International Performance Measurement and Verification Protocol

Statistics and Uncertainty for IPMVP

Prepared by Efficiency Valuation Organization
www.evo-world.org

June 2014

EVO 10100 – 1:2014

EVO Vision

A global marketplace that correctly values the efficient use of natural resources and utilizes end-use efficiency options as a viable alternative to supply options

EVO Mission

To develop and promote the use of standardized protocols, methods and tools to quantify and manage the performance risks and benefits associated with end-use energy-efficiency, renewable-energy, and water-efficiency business transactions

Contents   Page

1 Introduction

The objective of M&V is to reliably determine energy savings. In order for savings reports to be reliable, they need to have a reasonable level of uncertainty. The uncertainty of a savings report can be managed by controlling random errors and data bias. Random errors are affected by the quality of the measurement equipment, the measurement techniques, and the design of the sampling procedure. Data bias is affected by the quality of measurement data, assumptions and analysis. Reducing errors usually increases M&V cost so the need for improved uncertainty should be justified by the value of the improved information.

Energy savings computations involve a comparison of measured energy data, and a calculation of "adjustments" to convert both measurements to the same set of operating conditions. Both the measurements and the adjustments introduce error. Errors may arise, due to meter inaccuracy, sampling procedures or adjustment procedures. These processes produce statistical "estimates" with reported or expected values, and some level of variation. In other words, true values are not known, only estimates with some level of uncertainty. All physical measurement and statistical analysis is based on estimation of central tendencies, such as mean values, and quantification of variations such as range, standard deviation, standard error, and variance.

Statistics is the body of mathematical methods that can be applied to data to help make decisions in the face of uncertainty. Statistics provide ways of checking results to see if the reported savings are "significant," i.e. likely to be a real effect of the ECM rather than random behavior.
Errors occur in three ways: modeling, sampling, and measurement.

- **Modeling**. Errors in mathematical modeling due to inappropriate functional form, inclusion of irrelevant variables, exclusion of relevant variables, etc.
- **Sampling**. Sampling error arises when only a portion of the population of actual values is measured, or a biased sampling approach is used. Representation of only a portion of the population may occur in either a physical sense or in the time sense.
- **Measurement**. Measurement errors arise from the accuracy of sensors, data tracking errors, drift since calibration, imprecise measurements, etc. The magnitude of such errors is largely given by manufacturer's specifications and managed by periodic re-calibration.

This document gives guidance on quantifying the uncertainties created by these three forms of error. Some sources of error are unknown and unquantifiable:

- ➢ poor meter selection or placement,
- ➢ inaccurate estimates in Option A, or
- ➢ mis-estimation of interactive effects in Options A or B.

Unknown or unquantifiable uncertainties can only be managed by following industry best practices.

1.1 Expressing Uncertainty

In order to communicate savings in a statistically valid manner, savings need to be expressed along with their associated confidence and precision levels. Confidence refers to the likelihood or probability that the estimated savings will fall within the precision range.[1]

EXAMPLE The savings estimation process may lead to a statement such as: "the best estimate of savings is 1,000 kWh annually (point estimate) with a 90% probability (confidence) that the true-average savings value falls within ±20% of 1,000." A graphical presentation of this relationship is shown in Figure 1.

[1] Selected statistical terms are defined in section 1.3

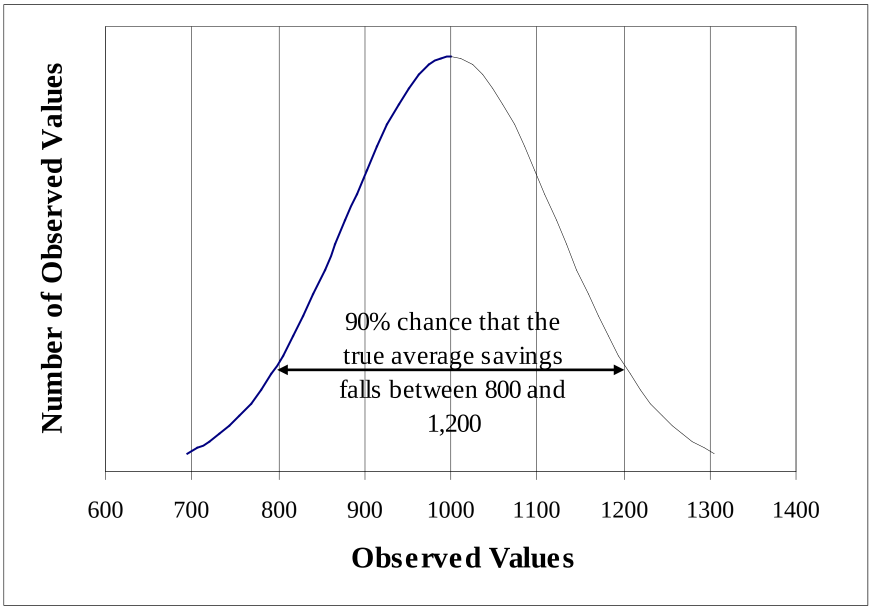

Figure 1 Normally Distributed Population

A statistical precision statement (the ±20% portion) without a confidence level (the 90% portion) is imprecise. The M&V process may yield extremely high precision with low confidence.

Example: The savings may be stated with a precision of ±1%, but the associated confidence level may drop from 95% to 35%.

1.2 Acceptable Uncertainty

Savings are deemed to be statistically valid if they are large relative to the statistical variations. Specifically, the savings need to be larger than twice the standard error of the baseline value. If the variance of the baseline data is excessive, the unexplained random behavior in energy use of the facility or system is high, and any single savings determination is unreliable.

Where you cannot meet this criterion, consider using:

- more precise measurement equipment,
- more independent variables in the mathematical mode,
- larger sample sizes, or
- an IPMVP Option that is less affected by unknown variables.

1.3 Definitions of Statistical Terms

Sample Mean ($\bar{Y}$):determined by adding up the individual data points (Y_i) and dividing by the total number of these data points (n), as follows:

$$\bar{Y} = \frac{\sum Y_i}{n} \qquad (1)$$

Sample Variance (S^2): Sample variance measures the extent to which observed values differ from each other, i.e., variability or dispersion. The greater the variability, the greater the uncertainty in the mean. Sample variance is found by averaging the squares of the individual deviations from the mean. The reason these deviations from the mean are squared is simply to eliminate the negative values (when a value is below the mean) so they do not cancel out the positive values (when a value is above the mean). Sample variance is computed as follows:

$$S^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} \qquad (2)$$

Sample Standard Deviation (s): This is simply the square root of the sample variance. This brings the variability measure back to the units of the data (e.g., if the variance units are $(kWh)^2$, the standard deviation units would be kWh).

$$s = \sqrt{S^2} \qquad (3)$$

Sample Standard Error (SE): This is the sample standard deviation divided by $\sqrt{n}$. This measure is used in estimating precision of a sample mean. It is also denoted as $\bar{s}$, or the "sample standard deviation of the mean" in most statistics textbooks.

$$SE = \frac{s}{\sqrt{n}} \qquad (4)$$

Sample Standard Deviation of the Total (s_{tot}): Many times we are interested in the statistical properties of a total rather than a mean. The sample standard deviation of a total is used to define the precision about a sample total. It is defined as the square root of the sample size, $\sqrt{n}$ times the sample standard deviation:

$$S_{tot} = \sqrt{n} \bullet S \qquad (5)$$

Coefficient of Variation (cv): The coefficient of variation is simply the standard deviation of a distribution expressed as a percentage of the mean. For instance, the cv of a sample total would be the [stot] ÷ [sample total]; the cv of a sample mean would be the [SEȲ] ÷ [sample mean]; etc. The general formula is:

$$cv = \frac{s}{\bar{Y}} \qquad (6)$$

Precision: Precision is the measure of the absolute or relative range within which the true value is expected to occur with some specified level of confidence. Confidence level refers to the probability that the quoted range contains the estimated parameter.

Absolute precision is computed from sample standard error using a “t” value from a “t-distribution” Table. A t-distribution table is provided below, but can be found in statistic tables, books or on-line resources.

$$t \bullet SE_{\bar{Y}} \quad (7)$$

Table 1 t-Table

Degrees of Freedom **DF**	Confidence Level				Degrees of Freedom **DF**	Confidence Level			
	95%	90%	80%	50%		95%	90%	80%	50%
1	12.71	6.31	3.08	1.00	16	2.12	1.75	1.34	0.69
2	4.30	2.92	1.89	0.82	17	2.11	1.74	1.33	0.69
3	3.18	2.35	1.64	0.76	18	2.10	1.73	1.33	0.69
4	2.78	2.13	1.53	0.74	19	2.09	1.73	1.33	0.69
5	2.57	2.02	1.48	0.73	21	2.08	1.72	1.32	0.69
6	2.45	1.94	1.44	0.72	23	2.07	1.71	1.32	0.69
7	2.36	1.89	1.41	0.71	25	2.06	1.71	1.32	0.68
8	2.31	1.86	1.40	0.71	27	2.05	1.70	1.31	0.68
9	2.26	1.83	1.38	0.70	31	2.04	1.70	1.31	0.68
10	2.23	1.81	1.37	0.70	35	2.03	1.69	1.31	0.68
11	2.20	1.80	1.36	0.70	41	2.02	1.68	1.30	0.68
12	2.18	1.78	1.36	0.70	49	2.01	1.68	1.30	0.68
13	2.16	1.77	1.35	0.69	60	2.00	1.67	1.30	0.68
14	2.14	1.76	1.35	0.69	120	1.98	1.66	1.29	0.68
15	2.13	1.75	1.34	0.69	∞	1.96	1.64	1.28	0.67

Note: Calculate DF using the following,

- DF = n – 1 (for a sample distribution)
- DF = n - p – 1 (for a regression model)

Where,
n = sample size
p = number of regression model variables

In general, the true value of any statistical estimate is expected with a given confidence level, to fall with the range defined by

Range = estimate ± *absolute precision* (8)

Where “estimate” is any empirically derived value of a parameter of interest (e.g., total consumption, average number of units produced, etc.).

Relative precision is the absolute precision divided by the estimate:

$$\frac{t \bullet SE}{Estimate} \qquad (9)$$

Example: Consider the data in Table 2 from 12 monthly readings of a meter, and related analysis of the difference between each reading.

Table 2 Data Analysis Example

	Actual	Computed Differences From the Mean	
	Reading	Raw	Squared
1	950	-50	2,500
2	1,090	90	8,100
3	850	-150	22,500
4	920	-80	6,400
5	1,120	120	14,400
6	820	-180	32,400
7	760	-240	57,600
8	1,210	210	44,100
9	1,040	40	1,600
10	930	-70	4,900
11	1,110	110	12,100
12	1,200	200	40,000
Total	12,000		246,600

The Mean value is: $\bar{Y} = \frac{\sum Y_i}{n} = \frac{12{,}000}{12} = 1{,}000$

The Variance is: $S^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} = \frac{246{,}600}{12-1} = 22{,}418$

The Standard Deviation is: $s = \sqrt{S^2} = \sqrt{22{,}418} = 150$

The Standard Error is: $SE = \frac{s}{\sqrt{n}} = \frac{150}{\sqrt{12}} = 43$

In Table 2, there are 12 data points. That means DF= 12-1=11. Using Table 1, for a confidence level of 90% the value for “t” is 1.80. Therefore:

the Absolute Precision is: $t \bullet SE = 1.80 \times 43 = 77$

the Relative Precision is: $\frac{t \bullet SE}{estimate} = \frac{77}{1{,}000} = 7.7\%$

So, there is 90% confidence that the true mean-monthly consumption lies in the range between 923 and 1,077 kWh. It can be said with 90% confidence that the mean value of the 12 observations is 1,000 ±7.7%. Similarly it could be said:

- with 95% confidence that the mean value of the 12 observations is 1,000 ±9.5%, or
- with 80% confidence that the mean value of the 12 observations is 1,000 ±5.8%, or
- with 50% confidence that the mean value of the 12 observations is 1,000 ±3.0%.

2 Modeling

Mathematical modeling is used in M&V to prepare the routine-adjustments term in the various versions for savings are discussed in IPMVP Core Concepts. They are summarized below:

1. Savings = (Baseline – Period Use or Demand – Reporting-Period Use or Demand) ± Adjustments
2. Savings = (Baseline Energy – Reporting-Period Energy) ± Routine Adjustments ± Non-Routine Adjustments
3. Avoided Energy Use (or Savings) = (Baseline Energy ± Routine Adjustments to Reporting Period Conditions ± Non-Routine Adjustments to Reporting Period Conditions) – Reporting-Period Energy.
4. Avoided Energy Use (or Savings)= Adjusted-Baseline Energy – Reporting-Period Energy ± Non-Routine Adjustments of Baseline Energy to Reporting –Period Conditions
5. Normalized Savings = (Baseline Energy ± Routine Adjustments to Fixed Conditions ± Non-Routine Adjustments to Fixed Conditions – (Reporting Period Energy ± Routine Adjustments to Fixed Conditions ± Non-Routine Adjustments to Fixed Conditions).
6. Option A Savings = Estimated Value x (Baseline-Period, Measured Parameter – Reporting-Period, Measured Parameter)
7. Option B Savings = Baseline Energy – Reporting-Period Energy
8. Savings = Baseline Energy from the Calibrated Model [hypothetical or without ECMs] – Reporting-Period Energy from the Calibrated Model [with ECMs]
9. Savings = Baseline Energy from the Calibrated Model [hypothetical or without ECMs] – Actual Calibration-Period Energy ± Calibration error in the Correspoinding Calibration Reading

Modeling involves finding a mathematical relationship between dependent and independent variables. The dependent variable, usually energy, is modeled as being governed by one or more independent variable(s) X_i, (also known as 'explanatory' variables). This type of modeling is called regression analysis.

In regression analysis, the model attempts to "explain" the variation in energy resulting from variations in the individual independent variables (X_i). For example, if one of the X's is production level, the model would assess whether the variation of energy from its mean is caused by changes in production level. The model quantifies the causation. For example, when production increases by one unit, energy consumption increases by "b" units, where "b" is called the regression coefficient.

The most common models are linear regressions of the form:

$$Y = b_o + b_1X_1 + b_2X_2 + ... + b_pX_p + e$$

where:

- Y is the dependent variable, usually in the form of energy use during a specific time period (e.g., 30 days, 1 week, 1 day, 1 hour, etc.)
- X_{it} (i = 1, 2, 3, … p) represents the 'p' independent variables such as weather, production, occupancy, metering period length, etc.
- b_i (i = 0, 1, 2, … p) represents the coefficients derived for each independent variable, and one fixed coefficient (b_0) unrelated to the independent variables
- e represents the residual errors that remain unexplained after accounting for the impact of the various independent variables. Regression analysis finds the set of b_i values that minimizes the sum of squared residual-error terms (thus regression models are also called least-squares models).

An example of the above model for a building's energy use is:

monthly energy consumption = 342,000 + (63 x *HDD*) + (103 x *CDD*) + (222 x Occupancy)

HDD and CDD are heating (HDD) and cooling (CDD) degree days. Occupancy is a measure of percent occupancy in the building. In this mode 342,000 is an estimate of base load in kWh, 63 measures the change in consumption for one additional HDD, 103 measures the change in consumption for one additional CDD, and 222 measures the change in consumption per 1% change in occupancy.

2.1 Modeling Errors

When using regression models, as described above, several types of errors may be introduced as listed below.

1. The model is built on values that are outside the probable range of the variables to be used. A mathematical model should only be constructed using reasonable values of the dependent and independent variables.
2. The mathematical model may not include relevant independent variables, introducing the possibility of biased relationships (omitted variable bias).
3. The model may include some variables that are irrelevant.
4. The model may use inappropriate functional form.
5. The model may be based on insufficient or unrepresentative data.

These errors are discussed in more detail in text below.

2.1.1 Using Out of Range Data

If the model is built on data that are not representative of the normal energy behavior of the facility, then the predictions may not be relied upon. This may include inclusion of outliers, or values that are well outside the range of reasonableness. Data should be screened before building the model.

2.1.2 Omission of Relevant Variables

In M&V, regression analysis is used to account for changes in energy use. Most complex energy using systems are affected by innumerable independent variables. Regression models cannot hope to include all independent variables. Even if it were possible, the model would be too complex to be useful and would require excessive data gathering activities. The practical approach is to include only independent variable(s) thought to significantly impact energy.

Omission of a relevant independent variable may be an important error. If a relevant independent variable is missing (e.g., HDD, production, occupancy), then the model will fail to account for a significant portion of the variation in energy. The deficient model will also attribute some of the variation that is due to the missing variable to the variable(s) that are included in the model. The effect will be a less accurate model.

There are no obvious indications of this problem in the standard statistical tests (except maybe a low R^2). Experience and knowledge of the engineering of the system whose performance is being measured is valuable in addressing this issue.

There may be cases where a relationship is known to exist with a variable recorded during the baseline period. However the variable is not included in the model due to lack of budget to continue to gather the data in the reporting period. Such omission of a relevant variable should be noted and justified in the M&V Plan.

2.1.3 Inclusion of Irrelevant Variables

Sometimes models include irrelevant independent variable(s). If the irrelevant variable has no relationship (correlation) with the included relevant variables, then it will have minimal impact on the model. However, if the irrelevant variable is correlated with other relevant variables in the model, it may bias the coefficients of the relevant variables.

Use caution in adding more independent variables into a regression analysis just because they are available. To judge the relevance of independent variables requires both experience and intuition. However, the associated t-statistic is one way of confirming the relevance of particular independent variables included in a model. Experience in energy analysis for the type of facility involved in any M&V program is necessary to determine the relevance of independent variables.

2.1.4 Functional Form

It is possible to model a relationship using the incorrect functional form. For example, a linear relationship might be incorrectly used in modeling an underlying physical relationship that is non-linear. For example, electricity consumption and ambient temperature tend to have a non-linear (often 'U' shaped) relationship with outdoor temperature over a one-year period in buildings that are both heated and cooled electrically. (Electricity use is high for both low and high ambient temperatures, while relatively low in mid seasons.) Modeling this non-linear relationship with a single linear model would introduce unnecessary error. Instead, separate linear models should be derived for each season.

It may also be appropriate to try higher order relationships, e.g., $Y = f(X, X^2, X^3)$.
The modeler needs to assess different functional forms and select the most appropriate among them using evaluation measures.

2.1.5 Data Shortage

Errors may also occur from insufficient data either in terms of quantity (i.e., too few data points) or time (e.g., using summer months in the model and trying to extrapolate to winter months). The data used in modeling should be representative of the range of operations of the facility. The time period covered by the model needs to include various possible seasons, types of use, etc. This may call for either extension of the time periods used or increasing sample sizes.

2.2 Evaluating Regression Models

In order to evaluate how well a particular regression model explains the relationship between energy use and independent variable(s), three tests may be performed as described below.

2.2.1 Coefficient of Determination (R2)

The first step in assessing the accuracy of a mode is to examine the Coefficient of Determination, R2, a measure of the extent to which variations in the dependent variable Y from its mean value are explained by the regression model. Mathematically, R^2 is:

$$R^2 = \frac{Explained\ \ Variation\ \ in\ \ Y}{Total\ \ Variation\ \ in\ \ Y}$$

or more explicitly:

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \qquad (10)$$

where:

- $\hat{Y}_i$ = model predicted energy value for a particular data point using the measured value of the independent variable (i.e., obtained by plugging the X values into the regression model)
- $\bar{Y}$ = mean of the n measured energy values, found using equation (1)
- Y_i = actual observed (e.g., using a meter) value of energy

All statistical packages and spreadsheet regression-analysis tools compute the value of R^2.

The range of possible values for R^2 is 0.0 to 1.0. An R^2 of 0.0 means none of the variation is explained by the mode, therefore the model provides no guidance in understanding the variations in Y (i.e., the selected independent variable(s) give no explanation of the causes of the observed variations in Y). On the other hand, an R^2 of 1.0 means the model explains 100% of the variations in Y, (i.e., the model predicts Y with total certainty, for any given set of values of the independent variable(s)). Neither of these limiting values of R^2 is likely with real data.

In general, the greater the coefficient of determination, the better the model describes the relationship of the independent variables and the dependent variable. Though there is no universal standard for a minimum acceptable R^2 value, 0.75 is often considered a reasonable indicator of a good causal relationship amongst the energy and independent variables.

The R^2 test should only be used as an initial check. Models should not be rejected or accepted solely on the basis of R^2. Finally, a low R^2 is an indication that some relevant variable(s) are not included, or that the functional form of the model (e.g., linear) is not appropriate. In this situation it would be logical to consider additional independent variables or a different functional form.

2.2.2 Standard Error of the Estimate

When a model is used to predict an energy value (Y) for given independent variable(s), the accuracy of the prediction is measured by the standard error of the estimate (SE$\widehat{Y}$). This accuracy measure is provided by all standard regression packages and spreadsheets.
Once the value(s) of independent variable(s) are plugged into the regression model to estimate an energy value ($\widehat{Y}$), an approximation of the range of possible values for $\widehat{Y}$ can be computed using equation 8 as:

$$\hat{Y} \pm t \times SE_{\hat{Y}}$$

where:

- $\hat{Y}$ is the predicted value of energy (Y) from the regression model
- t is the value obtained from the t-tables (see Table 1)
- $SE_{\hat{Y}}$ is the standard error of the estimate (prediction). It is computed as:

$$SE_{\hat{Y}} = \sqrt{\frac{\sum(\hat{Y}_i - Y_i)^2}{n - p - 1}} \qquad (11)$$

where p is the number of independent variables in the regression equation.

This statistic is often referred to as the root-mean squared error (RMSE). Dividing the RMSE by the average energy use produces the coefficient of variation of RMSE, or the CV(RMSE).

$$CV(RMSE) = \frac{SE_{\hat{Y}}}{\overline{Y}} \qquad (12)$$

A similar measure is the mean bias error (MBE) defined as:

$$MBE = \frac{\sum(\hat{Y}_i - Y_i)}{n} \qquad (13)$$

The MBE is a good indicator of overall bias in the regression estimate. Positive MBE indicates that regression estimates tend to overstate the actual values. Overall positive bias does tend to cancel out negative bias. The RMSE does not suffer from this cancellation problem.
All three measures may be used in evaluating the calibration of simulation models in Option D.

2.2.3 t-statistic

Since regression-model coefficients (b_k) are statistical estimates of the true relationship between an individual X variable and Y, they are subject to variation. The accuracy of the estimate is measured by the standard error of the coefficient and the associated value of the t-statistic. A t-statistic is a statistical test to determine whether an estimate has statistical significance. Once a value is estimated using the test, it can be compared against critical t-values from a t-table (Table 1).

The standard error of each coefficient is computed by regression software. The following equation applies for the case of one independent variable.

$$SE_b = \sqrt{\frac{\sum (Y_i - \hat{Y})^2 / (n-2)}{\sum (X_i - \bar{X})^2}} \quad (14)$$

For cases with more than one independent variable, the equation provides reasonable approximation when the independent variables are truly independent (i.e., not correlated). Otherwise, the equation gets very complex and the M&V analyst is better off using a software package to compute the standard errors of the coefficients. The range within which the true value of the coefficient, b falls is found using equation (8) as:

$$b \pm t \times SE_b$$

The standard error of the coefficient, b, also leads to the calculation of the t-statistic. This test ultimately determines if the computed coefficient is statistically significant. The t-statistic is computed by all statistical software using the following equation:

$$\text{t-statistic} = \frac{b}{SE_b} \qquad (15)$$

Once the t-statistic is estimated, it can be compared against critical t values from Table 1. If the absolute value of the t-statistic exceeds the appropriate number from Table B-1, then it should be concluded that the estimate is statistically valid.

A rule of thumb states that the absolute value of a t-statistic result of 2 or more implies that the estimated coefficient is significant relative to its standard error, and therefore that a relationship does exist between Y and the particular X related to the coefficient. It can then be concluded that the estimated b is not zero. However, at a t-statistic of about 2, the precision in the value of the coefficient is about ±100%: not much of a vote of confidence in the value of b. To obtain a better precision of say ±10%, the t-statistic values must be around 20, or the standard error of b has to be no more than 0.1 of b itself.

To improve the t-statistic result s considered the actions below:

- select independent variable(s) with the strongest relationship to energy;
- select independent variable(s) whose values span the widest possible range (if X does not vary at all in the regression model, b cannot be estimated and the t-statistic will be poor);
- gather and use more data points to develop the model; or
- select a different functional form for the model; for example, one which separately determines coefficient(s) for each season in a building that is significantly affected by seasonal weather changes.

3 Sampling

Sampling creates errors because not all units under study are measured. The simplest sampling situation is that of randomly selecting n units from a total population of N units. In a random sample, each unit has the same probability $\left(n/N\right)$ of being included in the sample.

In general, the standard error is inversely proportional to $\sqrt{n}$. That is, increasing the sample size by a factor "f" will reduce the standard error (improve the precision of the estimate) by a factor of $\sqrt{f}$.

3.1 Sample Size Determination

You can minimize sampling error by increasing the fraction of the population that is sampled $\left(n/N\right)$.

Increasing the sample size typically increases cost. Several issues are critical in optimizing sample sizes. The following steps should be followed in setting the sample size.

1. **Select a homogeneous population.** In order for sampling to be cost effective, the measured units should be expected to be the same as the entire population. If there are two different types of units in the population, they should be grouped and sampled separately. For example, when designing a sampling program to measure the operating periods of room lighting controlled by occupancy sensors, rooms occupied more or less continuously (e.g., multiple person offices) should be separately sampled from those which are only occasionally occupied (e.g., meeting rooms).
2. **Determine the desired precision and confidence levels** for the estimate (e.g., hours of use) to be reported. Precision refers to the error bound around the true estimate (i.e., ±x% range around the estimate). Higher precision requires larger sample. Confidence refers to the probability that the estimate will fall in the range of precision (i.e., the probability that the estimate will indeed fall in the ±x% range defined by the precision statement). Higher probability also requires larger samples. For example, if you want 90% confidence and ±10% precision, you mean that the range defined for the estimate (±10%) will contain the true value for the whole group (which is not observed) with a probability of 90%. As an example, in estimating the lighting hours at a facility, it was decided to use sampling because it was too expensive to measure the operating hours of all lighting circuits. Metering a sample of circuits provided an estimate of the true operating hours. To meet a 90/10 uncertainty criterion (confidence and precision) the sample size is determined such that, once the operating hours are estimated by sampling, the range of sample estimate (±10%) has to have a 90% chance of capturing the true hours of use. The conventional approach is to design sampling to achieve a 90% confidence level and ±10% precision. However, the M&V Plan needs to consider the limits created by the budget. Improving precision from say ±20% to ±10% will increase sample size by 4 times, while improving it to ±2% will increase sample size by 100 times (This is a result of the sample error being inversely proportional to $\sqrt{n}$.). Selecting the appropriate sampling criteria requires balancing accuracy requirements with M&V costs.
3. **Decide on the level of disaggregation**. Establish whether the confidence and precision level criteria should be applied to the measurement of all components, or to various sub-groups of components.
4. **Calculate Initial Sample Size**. Based on the information above, an initial estimate of the overall sample size can be determined using the following equation:

$$n_0 = \frac{z^2 * cv^2}{e^2} \qquad (16)$$

where:

- n_o is the initial estimate of the required sample size, before sampling begins
- cv is the coefficient of variance, defined as the standard deviation of the readings divided by the mean. Until the actual mean and standard deviation of the population can be estimated from actual samples, 0.5 may be used as an initial estimate for cv.
- e is the desired level of precision.
- z is the standard normal distribution value from Table 1, with an infinite number of readings, and for the desired confidence level. For example z is 1.96 for a 95% confidence level (1.64 for 90%, 1.28 for 80%, and 0.67 for 50% confidence).

NOTE: When n< 30 use a t statistic, when n> 30 use a z statistic and when the sample size is infinite then the t statistic and z statistic are equal.

For example, for 90% confidence with 10% precision, and a cv of 0.5, the initial estimate of required sample size (n_o) is

$$n_o = \frac{1.64^2 \times 0.5^2}{0.1^2} = 67$$

In some cases (e.g., metering of lighting hours or use), it may be desirable to initially conduct a small sample for the sole purpose of estimating a cv value to assist in planning the sampling program. Also values from previous M&V work may be used as appropriate initial estimates of cv.

5. **Adjust initial sample size estimate for small populations**. The necessary sample size can be reduced if the entire population being sampled is no more than 20 times the size of the sample. For the initial sample size example, above, (no = 67), if the population (N) from which it is being sampled is only 200, the population is only 3 times the size of the sample. Therefore the "Finite Population Adjustment" can be applied. This adjustment reduces the sample size (n) as follows:

$$n = \frac{n_0 N}{n_0 + N} \qquad (17)$$

Applying this finite population adjustment to the above example reduces the sample size (n) required to meet the 90%/±10% criterion to 50.

6. **Finalize Sample Size**. Because the initial sample size (n_o) is determined using an assumed cv, it is critical to remember that the actual cv of the population being sampled may be different. Therefore a different actual sample size may be needed to meet the precision criterion. If the actual cv turns out to be less than the initial assumption in step 4, the required sample size will be unnecessarily large to meet the precision goals. If the actual cv turns out to be larger than assumed, then the precision goal will not be met unless the sample size increases beyond the value computed by Equations (16) and (17).

As sampling continues, the mean and standard deviation of the readings should be computed. The actual cv and required sample size (Equations 16 and 17) should be re-computed. This re-computation may allow early curtailment of the sampling process. It may also lead to a requirement to conduct more sampling than originally planned. To maintain M&V costs within budget it may be appropriate to establish a maximum sample size. If this maximum is actually reached after the above re-computations, the savings report(s) should note the actual precision achieved by the sampling.

4 Metering

Energy quantities and independent variables are often measured as part of an M&V program, using meters. No meter is 100% accurate, though more sophisticated meters may increase the accuracy towards 100%. The accuracy of selected meters is published by the meter manufacturer, from laboratory tests. Proper meter sizing, for the range of possible quantities to be measured, ensures that collected data fall within known and acceptable error limits (or precision).

Manufacturers typically rate precision as either a fraction of the current reading or as a fraction of the maximum reading on the meter's scale. In this latter case it is important to consider where the typical readings fall on the meter's scale before computing the precision of typical readings. Over-sizing of meters whose precision is stated relative to maximum reading will significantly reduce the precision of the actual metering.

The readings of many meter systems will 'drift' over time due to mechanical wear. Periodic re-calibration against a known standard is required to adjust for this drift. It is important to maintain the precision of meters in the field through routine maintenance, and calibration against known standards.

In addition to accuracy of the meter element itself, other possibly unknown effects can reduce meter system precision:

- poor placement of the meter so it does not get a representative 'view' of the quantity it is supposed to measure (e.g., a fluid flow meter's readings are affected by proximity to an elbow in the pipe)
- data telemetry errors which randomly or systematically clip off meter data

  As a result of such unquantifiable metering errors, it is important to realize that manufacturer-quoted precision probably overstates the precision of the actual readings in the field. However there is no way to quantify these other effects.

  Manufacturer precision statements should be in accordance with the relevant industry standard for their product. Care should be taken to determine the confidence level used in quoting a meter's precision. Unless stated otherwise, the confidence is likely to be 95%.

  When a single measurement is used in a savings computation, rather than the mean of several measurements, then independent components are combined to determine uncertainties. The standard error of the measured value is:

$$SE = \frac{meter\ relative\ precision \times measured\ value}{t} \qquad (18)$$

  Where t is based on the large sampling done by the meter manufacturer when developing its relative precision statement. Therefore the Table 1 value of t should be for infinite sample sizes.

  When making multiple readings with a meter, the observed values contain both meter error and variations in the phenomenon being measured. The mean of the readings likewise contains both effects. The standard error of the estimated mean value of the measurements is found using equation (4).

5 Combining Components of Uncertainty

Both the measurement and adjustment components in the equation:

Savings = (Baseline – Period Use or Demand – Reporting-Period Use or Demand) ± Adjustments

can introduce uncertainty in reporting savings. The uncertainties in the individual components can be combined to enable overall statements of savings' uncertainty. This combination can be performed by expressing the uncertainty of each component in terms of its standard error.

The components must be independent to use the following methods for combining uncertainties. Independence means that whatever random errors affect one of the components are unrelated to the errors affecting other components.

If the reported savings is the sum or difference of several independently determined components (C) (i.e., $Savings = C_1 \pm C_2 \pm ... \pm C_p$), then the standard error of the reported savings can be estimated by:

$$\text{SE}(\textit{Savings}) = \sqrt{SE(C_1)^2 + SE(C_2)^2 + ... + SE(C_p)^2} \qquad (19)$$

For example, if savings are computed using the equation:

Savings = (Baseline Energy – Reporting-Period Energy) ± Routine Adjustments ± Non-Routine Adjustments

as the difference between the adjusted-baseline energy and measured reporting-period energy, the standard error of the difference (savings) is computed as:

$$SE(\textit{Savings}) = \sqrt{SE(adjusted\ baseline)^2 + SE(reporting\ \ period\ energy)^2}$$

The SE (adjusted baseline) comes from the standard error of the estimate derived from Equation (11). The SE (reporting period energy) comes from the meter accuracy using Equation (18). If the reported savings estimate is a product of several independently determined components (C_i) (i.e. $Savings = C_1 * C_2 * ... * C_p$), then the relative standard error of the savings is given approximately by:

$$\frac{SE(Savings)}{Savings} \approx \sqrt{\left(\frac{SE(C_1)}{C_1}\right)^2 + \left(\frac{SE(C_2)}{C_2}\right)^2 + ... + \left(\frac{SE(C_p)}{C_p}\right)^2} \qquad (20)$$

A good example of this situation is the determination of lighting *savings* as:

Savings = Δ Watts x Hours

If the M&V Plan requires measurement of hours of use, then "Hours" will be a value with a standard error. If the M&V Plan also includes measurement of the change in wattage, then ΔWatts will also be a value with a standard error. Relative standard error of savings will be computed using the formula above as follows:

$$\frac{SE(Savings)}{Savings} = \sqrt{\left(\frac{SE(\Delta Watts)}{\Delta Watts}\right)^2 + \left(\frac{SE(Hours)}{Hours}\right)^2}$$

When a number of savings results are totaled and they all have the same Standard Error, equation (5) or (19) can be used to find the estimated Standard Error of the total reported.

Total SE(*Savings*) = $\sqrt{SE(savings_1)^2 + SE(savings_2)^2 + ... + SE(savings_N)^2}$ =

$$\sqrt{N} \times SE(Savings)$$

Where N is the number of savings results with the same Standard Error that are added together.
Once the standard error of the savings is determined from the above calculation, it is possible to make appropriate concluding statements about the relative amount of uncertainty inherent in the savings, using the mathematics of the standard normal distribution curve, Figure 1, or data in Table 1. For example, one can compute three values:

1. Absolute or relative precision of the total saving, for a given level of confidence (e.g., 90%), computed using the relevant t value from Table 1 and Equation (7) or (9), respectively.
2. Probable Error (PE), defined as the 50% confidence range. Probable Error represents the most likely amount of error. That is, it is equally likely that error will be larger or smaller than the PE. (ASHRAE, 1997). Table 1 shows that 50% confidence level is achieved at t = 0.67 for samples sizes larger than 120, or 0.67 standard errors from the mean value. So the range of probable error in reported savings using Equation (8) is ±0.67 x SE (Savings).
3. The 90% Confidence Limit (CL), defined as the range where we are 90% certain that random effects did not produce the observed difference. From Table 1 using Equation (8), CL is ±1.64 x SE (Savings) for sample sizes larger than 120.

5.1 Assessing Interactions of Multiple Components of Uncertainty

Equations (19) and (20) for combining uncertainty components can be used to estimate how errors in one component will affect the accuracy of the overall savings report. M&V resources can then be designed to cost-effectively reduce error in reported savings. Such design considerations would take into account the costs and the effects on savings precision of possible improvements in the precision of each component.

Software applications written for common spreadsheet tools allow for easy assessment of the net error associated with the combination of multiple components of uncertainty, using Monte Carlo techniques. Monte Carlo analysis allows the assessment of multiple “what if” scenarios revealing a range of possible outcomes, their probability of occurring, and which component has the most effect on the final output. Such analysis identifies where resources need to be allocated to control error. A simple illustration of “what if” analysis is presented below for a lighting retrofit. A nominally 96 watt light fixture is replaced with a nominal 64-watt fixture. If the fixture operates for 10 hours every day, the annual savings would be computed as:

$$Annual\ Savings = \frac{(96-64)\times 10\times 365}{1{,}000} = 117kWh$$

The new 64-watt fixture’s wattage is consistent and easily measured with accuracy. However there is much variation among the old-fixture wattages and among the hours of use in different locations. Old-fixture wattages and hours of use are not easily measured with certainty. Therefore the savings will also not be known with certainty. The M&V design challenge is to determine the impact on reported savings if the measurement of either of these uncertain quantities is in error by plausible amounts.

Figure 2 shows a sensitivity analysis of the savings for the two parameters, old-fixture watts, and hours of use. Each is varied by up to 30% and the impact on savings is shown. It can be seen that savings are significantly more sensitive to variation in old-fixture wattage than to hours of use. A 30% wattage error produces a 90% savings error, while a 30% error in operating hours produces only a 30% savings error.

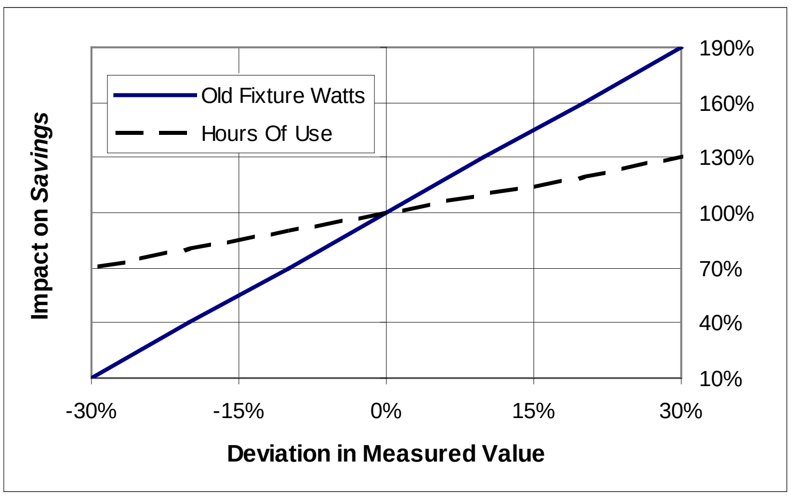

Figure 2 Example Sensitivity Analysis - Lighting Savings

If the proposed M&V method will yield readings of old-fixture wattage with a range of uncertainty of ±5%, the range of electricity savings uncertainty will be ±15%. In other words, if the old-fixture wattage could be between 91 and 101 watts, the savings could be between 99 and 135 kWh annually. The range of uncertainty on the savings is 36 kWh (135 - 99). If the marginal value of electricity is 10 cents per kWh, the uncertainty range is about $3.60 annually. If the wattage of the old fixture could be estimated with greater precision for significantly less than $3.60, then it may be worth enhanced-measurement efforts, depending on the number of years of savings being considered.

Figure 2 shows that the hours-of-use term has less of an impact on final savings in this example (the hours-of-use line is flatter indicating lower sensitivity). It is plausible that the error in measurement of operating hours is ±20%, so the energy-savings uncertainty range is also ±20% or ±23 kWh (= 20% of 117 kWh). The range in savings is about 46 kWh (= 2 x 23 kWh), worth $4.60 per year. Again it may be warranted to increase the accuracy in measuring the hours of use if it can be done for significantly less than $4.60, depending upon the number of years of savings being considered.

The range of possible savings errors from errors in measuring operating hours (46 kWh) is greater than from the error in measuring the old-fixture wattages (36 kWh). This is the opposite effect from what might be expected based on the greater sensitivity of savings to wattage than to hours of use, as seen in Figure 2. This difference arises because the plausible error of measuring operating hours (±20%) is much larger than the plausible error of measuring old-fixture wattages (±5%).

Sensitivity analysis such as the above can take many forms. The preceding simple example was used to show the principles. Monte Carlo simulation, allows complex consideration of many different parameters, allowing M&V design to focus expenditures where most needed to improve the overall accuracy of savings reports.

5.2 Establishing Targets for Quantifiable Savings Uncertainty

As discussed previously, not all uncertainties can be quantified. However, those that can be quantified provide guidance in M&V Planning. By considering the M&V cost of various optional approaches to uncertainty, the M&V program can produce the type of information that is acceptable to all readers of the savings report, including those who have to pay for the M&V reports. Ultimately, any M&V Plan should report the expected level of quantifiable uncertainty.

Determination of energy savings requires estimating a difference in energy levels, rather than simply measuring the level of energy itself. In general, calculating a difference to suit a target relative precision criterion requires better absolute precision in the component measurements than the absolute precision required of the difference. For example, suppose the average load is around 500 kW, and the anticipated savings are around 100 kW. A ±10% error with 90% confidence ("90/10") criterion can be applied two ways:

- If applied to the load measurements, absolute precision must be 50 kW (10% of 500 kW) at 90% confidence.
- If applied to the reported savings, absolute precision in the savings must be 10 kW (10% of 100 kW) at the same 90% confidence level. To achieve this 10 kW absolute precision in reported savings requires component measurement absolute precisions of 7 kW (using Equation (19), if both components are to have the same precision).

Clearly the application of the 90/10 confidence/precision criterion at the level of the savings requires much more precision in the load measurement than a 90/10 requirement at the level of the load.

The precision criterion may be applied not only to energy savings, but also to parameters that determine savings. For example, suppose the savings amount is the product of the number (N) of units, hours (H) of operation, and change (C) in watts: Savings = N x H x C. The 90/10 criterion could be applied separately to each of these parameters. However, achieving 90/10 precision for each of these parameters separately does not imply that 90/10 is achieved for the savings, which is the parameter of ultimate interest. In fact using Equation (20) the precision at 90% confidence would only be ±17%. On the other hand, if the number of units and change in watts are assumed to be known without error, 90/10 precision for hours implies 90/10 precision for savings.

The precision standard could be imposed at various levels. The choice of level of disaggregation dramatically affects the M&V design and associated costs. In general, data collection requirements increase if precision requirements are imposed on each component. If the primary goal is to control savings precision for a project as a whole, it is not necessary to impose the same precision requirement on each component.

6 Example Uncertainty Analysis

To illustrate the use of the various statistical tools for uncertainty analysis, Table 3 shows an example spreadsheet regression model output. It is a regression of a building's 12 monthly electric-utility consumption-meter values and cooling degree days (CDD) over a one-year period. This is just a partial spreadsheet output. Specific values of interest are highlighted in italics.

Table 3 Example Regression Analysis Spreadsheet Output

SUMMARY OUTPUT

Regression Statistics					
Multiple R	0.97				
R Square	0.93				
Adjusted R Square	0.92				
Standard Error	367.50				
Observations	12.00				
	Coefficients	Standard Error	T Stat	Lower 95%	Upper 95%
Intercept	5,634.15	151.96	37.08	5,295.56	5,972.74
CDD	7.94	0.68	11.64	6.42	9.45

For A Baseline of 12 Monthly kWh and CDD Data Points the derived regression model is:

Monthly electricity consumption = 5,634.15 + (7.94 x CDD)

The coefficient of determination, R^2, (shown as "R Square" in Table 3) is quite high at 0.93, indicating that 93% of the variation in the 12 energy data points is explained by the model using CDD data. This fact implies a very strong relationship and that the model may be used to estimate adjustment terms in the relevant form of Savings = (Baseline – Period Use or Demand – Reporting-Period Use or Demand) ± Adjustments. The estimated coefficient of 7.94 kWh per CDD has a standard error of 0.68. This SE leads to a t-statistic (shown as "T stat" in Table 3) of 11.64. This t-statistic is then compared to the appropriate critical t value in Table 1 (t = 2.2 for 12 data points and 95% confidence). Because 11.64 exceeds 2.2, CDD is a highly significant independent variable. The spreadsheet also shows that the range for the coefficient at the 95% level of confidence is 6.42 to 9.45, and implies a relative precision of ±19% (= (7.94 - 6.42) / 7.94). In other words, we are 95% confident that each additional CDD increases kWh consumption between 6.42 and 9.45 kWh.

The standard error of the estimate using the regression formula is 367.5. The average CDDs per month is 162 (not shown in output). To predict what electric consumption would have been under average cooling conditions, for example, this CDD value is inserted into the regression model:

Predicted consumption = 5,634 + (7.94 x 162) = 6,920 kWh per average cooling degree day month

Using a Table 1 *t*-value of 2.2, for 12 data points and a 95% *confidence* level, the range of possible predictions is:

Range of predictions = 6,920 ± (2.2 x 367.5) = 6,112 to 7,729 kWh.

The absolute precision is approximately ±809 kWh (= 2.2 x 367.5) and the relative precision is ±12% (= 809 / 6,920). The spreadsheet described value for the standard error of the estimate provided the information needed to compute the relative precision expected from use of the regression model for any one month, in this case 12%. If reporting-period consumption was 4,300 kWh, savings computed as: Savings = (6,920 – 4,300)= 2,600 kWh. Since the utility meter was used to obtain the reporting-period electricity value, it's reported values may be treated as 100% accurate (SE = 0%) because the utility meter defines the amounts paid, regardless of meter error. The SE of the savings number will be:

$$SE(monthly\ savings) = \sqrt{SE(adjusted\ baseline)^2 + SE(reporting\ \ period\ consumption)^2}$$

$$\text{SE} = \sqrt{367.5^2 + 0^2} = 367.5$$

Using t of 2.2, the range of possible monthly savings is

Range of savings = 2,620 ± (2.2 x 367.5) = 2,620 ± -810 = 1,810 to 3,430

To determine the precision of the annual total of monthly savings, it is assumed that the standard error\ of each month's savings will be the same. The annual reported savings then have a standard error of:

SE (annual *savings*) = $\sqrt{12 \times 367.5^2}$ = 1,273 kWh

Since t derives from the model of the baseline, it remains at the 2.2 value used above. Therefore the absolute precision in annual savings is 2.2 x 1,273 = 2,801 kWh. Assuming equal monthly savings of 2,620 kWh, annual savings are 31,440 kWh, and the relative precision of the annual savings report is 9% (= (2,801 / 31,440) x 100).

EVO Would Like to Thank its Organizational Subscribers:

Primary:

BC Hydro
Services Industriels de Genève

Senior:

EDF Electricité de France
Pacific Gas& Electric
Southern California Edison
Schneider- Electric

Large Government:

Ontario Power Authority
Bonneville Power Administration

Small Government:

ADENE – Agência para a Energia(Portuguese Energy Agency)

Educational:

Université de Genève

Associate Annex 1:

Navigant Consulting Inc.

Quantum Energy Services &Technologies, Inc. (QuEST)

EEVS – Energy Efficiency Verification Specialists

HEP-ESCO d.o.o.

Non Profit Non Annex 1:

Taiwan Green Productivity Foundation (TGPF)