

## Enquête sur le milieu de travail et les employés – Exactitude des données

Bien que des efforts considérables soient faits pour assurer un haut standard de qualité à travers toutes les étapes de l'enquête, les estimations sont inévitablement entachées d'une certaine marge d'erreur. Cela vaut pour toutes les enquêtes. L'erreur totale dans l'enquête se divise en deux grandes composantes : l'erreur d'échantillonnage et les erreurs non dues à l'échantillonnage. L'erreur d'échantillonnage vient de ce que les estimations sont établies uniquement à partir d'un échantillon de l'ensemble de la population plutôt que d'un recensement intégral, tandis que les erreurs non dues à l'échantillonnage viennent de toutes les autres causes, comme l'imperfection de la base de sondage, les erreurs de mesure ou la non-réponse. Par exemple, les erreurs de mesure peuvent venir d'erreurs commises par les répondants ou les intervieweurs à l'étape de la collecte des données, d'erreurs commises pendant la saisie au clavier, ou d'autres sources. Ce type d'erreur peut être à l'origine de l'imputation de valeurs cohérentes mais pas nécessairement correctes.

Dans l'EMTE, l'erreur d'échantillonnage et une partie des erreurs de non-réponse et de base de sondage sont traitées par l'attribution d'un poids d'estimation, appelé le poids final, à chaque unité échantillonnée (emplacement ou employé) pour laquelle nous avons des données, imputées ou pas. Le reste de l'erreur de non-réponse est traité par l'imputation des données manquantes. L'étape du contrôle de l'enquête a pour objet de limiter l'effet des erreurs de mesure. Elle comporte la détection des valeurs aberrantes et différentes étapes de validation. La démarcation entre le contrôle et la qualité des données est floue. La première est là pour garantir la seconde.

S'il n'y avait pas d'erreur non due à l'échantillonnage, la stratégie de pondération ferait que les estimations ne seraient à peu près pas entachées d'un biais par rapport au plan de sondage, au sens que l'espérance de l'erreur sur l'ensemble des échantillons possibles de l'enquête serait à peu près égale à zéro. Pour évaluer la qualité d'une estimation et obtenir des inférences valides, on calcule habituellement des mesures de précision, comme le coefficient de variation estimé. Le coefficient de variation estimé est défini comme la racine carrée de la variance estimée d'une estimation sur l'estimation elle-même. La variance est la variabilité hypothétique des estimations prises sur tous les échantillons qui auraient pu être tirés selon le plan d'échantillonnage. Puisque, dans la pratique, un seul échantillon est choisi, la variance est inconnue. Cependant, on peut l'estimer à l'aide d'un seul échantillon (l'EMTE utilise la technique du « bootstrap » moyen), ce qui permet d'obtenir les mesures de précision désirées. Noter qu'à des coefficients de variation plus petits correspondent des estimations de meilleure qualité.

On a conçu l'échantillon EMTE pour pouvoir estimer les totaux avec efficacité au niveau industrie-région-taille compte tenu du budget disponible. Les coefficients de variation projetés -- qui sont une mesure de la qualité (des valeurs plus basses conviennent mieux) -- étaient d'environ 5 % pour l'industrie et de 10 % pour l'industrie par région dans le cas des variables en étroite corrélation avec l'emploi. Lorsqu'on produit des estimations, on les met en comparaison selon ce qu'on prévoit comme précision. Dans une proportion approximative de 60 %, les estimations de totaux excédaient ce qui était prévu et, dans une autre proportion de 25 %, se situaient en deçà de la valeur limite de 33 % pour le critère de « publiabilité » à Statistique Canada. La proportion restante (15 %) n'était pas publiable suivant les normes. Il s'agissait en majeure partie d'estimations qui n'étaient pas en étroite corrélation avec l'emploi. Toutes les estimations appartenant à la catégorie « non publiable » ont fait l'objet d'une validation. Toute estimation possédant un cv entre 25 à 33 % est annotée d'un avertissement indiquant une estimation dont le taux de variabilité est relativement élevé.