

L'ENQUÊTE SUR LE MILIEU DE TRAVAIL ET LES EMPLOYÉS – DÉTECTION DES ERREURS

Vérification des données

L'outil de saisie de données d'ITAO (interview assistée par ordinateur) pour les employeurs effectue les vérifications de la validité, de l'étendue et entre les champs. Ce sont là les types de vérifications effectués durant la collecte des données du premier cycle. On a élaboré un ensemble approprié de vérifications longitudinales pour les cycles subséquents. La majorité des vérifications entre les champs est confinée à un seul bloc du contenu. En cas de rejet à la vérification entre blocs, on demande alors au répondant principal de confirmer l'information. Le calcul obligatoire d'un total positif des dépenses annuelles est un exemple de vérification de la validité. La vérification de l'étendue correspondante exige que les dépenses ne dépassent pas une limite supérieure élevée. Une vérification entre les champs connexes pour le total des dépenses annuelles offre l'assurance que la somme de la rémunération brute totale et des dépenses totales en avantages sociaux ne dépasse pas le total des dépenses annuelles.

L'application d'ITAO pour les employés effectue les vérifications de la validité, de l'étendue, entre les champs et historique. On règle les cas de rejet à la vérification durant l'interview téléphonique.

Détection des erreurs aberrantes

Une collecte de données ITAO réduit largement le nombre d'erreurs de réponse et d'erreurs typographiques. Le système prévoit une validation des données de base et une vérification des relations connues (il faut, par exemple, que l'addition de l'emploi à plein temps et de l'emploi à temps partiel ne dépasse pas l'emploi global). Pour détecter les erreurs qui auraient échappé à l'application ITAO, on soumet les réponses douteuses à une micro-analyse et à une macro-analyse afin de sauvegarder la cohérence des données.

Au niveau global, on examine les dix principaux apports aux estimations respectives, ainsi que les enregistrements où l'estimation accuse une variation relativement importante d'une année à l'autre, qu'elle soit positive ou négative. Les deux techniques sont liées, puisqu'un apport particulièrement important à une estimation peut aussi être la cause de son ample variation. Pour rendre l'analyse plus efficiente, on calcule l'apport prévu d'une unité à une estimation en se reportant à l'emploi déclaré. On compare ensuite l'apport prévu à l'apport observé. À l'aide d'un test, on établit si la différence est significative entre les deux. Cette méthode donne de bons résultats dans le cas de variables qui sont en bonne corrélation avec l'emploi. Elle indique bien les problèmes possibles même pour des variables qui sont plus faiblement en corrélation avec l'emploi.

Si on détecte d'importantes variations d'une année à l'autre dans les estimations, on regarde tous les enregistrements correspondants. Souvent, la variation sera bien réelle si le secteur connaît une période de croissance ou de décroissance marquée. Il n'y a pas d'enregistrement qui, à lui seul, apporte beaucoup à l'estimation, mais si de petites variations s'accumulent, cela peut donner une variation considérable. La macro-analyse est unidimensionnelle et pourrait donc ne pas déceler les problèmes entre variables.

Au niveau des microdonnées, une routine de détection univariée des valeurs aberrantes est appliquée à tous les répondants (complets ou partiels) avant l'imputation. La détection des valeurs aberrantes est effectuée sur des variables individuelles ou des rapports de variables, et ce, par méthode transversale et longitudinale. La méthode utilisée pour la détection des valeurs aberrantes normalise la ou les variables d'intérêt en soustrayant un paramètre de position et en divisant par un paramètre d'échelle. Dans l'EMTE, le paramètre de position est la médiane et le paramètre d'échelle est l'écart interquartile. Ce type de détection des valeurs aberrantes s'effectue pour les lieux de travail au niveau des microdonnées. La sensibilité du procédé peut être modifiée pour répondre aux besoins de l'enquête.

Pour pouvoir appliquer avec succès le programme de détection des valeurs aberrantes aux données des enquêtes-entreprises, on doit respecter les deux critères de a) l'homogénéité et b) de la symétrie des données. Avec des données homogènes, on n'a plus à utiliser la pondération d'échantillonnage lorsqu'on combine des strates voisines pour que le programme de détection des valeurs aberrantes ait une « résolution » plus fine. L'homogénéité des données diminue l'effet du plan d'échantillonnage et les problèmes complexes d'une constatation des observations aberrantes dans un échantillon prélevé sur une population finie se ramènent au problème bien plus simple du traitement des valeurs extrêmes pour une population infinie. On parvient à l'homogénéité en soumettant une ou plusieurs variables à une transformation appropriée. On vérifie ensuite la symétrie approximative des données ainsi transformées.