage du même jour.
- ▶ Reportez-vous à la table visant le CANADA.
- ▶ Le numérateur de cette estimation par ratio est 3 998 785. Le chiffre qui s'en rapproche le plus est 4 000 000. Le coefficient de variation de cette estimation correspond au premier chiffre suivant les astérisques dans cette rangée, 3,1%.

- ▶ Le dénominateur de cette estimation par ratio est 2 951 511. Le chiffre qui s'en rapproche le plus est 3 000 000. Le coefficient de variation de cette estimation correspond au premier chiffre suivant les astérisques dans cette rangée, 3,6%.
- ▶ Conformément à la règle 4, le coefficient de variation approximatif de cette estimation par ratio se calcule comme suit :

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 représentent les coefficients de variation de \bar{X}_1 et de \bar{X}_2 respectivement.

Par conséquent :

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(.031)^2 + (.036)^2} \\ &= 0.047\end{aligned}$$

- ▶ Le rapport obtenu entre les voyageurs de mars 1996 ayant fait au moins un voyage d'une nuit ou plus et les voyageurs de mars 1996 ayant fait au moins un voyage du même jour est de 3 998 785/ 2 951 511, soit 1.35:1. Le coefficient de variation de cette estimation s'établit à 4,7%, et l'estimation peut donc être diffusée sans restriction.

Règle 5 : Estimations de différences entre des rapports

Dans ce cas, on combine les règles 3 et 4. On calcule d'abord les CV des deux rapports au moyen

de la règle 4, puis le CV de leur différence conformément à la règle 3.

**Utilisation
des tables
de CV pour
déterminer
les
intervalles
de confiance**

Bien que les coefficients de variation soient largement utilisés, on peut mesurer de façon significative l'erreur d'échantillonnage par une mesure plus intuitive, soit l'intervalle de confiance d'une estimation. L'intervalle de confiance constitue une assertion quant à la probabilité de confiance que la valeur réelle de la population se situe dans une fourchette de valeurs donnée. Par exemple, un intervalle de confiance de 95 % peut se définir comme suit :

Si l'échantillonnage de la population se répète à l'infini, chacun des échantillons modifiant l'intervalle de confiance d'une estimation, l'intervalle contiendra la véritable valeur de la population dans 95 % des cas.

Une fois que l'erreur-type d'une estimation est déterminée, on peut calculer les intervalles de confiance d'une estimation en partant de l'hypothèse que les diverses estimations d'une caractéristique donnée d'une population, obtenues à partir d'un échantillonnage répété de cette population, se dispersent normalement autour de la valeur réelle de la population. Compte tenu de cette hypothèse, dans environ 68 cas sur 100, l'écart entre une estimation par échantillonnage et la véritable valeur de la population sera inférieur à une erreur-type, dans environ 95 cas sur 100, cet écart sera inférieur à deux erreurs-types et, dans environ 99 cas sur 100, l'écart sera inférieur à trois erreurs-types. On

appelle ces différents degrés de confiance des niveaux de confiance.

Les intervalles de confiance pour une estimation \hat{X} sont généralement exprimés sous la forme de deux nombres, l'un étant inférieur à l'estimation et l'autre, supérieur à celle-ci, sous la forme $(\hat{X}-k, \hat{X}+k)$ où k varie selon le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

On peut calculer directement les intervalles de confiance d'une estimation à partir des tableaux des degrés approximatifs de variabilité de l'échantillonnage en déterminant d'abord, au moyen du tableau approprié, le coefficient de variation de l'estimation \hat{X} , puis en utilisant la formule suivante pour obtenir l'intervalle de confiance CI correspondant :

$$CI_X = [\hat{X} - t \hat{X} \alpha_{\hat{X}}, \hat{X} + t \hat{X} \alpha_{\hat{X}}]$$

où α_X est le coefficient de variation obtenu pour \hat{X} ,

$t = 1$ si l'on désire un intervalle de confiance de 68 %

$t = 1.6$ si l'on désire un intervalle de confiance de 90 %

$t = 2$ si l'on désire un intervalle de confiance de 95 %

$t = 3$ si l'on désire un intervalle de confiance de 99 %.

Note : Les lignes directrices qui s'appliquent à la diffusion des estimations s'appliquent aussi aux intervalles de confiance. Par exemple, si l'estimation ne peut être diffusée, alors l'intervalle de confiance ne peut, lui non plus, être diffusé.

**Exemple
d'utilisation
des tables
de CV pour
déterminer
les
intervalles
de confiance**

Un intervalle de confiance de 95 % pour l'estimation de la proportion des personnes qui ont voyagé en mars et qui ont fait au moins un voyage du même jour (tiré de l'exemple illustrant la règle 2 du [chapitre 10](#)) se calculerait comme suit.

$$\hat{X} = 49 \% \text{ (ou, sous forme de proportion, } = .49)$$

$$t = 2$$

$\alpha_X = 2,8 \%$ (soit la proportion .028 exprimée en pourcentage) est le coefficient de variation de cette estimation déterminé à partir des tables.

$$CI_X = \{.49 - (2) (.49) (.028), .49 + (2) (.49) (.028)\}$$

$$CI_X = \{.49 - .027, .49 + .027\}$$

$$CI_X = \{.463, .517\}$$

On peut affirmer avec un degré de confiance de 95 % qu'entre 46,3 % et 51,7 % des personnes qui ont voyagé en mars ont fait au moins un voyage du même jour.

**Utilisation
des tables
de CV pour
effectuer un
test t**

On peut aussi utiliser les erreurs-types pour effectuer des tests d'hypothèses, une technique qui permet de faire la distinction entre les paramètres d'une population à l'aide d'estimations fondées sur un échantillon. Ces estimations peuvent être des nombres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification, le niveau de signification étant la probabilité de conclure que les

caractéristiques sont différentes quand, en fait, elles sont identiques.

Soit \mathbf{X}_1 et \mathbf{X}_2 des estimations fondées sur un échantillon visant deux caractéristiques données. Supposons que l'erreur-type de la différence $\mathbf{X}_1 - \mathbf{X}_2$ est σ_d .

Si
$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d}$$
 se situe entre -2 et 2, alors on ne

peut tirer aucune conclusion quant à la différence entre les caractéristiques à un seuil de signification de 5 %. En revanche, si ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative considérant un seuil de 0.05. En d'autres mots, les caractéristiques sont significatives.

Exemple d'utilisation des tables de CV pour effectuer un test t

Supposons que l'on désire tester, à un niveau de signification de 5 %, l'hypothèse qu'il n'existe pas de différence entre la proportion des voyageurs en mars qui ont fait au moins un voyage du même jour et la proportion de voyageurs en mars qui ont fait au moins un voyage d'une nuit ou plus. Dans l'exemple illustrant la règle 3, l'erreur-type de la différence entre ces deux estimations a été établie à .019. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d} = \frac{.49 - .663}{.019} = \frac{-.173}{.019} = -9.10.$$

Puisque $t = -9.10$ est inférieur à -2, on doit conclure qu'il existe une différence significative entre les deux estimations à un seuil de signification de 0.05.

CV des estimations quantitatives

Pour calculer l'erreur d'échantillonnage d'estimations quantitatives, il faudrait élaborer des tableaux spéciaux, ce qui n'a pas été fait puisque la plupart des variables de l'Enquête sur les voyages des Canadiens sont essentiellement de type nominal. Les tableaux sont compris dans la documentation et peuvent être utilisés dans le cadre de l'analyse des variables qui nécessitent l'application des poids du voyage-personne, du voyage-ménage, de la nuitée-personne et des dépenses.

En règle générale, le coefficient de variation d'un total quantitatif est supérieur à celui de l'estimation nominale correspondante (c.-à-d, l'estimation du nombre de personnes contribuant à l'estimation quantitative). Si l'estimation nominale correspondante ne peut être diffusée, il en sera de même pour l'estimation quantitative. Par exemple, le coefficient de variation du nombre total de voyages effectués en mars sera supérieur au coefficient de variation de la proportion correspondante des personnes qui ont fait au moins un voyage en mars. Par conséquent, si le coefficient de variation de la proportion ne peut être diffusé, il en sera de même pour le coefficient de variation de l'estimation quantitative correspondante.

Dans le cas d'estimations de ce genre, on peut déterminer au besoin les coefficients de variation d'une estimation particulière à l'aide d'une technique appelée pseudo-réplication qui consiste à diviser les enregistrements des fichiers de microdonnées en sous-groupes (ou sous-échantillons) et à calculer la variabilité de l'estimation d'un sous-échantillon à l'autre. Les

utilisateurs qui désirent calculer les coefficients de variation d'estimations quantitatives peuvent demander conseil à Statistique Canada relativement à la répartition des enregistrements en sous-échantillons appropriés et aux formules à utiliser pour ces calculs.