A

Appendix A: Sampling

A company needing to collect a large quantity of data for a particular scope 3 category may find it impractical or impossible to collect the data from each activity in the category. In such cases, companies may use appropriate sampling techniques to extrapolate data from a representative sample of activities within the category.

Companies may also choose to categorize activities into similar groups for data collection. This strategy should group activities with similar anticipated emissions intensities. For example:

- Companies with a large number of leased assets (Categories 8 and 13) or franchises (Category 14) may group buildings by building type or floor area and vehicles by vehicle type
- Companies with a large number of employees collecting data on employee commuting (Category 7) may wish to extrapolate data from a representative sample of employees
- Companies with a large number of distribution channels may use sampling when calculating the emissions associated with Categories 4 and 9 (Transportation and Distribution).

Companies should choose a sampling method that aligns with their business goals and document and justify their choice. The choice of sampling method will depend on factors including, but not limited to:

- Available resources
- Number of data points
- Expected level of homogeneity between samples
- Geographical spread of data points
- Ease of data collection
- Timeframe available.

Ultimately, the use of sampling and choice of a specific sampling method aims to optimize the trade-off between cost and accurately representing all emission sources in the scope 3 category. Companies may use a variety of sampling methods, as appropriate for each specific emissions activity.

Sampling methods

Sampling methods available to companies include, but are not limited to:

- Simple random sampling
- Systematic sampling
- Stratified sampling

Each approach is summarized below. Alternative methods for sampling may also be used.

Simple Random Sampling

Simple random sampling involves randomly selecting activities (i.e., a sample) from a larger set of activities (i.e., the entire population).

If the total number of activities from which a sample is selected is small, simple random sampling may be performed at its most basic level by selecting activities at random. If the total number of activities is large, for example with hundreds or thousands of activities, then random sampling is better performed by computer.

Advantages of simple random sampling include:

- With an appropriate sample size, simple random sampling creates a representative view of the entire population. (For example, if a company has fifty employees located within a close geographical area and wants to determine the average commuting distance, it may choose to collect data from ten randomly selected employees as a representative sample.)
- It is relatively straightforward to construct the sample.

Disadvantages of simple random sampling include:

- The sample size needed to generate appropriately representative results may be prohibitively large and cumbersome to sample. (For example, if a retail organization has thousands of stores in many countries, randomly selecting individual stores may result in a difficult and time-consuming data collection process.)
- It may not be possible to obtain a complete list of all activities from the sample size, which is a prerequisite for simple random sampling. (For example, if a distribution company wants to determine the average backhaul capacity of its trucks, it would have to list every journey before a random sample could be selected.)

Systematic Sampling

Systematic sampling involves randomly selecting the first item to sample and then selecting subsequent activities at regular intervals.

An appropriate sampling interval should be chosen so that the company achieves the desired sample size. For example, if a company sourced agricultural products from 100 farms but only wanted to sample 20 farms, an appropriate sampling interval would be every 5 farms. If the first farm to be sampled was picked as Farm 3, the company would subsequently sample from Farms 8, 13, 18, 23,…, 93, 98.

Calculation formula [A.1] Selecting an appropriate systematic sampling interval

Systematic sampling interval =
sampling interval = total population size / desired sample size

Advantages of systematic sampling include:

- Simple to implement
- The population is guaranteed to be evenly sampled without risk that the sample points are clustered together.

Disadvantages of systematic sampling include:

- If there is a periodic pattern in the population to be sampled, it could lead to biased sampling
- As with simple random sampling, it may not be possible to obtain a complete list of all activities in the population.

Stratified Sampling

Stratified sampling initially groups the population's activities into categories with similar characteristics. Random sampling is subsequently performed within these homogeneous groups.

The company should initially create population groups containing activities with characteristics likely to offer similar intensities of GHG emissions. Grouping variables could include location, size, building type, manufacturing technique, age, etc.

For example, if an agricultural produce company was assessing emissions from its farms, it may use the following variable to create initial groupings of all farms: high / low rainfall; smaller than 100 hectares / larger than 100 hectares; north-facing-hill / south-facing-hill / neither.

Stratified sampling is particularly useful when the variability in GHG emissions within groups is small, but the variability between groups is large.

Advantages of stratified sampling:

- Can lead to higher precision because there is less variability within the groups given that similar characteristics are grouped together.
- The necessary sample size can be reduced due to lower variability within groups, therefore saving time and money.
- Allows companies to draw insights into the source and level of emissions among different groups. This level of detail may be lost with simple random sampling.
- Different random sampling techniques may be employed for different groups as appropriate.

Disadvantages of stratified sampling:

- Identifying appropriate variables and forming sampling groups may be difficult and complex.

Sample Size

Determining sample size is fundamental to any sampling activity. The choice of sample size will be influenced by several factors, including the likely significance of GHG emissions from the sources in question, the size of the population, the variability of the emission sources, and the necessary degree of precision.

Determining sample size

There are several approaches to determining sample size. In particular, four alternative approaches may prove useful for companies:

- Using the sample size of a similar inventory
- Using online calculators
- Using published tables
- Using formulas.

Using the sample size of a similar inventory

Companies may refer to similar inventories for guidance on appropriate sample size and sampling technique. When using this approach, companies should justify the similarity and appropriateness for the comparison. Companies may refer to similar inventories that have been externally verified for guidance on appropriate sample size and sampling technique.

Using online calculators

Online calculators are a quick and easy way to assess sample size.

For example:

- http://www.research-advisors.com/tools/SampleSize.htm provides a downloadable spreadsheet to calculate necessary sample size with the ability to tailor the sampling criteria.
- http://www.surveysystem.com/sscalc.htm provides an interactive online calculator for sample size; however, the choices for confidence level are fixed.

Using published tables

Many published tables give the necessary sample size for a specific set of criteria. Such criteria include precision, confidence levels, and variability for a given population size.

Users should refer to standard statistics texts or search online for a table matching their specific sampling criteria.

Using formulas

Companies that want greater assurance for their choice of sample size may turn to established formulas. Formulas for the calculation of sample size are available in all standard statistics and sampling text books, as well as via the internet.

When applying sample size formulas, users may find it advantageous to seek the advice of a person with experience of statistics.

Level of accuracy

The level of accuracy is related to the sample size, sampling strategy, and the measurement system. Assuming a normal distribution, increasing the sample size is likely to reduce the sampling error using the relationship $v = \sqrt{n}$. In this relationship "v" represents the variability of the data values. It is important to recognize that all measurements contain some level of uncertainty. An estimate of the measurement uncertainty should be obtained, particularly for parts of the assessment that contribute significantly to the organization and/ or if subsequent investment decisions are made based on the measurement.

Confidence level

An estimate of the uncertainty, which should include both precision and bias from random error and systematic error respectively, will enable an interpretation of the measurement. For example, a level of uncertainty of ±5 percent would imply for an emissions estimate of 100 tonnes CO_2e, that the actual emissions lie somewhere between 95 and 105 tonnes CO_2e. The confidence level associated with the uncertainty normally corresponds to a 95 percent confidence level, that is, 2 standard deviations. For example, the true value lies in the range of 95 and 105 tonnes with 95 percent confidence.

Variability

Variability refers to the degree of difference between activities within the population. A population that is more heterogeneous (more variable) will require a larger sample size. A variability of 50 percent is the maximum level of variability in a population. Therefore, a variability assumption of 0.5 is often used as a conservative estimate.