Farm Environmental Management Survey (FEMS) – 2006 – Data Accuracy

While non-sampling errors are difficult to quantify, sampling errors can be estimated from the sample itself using the standard error (SE) of estimated values also referred to as an absolute sampling error. For level estimates (e.g. totals and averages), a statistical measure called the coefficient of variation (CV) is normally used. The CV, defined as the standard error divided by the survey estimate, is a measure of precision in relative terms and is expressed as a percentage.

For level estimates, the CV is the appropriate measure of the sampling error. For proportions, however, an absolute sampling error – such as the SE itself – is preferred. In the case of the Farm Environmental Management Survey, most of the estimates deal with proportions.

The SE (which is a function of the population size, the sample size, and the estimate), along with the confidence level, can be used to calculate the margin of error. This measure is straightforward to interpret, since it is on the same scale as the estimate itself. For example, an estimated proportion of 80% might have a margin of error of 3%, meaning that we would conclude (with the appropriate confidence level, usually 19 times out of 20) that the true proportion is between 77% and 83%.

Suppose we want to estimate the proportion of Canadian livestock farms that store liquid manure. The estimated proportion is 14% with a standard error of 2.54. It can be deduced that the proportion of farms that do not store liquid manure is 86% and that the quality of the estimate is the same (i.e., the standard error is still 2.54). The standard error is an absolute error that applies to both the 14% and 86% estimates. The CV, being a relative error, would be different for the two estimates. It can even appear good for one proportion (86% for CV1) and bad for the complementary proportion (14% for CV2) as shown below:

CV1 = 100 * 2.54/86 = 3 (for farms which do not store liquid manure)
CV2 = 100 * 2.54/14 = 18 (for farms which store liquid manure)

Though the quality of the estimates is the same, the CV2 implies that the quality of the estimated proportion of farms which store liquid manure is much lower. In this case as with all proportion estimates, the CV can be misleading.

The following is a suggested CV rating system for level estimates, and a standard error (SE) rating system for proportion estimates:

CV Rating

0.01% - 4.99% A -- excellent
5.0% - 9.99% B -- very good
10.0% - 14.99% C -- good

15.0% - 24.99% D -- acceptable
25.0% - 34.99% E -- use with caution
35.0% and more F – too unreliable to be published

SE Rating

0.01% - 2.49% A -- excellent
2.5% - 4.99% B -- very good
5.0% - 7.49% C -- good
7.5% - 12.49% D -- acceptable
12.5% - 17.49% E -- use with caution
17.5% and more F – too unreliable to be published

The margin of error (called “e”) is calculated as follows:

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

where “z” is the percentile of the standard normal distribution that corresponds to the desired confidence level, “n” is the sample size, “N” is the population size, and “p” is our estimated proportion.

Consider an example to illustrate the link between the SE and the margin of error. The desired confidence level is typically set to 95%, meaning “z” takes a value of approximately 1.96. Suppose we have a sample of 450 farms, selected from a population 5000 farms, and we estimate a proportion of 75%.

The standard error is $SE = \sqrt{\left(1 - \frac{450}{5000}\right)} * \sqrt{\frac{(0.75)(0.25)}{450}} = 0.0195$ (or 1.95%), so we are within the “excellent” range for the SE. If we use this information to build a margin of error, we get:

$$e = (1.96)(0.0195)$$
$$e = 0.0382$$

We thus conclude that, 19 times out of 20, the true proportion will be within 3.82% of our estimate: that is, between 71.18% and 78.82%.