

VARIABILITÉ D'ÉCHANTILLONNAGE

L'Enquête sur la population active, sur laquelle reposent les estimations disponibles sur le disque compact, recueille des renseignements auprès d'un échantillon de ménages. On pourrait obtenir des estimations différentes d'un recensement complet des ménages réalisé dans un environnement identique à celui de l'Enquête sur la population active c.à.d. basé sur les mêmes questionnaires, intervieweurs, méthodes de traitement, etc.

L'erreur d'échantillonnage, ou erreur-type, est une mesure permettant de quantifier l'écart entre une estimation fondée sur un échantillon et la valeur tirée d'un recensement. Elle se base sur la notion du prélèvement de plusieurs échantillons, bien que, dans une enquête, on prélève un seul échantillon et on recueille des renseignements se rapportant aux unités de cet échantillon. Le prélèvement d'un grand nombre d'échantillons à partir d'une population donnée, selon le même plan d'échantillonnage, produirait une estimation située à une erreur-type de la valeur censitaire dans environ 68 % des échantillons et à deux erreurs-type de la valeur censitaire dans environ 95% des échantillons.

On peut aussi exprimer la variabilité d'échantillonnage en fonction de l'estimation. Le coefficient de variation (CV) est une mesure de l'erreur d'échantillonnage définie en pourcentage de l'estimation. En fait, il s'agit d'une erreur-type relative. Il est également possible d'établir des intervalles de confiance en utilisant le CV. Par exemple, avec un CV de 7 %, on peut dire que dans 68 % des échantillons, la valeur censitaire se trouvera au maximum à plus ou moins 7 % (la valeur du CV) de l'estimation alors que dans 95 % des échantillons, elle se situera au maximum à plus ou moins 14 % (le double du CV) de l'estimation. En ce qui concerne les estimations contenues sur ce disque compact, le CV est utilisé afin de donner une idée de l'exactitude associée aux estimations.

INDICATEURS DE LA VARIABILITÉ D'ÉCHANTILLONNAGE

Il est préférable d'obtenir des CV peu élevés puisque de tels CV indiquent que la variabilité d'échantillonnage est faible par rapport à l'estimation. Le CV dépend de la valeur des estimations, de la taille de l'échantillon à partir duquel l'estimation est établie, de la répartition de l'échantillon et de l'utilisation d'estimations postcensitaires de la population dans la procédure d'estimation. La valeur des estimations est importante parce que le CV représente l'erreur d'échantillonnage exprimée en pourcentage de l'estimation. Plus la valeur de l'estimation est faible, plus le CV est élevé (toutes choses étant égales par ailleurs). Par exemple, lorsque le taux de chômage est élevé, le CV peut être faible. Si le taux de chômage baisse en raison de l'amélioration de la situation économique, le CV correspondant augmentera. Normalement, dans le cas d'estimations similaires, l'estimation fondée sur l'échantillon le plus grand est associée au CV le plus bas, car l'erreur d'échantillonnage est plus petite.

Par ailleurs, les estimations se rapportant à des caractéristiques plus groupées donnent lieu à un CV plus élevé. Ainsi, les personnes employées dans les secteurs de la foresterie, de la pêche, de l'exploitation minière et de l'exploitation pétrolière et gazière au Canada sont plus regroupées sur le plan géographique que les femmes de 55 à 64 ans employées en Ontario. La variabilité d'échantillonnage correspondant à ce dernier groupe sera plus faible, même si la valeur des estimations est similaire.

Enfin, les estimations se rapportant à l'âge et au sexe sont généralement plus fiables que d'autres estimations analogues parce que, dans le cas des estimations de l'EPA, l'échantillon est calibré en fonction des estimations postcensitaires de la population selon diverses catégories d'âge et de sexe. Pour poursuivre l'exemple précédent, les personnes employées à temps partiel en Alberta seront associées à une variabilité d'échantillonnage plus forte que les hommes de 35 à 44 ans employés en Colombie-Britannique, même si la valeur des estimations est similaire.

Variabilité des estimations mensuelles

Pour déterminer le CV approximatif d'une estimation mensuelle, veuillez consulter le tableau A, lequel présente l'estimation en fonction de la région géographique et du CV. Les lignes correspondent à la région géographique à laquelle l'estimation est associée et les colonnes, au degré de précision exprimé en CV, selon la grandeur de l'estimation. Pour déterminer le CV d'une estimation X dans la région A, suivez la ligne de la région A jusqu'à l'estimation la plus proche de X sans la dépasser. Le titre de la colonne indiquera le CV approximatif. Par exemple, pour connaître l'erreur d'échantillonnage d'une estimation de 40 200 chômeurs à Terre-Neuve-et-Labrador en septembre 2004, on constate que l'estimation inférieure la plus proche, de 27 100, donne un CV de 5 %. L'estimation de 40 200 chômeurs à Terre-Neuve-et-Labrador aura donc un CV d'environ 5 %.

Tableau A: CV pour les estimations* mensuelles, Canada et provinces

Région géographique	Coefficient de variation								
	1,0 %	2,5 %	5,0 %	7,5 %	10,0 %	15,0 %	20,0 %	25,0 %	30,0 %
Canada	1 112,2	324,7	142,3	82,6	46,4	27,0	18,0	13,0	9,9
Terre-Neuve-et-Labrador	241,7	66,2	27,1	15,2	8,6	4,8	3,2	2,2	1,7
Île-du-Prince-Édouard	59,7	18,0	7,9	4,6	2,7	1,6	1,1	0,8	0,6
Nouvelle-Écosse	278,4	79,8	33,7	19,3	11,1	6,4	4,2	3,1	2,3
Nouveau-Brunswick	242,7	67,0	27,0	15,2	8,9	5,0	3,3	2,4	1,8
Québec	1 173,9	327,3	135,9	77,0	43,8	24,8	16,3	11,7	8,9
Ontario	1 060,6	304,0	131,5	75,8	42,2	24,3	16,1	11,6	8,8
Manitoba	231,5	68,3	30,1	17,6	9,9	5,8	3,9	2,8	2,1
Saskatchewan	193,6	56,6	24,8	14,4	8,1	4,7	3,1	2,3	1,7
Alberta	530,2	158,6	72,2	42,6	23,5	13,8	9,3	6,8	5,2
Colombie-Britannique	761,8	222,8	97,7	56,8	31,9	18,6	12,4	9,0	6,8

* Les estimations sont en milliers.

Le tableau A donne une idée approximative de la variabilité d'échantillonnage. Cette dernière est modélisée pour que le CV de l'estimation soit inférieur ou égal au CV du tableau dans environ 75 % des cas. Néanmoins, dans 25 % des cas, le CV de l'estimation sera plus élevé que celui indiqué par le tableau.

Les valeurs de CV qui apparaissent au tableau A sont dérivées d'un modèle établi à partir des données de l'échantillon de l'EPA de 2001, 2002, 2003 et de la majeure partie de 2004. Il faut bien se rappeler que ces valeurs ne sont que des approximations. Pour une mesure plus précise de la variabilité, veuillez composez notre numéro sans frais 1 866 873-8788 ou communiquez avec nous par courriel à travail@statcan.ca.

Variabilité des estimations annuelles

Pour déterminer le CV approximatif d'une estimation d'une moyenne annuelle, veuillez consulter le tableau B, lequel présente l'estimation en fonction de la région géographique et du CV. Les lignes correspondent à la région géographique à laquelle l'estimation est associée et les colonnes, au degré de précision exprimé en CV, selon la grandeur de l'estimation. Pour déterminer le CV d'une estimation X dans la région A, suivez la ligne de la région A jusqu'à l'estimation la plus proche de X sans la dépasser. Le titre de la colonne indiquera le CV approximatif. Par exemple, pour connaître l'erreur d'échantillonnage d'une estimation annuelle de 39 800 chômeurs à Terre-Neuve-et-Labrador, on constate que l'estimation inférieure la plus proche, de 23 500, donne un CV de 2,5 %. L'estimation de 39 800 chômeurs à Terre-Neuve-et-Labrador aura donc un CV d'environ 2,5 %.

Tableau B: CV pour les estimations* de moyennes annuelles, Canada et provinces

Région géographique	Coefficient de variation								
	1,0 %	2,5 %	5,0 %	7,5 %	10,0 %	15,0 %	20,0 %	25,0 %	30,0 %
Canada	424,4	123,8	56,6	33,2	17,5	10,3	6,9	5,0	3,8
Terre-Neuve-et-Labrador	82,6	23,5	10,6	6,1	3,2	1,9	1,2	0,9	0,7
Île-du-Prince-Édouard	20,1	6,2	3,0	1,8	1,0	0,6	0,4	0,3	0,2
Nouvelle-Écosse	99,5	29,1	13,4	7,9	4,2	2,4	1,6	1,2	0,9
Nouveau-Brunswick	75,2	21,6	9,8	5,7	3,0	1,8	1,2	0,8	0,6
Québec	307,3	92,3	44,8	27,0	13,7	8,2	5,6	4,1	3,1
Ontario	335,4	102,1	50,0	30,3	15,5	9,4	6,4	4,7	3,6
Manitoba	82,0	24,2	11,7	7,0	3,5	2,1	1,4	1,0	0,8
Saskatchewan	70,1	20,0	9,4	5,5	2,7	1,6	1,1	0,8	0,6
Alberta	205,1	60,5	29,1	17,4	8,7	5,2	3,5	2,6	1,9
Colombie-Britannique	246,6	74,7	36,4	22,0	11,2	6,8	4,6	3,4	2,6

* Les estimations sont en milliers.

Le tableau B donne une idée approximative de la variabilité d'échantillonnage. Cette dernière est modélisée pour que le CV de l'estimation soit inférieur ou égal au CV du tableau dans environ 75 % des cas. Néanmoins, dans 25 % des cas, le CV de l'estimation sera plus élevé que celui indiqué par le tableau.

Les valeurs de CV qui apparaissent au tableau B sont dérivées d'un modèle établi à partir des données de l'échantillon de l'EPA de 2000, 2001, 2002, 2003 et de la majeure partie de 2004. Il faut bien se rappeler que ces valeurs ne sont que des approximations. Pour une mesure plus précise de la variabilité, veuillez composez notre numéro sans frais 1 866 873-8788 ou communiquez avec nous par courriel à travail@statcan.ca.

Variabilité des taux

Pour les estimations exprimées sous forme de taux ou de pourcentages, la variabilité d'échantillonnage dépend de la variabilité du numérateur et du dénominateur du rapport. Les divers taux donnés sur le disque compact sont traités différemment, car certains dénominateurs sont des valeurs étalonnées pour lesquelles il n'y a pas de variabilité d'échantillonnage.

Taux de chômage

Dans un groupe de personnes, le taux de chômage est défini comme le rapport du nombre de chômeurs, soit X, sur le nombre total de personnes dans la population active, soit Y. Le groupe peut désigner une province ou une RMR et(ou) un groupe d'âge-sexe. Par exemple, en septembre 2004, on dénombrait environ 40 200 chômeurs à Terre-Neuve-et-Labrador, tandis que la population active de la province comptait 253 700 personnes, ce qui donnait un taux de chômage de 15,8 %.

Pour déterminer le CV du taux de chômage, on se sert de la formule suivante :

$$CV(X/Y) = \sqrt{[CV(X)]^2 - [CV(Y)]^2}$$

où CV(X) est la valeur du CV associée au nombre total de chômeurs d'un sous-groupe régional ou démographique et CV(Y), la valeur du CV associé au nombre total de personnes faisant partie de la population active dans le même sous-groupe. Pour reprendre l'exemple de Terre-Neuve-et-Labrador, le CV du taux de chômage de 15,8 % serait égal à $\sqrt{5^2 - (1)^2} = 4,9\%$; le CV associé aux estimations mensuelles de 40 200 et de 253 700 à Terre-Neuve-et-Labrador est respectivement de 5 % et de 1 %, selon le tableau A.

Taux d'activité et taux d'emploi

Le taux d'activité représente la population active exprimée en pourcentage de la population totale. Le taux d'emploi est le nombre total de personnes occupées divisé par la population totale. Le numérateur et le dénominateur de ces deux rapports se rapportent au même groupe régional et démographique.

Les estimations de la population de l'EPA à l'échelle du Canada, des provinces, des RMR et par groupes d'âge-sexe ne sont pas sujettes à la variabilité d'échantillonnage, car elles sont étalonnées selon des sources indépendantes. Par conséquent, le CV correspondant du taux d'activité et du taux d'emploi est équivalent au CV du numérateur.

On appelle «domaine» les sous-groupes se trouvant à l'intérieur du Canada, des provinces et des groupes d'âge-sexe. Ainsi, les personnes occupées dans le secteur de l'agriculture au Manitoba forment un domaine. Pour déterminer le CV des taux se rapportant à un domaine, on doit tenir compte de la variabilité du numérateur et du dénominateur, car le dénominateur n'est plus un total contrôlé. Il est donc sujet à la variabilité d'échantillonnage. On peut calculer le CV associé au taux d'activité et au taux d'emploi dans un domaine de la même façon que celle montrée précédemment pour le taux de chômage. Les totaux figurant au numérateur et au dénominateur d'un taux en particulier devraient se rapporter au même domaine ou au même sous-groupe.

Variabilité de l'estimation d'un changement

La différence entre des estimations provenant de deux périodes différentes donne une estimation du changement, laquelle est sujette aussi à la variabilité d'échantillonnage. L'estimation d'un changement entre deux années ou entre deux mois repose sur deux échantillons qui peuvent avoir des ménages en commun. Ainsi, le CV du changement dépend à la fois du CV des estimations observées à chacune des deux périodes et du chevauchement entre les échantillons des deux périodes, ρ . La formule qui suit peut être utilisée pour calculer le CV de l'estimation du changement :

$$CV(Y_2 - Y_1) = \sqrt{1 - \rho} \frac{\sqrt{Y_1^2 CV(Y_1)^2 + Y_2^2 CV(Y_2)^2}}{(Y_2 - Y_1)}$$

où Y_1 et Y_2 correspondent à l'estimation respective des deux périodes; la valeur de ρ est 0,5 pour un changement entre deux mois consécutifs et zéro pour un changement couvrant toute autre période. Lorsqu'on compare les moyennes annuelles de deux années, les CV des estimations annuelles devraient être dérivées à partir du tableau B. Pour un changement entre deux mois, les estimations désaisonnalisées devraient être utilisées, conjointement avec les CV des estimations mensuelles dérivées à partir du tableau A. Veuillez prendre note que la formule ci-dessus produit des estimations approximatives de la variabilité de l'échantillonnage associées à l'estimation d'un changement.

Comment interpréter la valeur du CV

Les Enquêtes-ménages de Statistique Canada utilisent généralement les lignes directrices et les catégories de fiabilité suivantes pour interpréter les valeurs de CV pour l'exactitude des données et la diffusion de l'information statistique. Ces lignes directrices sont transmises à titre de référence seulement.

Catégorie 1 - Si le CV est $\leq 16,5\%$ - Aucune restriction de diffusion: Les données sont suffisamment exactes, si bien qu'un avertissement particulier aux utilisateurs ou d'autres restrictions ne sont pas nécessaires.

Catégorie 2 - Si le CV est $> 16,5\%$ et $\leq 33,3\%$ - Diffusion avec mise en garde: Les données sont potentiellement utiles pour certaines fins, mais devraient être accompagnées d'un avertissement aux utilisateurs concernant leur exactitude.

Catégorie 3 - Si le CV est $> 33,3\%$ - Diffusion déconseillée: Les données contiennent un niveau d'erreur élevé au point qu'elles ne devraient pas être diffusées dans la plupart des circonstances afin d'éviter de tromper les utilisateurs. Si les utilisateurs insistent pour inclure les données de la catégorie 3 dans un produit non normalisé, même après avoir été informés de leur exactitude, les données devraient être accompagnées d'un avis de non-responsabilité. L'utilisateur devrait reconnaître les mises en garde reçues et s'engager à ne pas diffuser, présenter ni déclarer les données, directement ou indirectement, sans cet avis de non-responsabilité.