

## Enquête sur la gestion agroenvironnementale – 2006 – Exactitude des données

Bien qu'il soit difficile de quantifier les erreurs non dues à l'échantillonnage, on peut estimer les erreurs d'échantillonnage à partir de l'échantillon à l'aide d'une mesure statistique appelée l'erreur-type (ET) d'une valeur estimée également appelée erreur d'échantillonnage absolue. Pour les estimations de niveau (par ex. totaux ou moyennes), une mesure statistique appelée le coefficient de variation (CV) est normalement utilisée. Le CV, soit l'erreur-type divisée par l'estimation de l'enquête, est une mesure de précision en termes relatifs et est exprimé en pourcentage.

Pour des estimations de niveaux, le CV est la mesure convenable de l'erreur d'échantillonnage. Cependant, dans le cas de proportions, il est préférable d'utiliser une erreur d'échantillonnage absolue telle que l'ET. Dans le cas de l'Enquête sur la gestion agroenvironnementale, la plupart des estimations portent sur des proportions.

L'ET (qui dépend de la taille de la population, de la taille de l'échantillon et de l'estimation), jumelée avec un niveau de confiance peuvent être utilisés pour calculer la marge d'erreur. Cette mesure est assez simple à interpréter puisqu'elle est établie sur la même base que la valeur estimée. Par exemple, une proportion estimée de 80% peut avoir une marge d'erreur de 3 %, ce qui nous amènera à conclure (avec un niveau de confiance approprié, habituellement 19 fois sur 20) que la vraie proportion est entre 77% et 83%.

Supposons que nous voulons estimer la proportion des fermes canadiennes d'élevage qui entreposent du fumier liquide. La proportion estimée est de 14 % avec une erreur-type de 2,54. On peut en déduire que la proportion des fermes qui n'entreposent pas de fumier liquide est de 86 % et que la qualité de l'estimation est la même (c.-à.-d. que l'erreur-type est toujours de 2,54). L'erreur-type est une erreur absolue qui s'applique aux estimations de 14 % et de 86 %. Étant une erreur relative, le CV est différent dans les deux estimations. Il peut même sembler positif pour une proportion (86 % pour le CV1) et négatif pour la proportion complémentaire (14 % pour le CV2) tel qu'illustré ci dessous :

$CV1 = 100 * 2,54/86 = 3$  (pour les fermes qui n'entreposent pas de fumier liquide)

$CV2 = 100 * 2,54/14 = 18$  (pour les fermes qui entreposent du fumier liquide)

Bien que la qualité des estimations est la même, le CV2 sous-entend que la qualité de la proportion estimée des fermes qui entreposent du fumier liquide est bien inférieure. Dans ce cas, comme pour toutes les estimations de proportions, le CV peut être trompeur.

Voici les cotes de CV proposées pour des estimations de niveaux et des cotes d'erreur-type (ET) pour des estimations de proportions :

Cote de CV

0,01 % à 4,99 % A -- excellent

5,0 % à 9,99 % B -- très bon  
 10,0 % à 14,99 % C -- bon  
 15,0 % à 24,99 % D -- acceptable  
 25,0 % à 34,99 % E -- à utiliser avec prudence  
 35,0 % et plus F – trop peu fiable pour être publié

#### Cote d'ET

0,01 % à 2,49 % A -- excellent  
 2,5 % à 4,99 % B -- très bon  
 5,0 % à 7,49 % C -- bon  
 7,5 % à 12,49 % D -- acceptable  
 12,5 % à 17,49 % E -- à utiliser avec prudence  
 17,5 % et plus F – trop peu fiable pour être publié

La marge d'erreur (appelée "e") se calcule comme suit:

$$e = z * SE = z * \sqrt{\left(1 - \frac{n}{N}\right) * \frac{(p)(1-p)}{n}}$$

ou "z" est le percentile de la répartition normalisée qui correspond à un niveau de confiance désiré, "n" est la taille de l'échantillon, "N" est la taille de la population, et "p" est la proportion estimée.

L'exemple suivant illustre le lien entre l'ET et la marge d'erreur. Le niveau de confiance de 95 % est habituellement choisi, ce qui implique que "z" prend une valeur approximative de 1.96. Supposons un échantillon de 450 fermes, choisies parmi une population de 5 000 fermes, et une proportion estimée de 75 %.

L'erreur-type est  $ET = \sqrt{\left(1 - \frac{450}{5000}\right) * \frac{(0.75)(0.25)}{450}} = 0.0195$  (ou 1,95 %), ce qui donne la

cote "excellent" pour une ET. En utilisant cette information pour déterminer la marge d'erreur, on obtient:

$$e = (1.96)(0.0195)$$

$$e = 0.0382$$

Nous pouvons donc conclure, 19 fois sur 20, que la vraie proportion sera à un intervalle de 3,82 % de l'estimation: soit entre 71,18 % et 78, 82%.