Apples to Apples: Leveraging Building Space-Type Breakdowns to Design a More Efficient Study

Clark Bernier, KEMA Inc., San Francisco, CA

ABSTRACT

The heterogeneous nature of non-residential building stocks means that stratified sampling by building type offers significant efficiencies in the precision of the population projections of studies of these buildings. This works well when buildings can easily be grouped into strata that have similar energy usage profiles to one another: offices, grocery stores, retail stores, etc. The similarity between these buildings means that as groups, they tend to have lower coefficients of variation than the population taken as a whole. These efficiencies are less pronounced, however, when variations in space usage *within* buildings results in buildings of the same nominal categories having significantly different energy characteristics; an "office" building that is all offices is a very different animal than an "office" building that is 20% laboratory space. In these cases, the coefficient of variation of building strata may be no better than a random sample of the population as a whole. Where information exists about the distribution of space types within the population, however, an analytical strategy that emphasizes space categories rather than building categories may offer significant improvement in precision for the same-sized sample of buildings; comparing apples to apples and oranges to oranges instead of relying on the similarities between varieties of "fruit punch."

The Project

We are working with a major state university to assess and quantify the potential energy savings in their current stock of 86 non-healthcare campus buildings over 50,000 ft^2. Building energy simulation models, developed in DOE 2 using the Survey-IT front end and calibrated to billing data, were used to access the potential energy savings of three levels of capital investment by the university. The university maintains a detailed database of its entire building stock as part of its state-mandated record keeping, including the total floor area in each building of eleven standardized categories of space type. A building-level analysis, projecting up to the population on the basis of total building energy usage or floor area, would include two sources of variation: the variation of space-type to space-type (occupancy, vintage, usage patterns) and the variation of the space type mixes from building to building. An ideal sample design might leverage the knowledge we have about the mix of space types in the population in order to remove this second source of sampling variation by treating each of these space categories as a stratum, and randomly select a sample from each. However, since projecting savings for a given part of a building often requires modeling the entire building, this would increase the total number of modeled buildings required to reach a target relative precision more than the increased precision would allowed us to reduce it.

Instead, we devised a sampling and analytical methodology and stratification variable that uses the university's space-type totals and two years of energy usage data to produce a building-focused sample which results in an optimally representative mix of space types. The approach allows us to minimize the number of buildings modeled, while maximizing the precision of the study by allowing us to project the sample to the campus using ratio estimates of savings by space type.[1] Sample selection is

[1] Preliminary estimates indicate that the same estimated precision will be achieved with a 40% smaller sample. The final assessment of the impact will be presented in the final draft of the paper, once we have data. The contractor conducting the building analyses projects data to be available in late April.

still done at the building level, but analysis and projection to the population are done at the space-type level. A revised method of calculating the error is used to account for the building-level sample design and inclusion probabilities.

Taxonomy of Space Types

The university keeps records on the total square footage of 11 different space types in each of their campus buildings. The space types are grouped in the building database using the State Commission on Higher Education Facilities (HEFC) classifications. We mapped these 11 spaces into six groupings as shown in
Table 1.

Table 1: Space Type Mappings

HEFC Space Label	Study Classification	Abbreviation
Classroom Facilities	General & Study	GS
General Use Facilities		
Study Facilities		
Laboratory Facilities	Laboratory Facilities	LF
Office Facilities	Office Facilities	OF
Residential Facilities	Residential Fac.	RF
Special Use Facilities	Special & Support	SS
Support Facilities		
Nonassignable Areas	Other	Other
Unclassified Assignable		
Health Care Facilities	EXCLUDED	EXCLUDED

The Laboratory Facilities were further broken down into "wet" research labs (WLF) and "dry" non-research labs (DLF; which include language labs, some computer labs, and teaching labs) based on classifications provided by university staff.

Table 2 shows the total number of buildings with each space type, the total floor area that mapped into each space type, and the average floor area per building of each space type.

Table 2: Number of Space Types and Floor Area (ft^2) in the university Building Population[2]

	GS	WLF	DLF	Other	OF	RF	SS
Number of Buildings	84	11	32	86	90	17	65
Mean Space Size (ft2)	17,963	43,413	15,846	12,942	16,240	38,842	9,859
Total Pop Area (ft2)	1,508,869	477,548	507,073	1,113,045	1,461,611	660,307	640,832

[2] Four buildings were removed from the analysis population by the university after the sample design was well under way; two because the university planned to demolish them, one because of access issues, and the fourth because the university did not have unrestricted jurisdiction over the building. They still informed the initial energy estimates, but were left out of the final sample draw. On discussion with the university, we determined that 1) no other buildings in the population frame fit the three criteria that excluded these buildings and that 2) the population frame could be reframed to focus on only the remaining buildings that would be viable candidates for energy measures. This allowed us to estimated the potential savings of the remaining population-frame buildings unbiased.

Including Energy Usage in the Stratification

When choosing the distribution of buildings (and consequently space types) to include in the sample, we want to take into account the estimated contribution of each building to the variability in the savings estimate. Generally speaking, a group of buildings with higher energy usage will have a wider range of savings estimates than a group with lower usage—even if both groups have roughly the same conditioned area. This reasoning extends to space types with higher usage as well; labs, which are much more energy intensive than storage space, will tend to have a much wider range (in absolute terms) of potential energy savings. Under the model-based statistical sampling (MBSS) methodology around which this sample design methodology is based, we sample higher variability sites at a higher rate than lower variability sites to reduce the overall variability of the sample. For this reason, we need a stratification variable that takes into account the variable energy usages across space types when grouping buildings into strata. A first step towards producing this was to estimate the amount of energy used by each building's component spaces. An ordinary least squares regression model of floor areas with energy usage as the dependent variable was used to make these estimates.

The Regression Model

The university provided us with two years (2006 and 2007) of aggregate energy usage information for each building in the study, resulting in two figures of total annual kBTU for each building. For buildings with two solid years of data, energy figures were averaged to produce a single estimate of annual energy usage for each building. A few buildings were opened in 2006 or had renovation work done that impacted one year of the usage information. For these buildings only the single complete year of data was used.

A number of buildings have significant amounts of vertical penetration, or open space above floor area that extends beyond a typical ceiling height. As this space must be conditioned along with the underlying floor area, it contributes to the amount of energy required to operate a given space. However, from the data provided for the population it is impossible to tell which specific areas in a building share in a building's vertical penetration. We worked under the assumption that within a specific building it was spread equally over all spaces. This was reflected in the model by using the gross energy consumption of the building and the net-of-vertical-penetration floor areas in our regression model.

The floor areas by space type ($GS_i, WLF_i, DLF_i, OF_i, RF_i, Other_i$, and SS_i respectively for each building *i*) and gross annual energy usages ($Energy_i$) were fed into an ordinary least squares estimate of the form:

$$Energy_i = \beta_1 \cdot GS_i + \beta_2 \cdot WLF_i + \beta_3 \cdot DLF_i + \beta_4 \cdot OF_i + \beta_5 \cdot RF_i + \beta_6 \cdot Other_i + \beta_7 \cdot SS_i$$

The intercept term was held to zero to reflect the assumption that all energy usage in each building could be attributed to one of the seven classes of space within that building. This assumption relies on the auditor/modeler to determine the proper allocation of common space and shared systems across the space types within a building. In many cases, proportionate floor area was used; in other cases, where the modeler thought that it had a large impact, occupation schedules were incorporated into the allocation decision.[3]

[3] This introduces a certain amount of measurement uncertainty into the results. For the buildings representing larger strata, this isn't a problem as the uncertainty ends up reflected in the reported sampling uncertainty. For the certainty strata, it could become an issue since the buildings contained therein are assumed to be measured with no error. This additional uncertainty must be taken into account when comparing the precision gains of the approach to a more traditional building-centric approach. As with all modeling-based studies, the possibility for bias, largely dealt with at the sampling level, has the potential to crop up in the modeling effort itself. It is important that the modeling team be aware of this potential and decide on non-arbitrary rules ahead of time to minimize its impact.

Table 3 shows the coefficients and associated statistics for this regression. The overall r-squared was 0.873, showing that roughly 87% of the variation in energy use between buildings was explained by the constituent floor areas. Every floor area has a coefficient that was statistically significantly different from zero at a 95% level of confidence. All space types except "Other" are statistically significant at a 99% level of confidence. The two laboratory floor areas dominate other areas in energy use per square foot: 1770 kBTU/yr/ft^2 for "wet" labs and 1110 kBTU/yr/ ft^2 for "dry" labs. Office space averages around 450 kBTU/yr/ft^2 while residential, general space, special/support, and other spaces average between 150 and 215 kBTU/yr/ft^2.

Table 3: Regression Results and Associated Statistics

	GS	RLF	TLF	OF	RF	Other	SS	R-squared
β	150.0	1773.5	1111.9	445.2	181.0	214.5	147.5	0.873
Std. Dev.	34.2	78.6	108.3	75.8	59.6	100.8	48.9	
T-stat	4.386	22.570	10.267	5.876	3.038	2.128	3.019	
p-value	0.000	0.000	0.000	0.000	0.003	0.035	0.003	

Table 4 appends the energy usage estimates to

Table 2 to create estimates of the average energy usage for each space type in the population. "Wet" Labs make up 31% of the energy usage, with offices and "dry" Labs contributing another 24% and 20% respectively. The remaining 25% of usage is spread across the other four space types.

Table 4: Estimated Energy Usage by Space Type across the Population

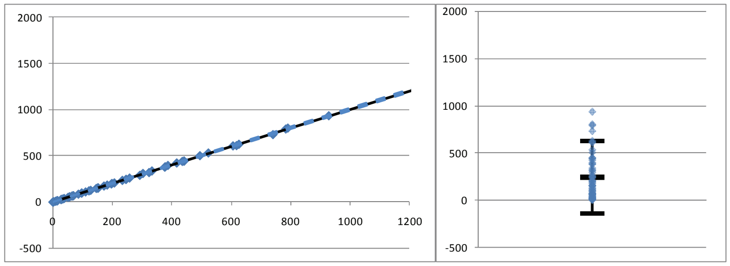
	GS	WLF	DLF	Other	OF	RF	SS
Number of Buildings	84	11	32	86	90	17	65
Mean Space Size (ft2)	17,963	43,413	15,846	12,942	16,240	38,842	9,859
Total Pop Area (ft2)	1,508,869	477,548	507,073	1,113,045	1,461,611	660,307	640,832
kBTU/yr/sf	150	1,774	1,112	214	445	181	148
Mean kBTU/yr	2,694,408	77,015,541	17,620,788	2,769,671	7,226,854	7,030,327	1,459,124
Total MBTU/yr	**226,330**	**847,171**	**563,865**	**238,192**	**650,417**	**119,516**	**94,843**

Why Ratio Analysis

Ratio analysis—the estimation of the ratio of some population total, Y, to another population total, X, is the central piece on which the MBSS methodology and this approach is built.[4] The motivation behind estimating a ratio instead of a population sum or population total directly is to leverage information we have about the population to make a more precise estimate given a sample of sites in that population. Figure 1, below, illustrates this at the extreme; if our goal is to estimate the population mean of Y, and we estimated the mean of Y directly, we would have the error bounds shown on the right. Instead, if we calculate the ratio of Y to X, the close and consistent relationship between them results in a very precise estimate of the ratio.

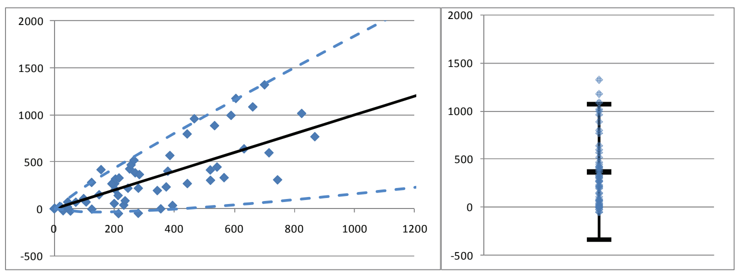
[4] The following is a high-level overview of the motivation and thinking behind the MBSS methodology, as outlined in Chapter 13 of The California Evaluation Framework (TecMarket Works, 2004). The uncertainty chapter itself builds upon the theoretical framework laid out in Cochran's *Sampling Technique* (Cochran 1977) and the techniques explored in Sarndal, Swensson and Wretman's *Model Assisted Survey Sampling* (Sarndal et al., 1992).

Figure 1: Ratio Estimate and Error Bound (left), Compared to Mean Estimate with Error Bounds (right). Beta=1, er=0.01, gamma =0.8



Even away from this extreme, ratio analysis still provides precision gains over mean-per-unit estimation as long as the Y whose mean is sought has a strong correlation with a variable X in the population. As Figure 2 shows, the error around the population points, estimated using a ratio, are narrower than those achieved using simple mean estimation.

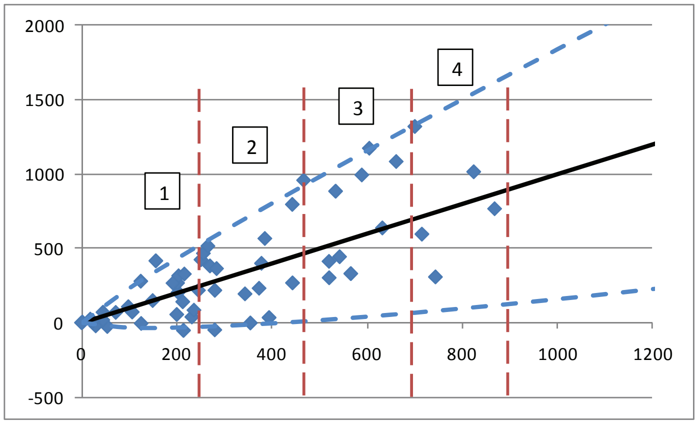
Figure 2: Ratio Estimate and Error Bound (left), Compared to Mean Estimate with Error Bounds (right). Beta=1, er=0.6, gamma =0.8



Model-Based Stratified Sampling (MBSS)

Ratio estimation in populations that, like buildings, have an inverse relationship between size and frequency in the population lends itself to additional improvement in precision through appropriate stratification techniques. When a stratification variable can be found in a population that is highly correlated with expected error of sites, and an equation (or model) specified to describe the relationship, then MBSS can be applied to achieve optimal sample distribution across the population, (TecMarket Works 2004, 332-338). In Figure 2 and Figure 3, the dotted lines depicting the expected standard error of the population, σ_i, are expressible as a function of X: $\sigma_i = \sigma_0 \cdot X_i^{\gamma}$, where gamma is less than 1 (typically 0.8). This is to say that the expected error of any site in the population increases with the X (often denoting, or a proxy for, the size) of the site. This makes sense as there are more factors that go into the estimates associated with a larger site than with a smaller site, and thus more ways that variability can enter the estimate.

Figure 3: Example of Model-based Stratification for Ratio Estimation



MBSS leverages a model of this sort to estimate the expected error from every site in the population before the sample is drawn and the sites are fielded. Strata are then chosen so that the *total expected error coming from the sites in each stratum are equal to one another*. Figure 3 shows an example of strata cut points chosen in this way. The sample sites are then spread evenly across the selected strata. The upshot is that strata with low X-values will have more members than strata with higher Xs, but the same number of sample points. Thus, there will be a higher rate of sampling among the larger sites and a lower rate among the smaller sites. This produces a sample design that better describes those sites that would introduce the most uncertainty into the analysis and produce a practically unbiased[5] estimate with relatively simple mathematics for planning and interpretation.

Creating the Stratification Variable

Returning to our energy usage model, where SF=square footage, we can rewrite the space types as $SF_1 = GS, SF_2 = WLF, ..., SF_7 = SS$ such that, for any building *i,* the total energy usage, E, can be written as $E_i = \sum_{j=1}^{7} \beta_j \cdot SF_{j,i}$ and the total floor area, FA, can be written as $FA_i = \sum_{j=1}^{7} SF_{j,i}$ where *j* represents the space type (1-7). Ultimately, we will calculate a savings ratio $\alpha_j = \dfrac{\sum_{i=1}^{n} Savings_{j,i}}{\sum_{i=n}^{n} SF_{j,i}}$ for each space type as the estimated savings per square foot based on the sample of selected buildings. This means that the estimated savings, *Y*, of any space, *j*, in a building, *i*, can be written: $Y_{j,i} = \alpha_j \cdot SF_{j,i} + \varepsilon_{j,i}$ where $\varepsilon_{j,i}$ is the error associated with each estimate and $E(\varepsilon_{j,i}) = 0$. Using the equations for the MBSS methodology outlined in Chapter 13 of the California Evaluation Framework (TecMarket Works, 2004), based on some γ_j (which we assume to be 0.8 for each space type), we need an estimate of the variance for any space type at a given site, $\sigma_{j,I}$ in order to build an optimal sample design.

Since we believe variation in savings will increase with the amount of energy used annually, we want to make sure to include the estimates of energy usage into the expected variance. If we assume

[5] Though technically biased by a negligible amount if the MBSS methodology is followed and the strata cut points chosen accordingly. See TecMarket Works 2004, 331 and Cochran 1977, 160-167.

that all $\gamma_j \equiv \gamma$, then we would expect that the variation of each space within each building would be proportionally related to the energy usage of that space raised to the power of γ. Therefore, $\sigma_{j,i} = k \cdot (\beta_j SF_{j,i})^{\gamma}$.

In order to estimate error, we need to aggregate back to the building level—not only is this the level at which the ultimate sampling decision must be made, but it is a level at which the errors are independent of one another. We would expect some unknown level of correlation between the errors of the space types within each building (they are affected by being from the same vintage, for instance), and thus cannot estimate their variance in the planning stage. Aggregating to the level of each building gives us an estimate of the total savings of any building in the population of $Y_i = \sum_{j=1}^{7} \alpha_j \cdot SF_{j,i} + \varepsilon_i$ where the building-level error is the sum of the space-level errors of that building, $\varepsilon_i = \sum_{j=1}^{7} \varepsilon_{j,i}$. Assuming independence of the errors between buildings, the variance of a sum of terms is equal to the square root of the sum of the squares of the summed terms' variances. Thus the variance associated with each building's estimate, σ_i, can be written as $\sigma_i = sd(\varepsilon_i) = k\sqrt{\sum_{j=1}^{7} (\beta_j SF_{j,i})^{2\gamma}}$. As outlined in the California Evaluation Framework, if we define $z_i = \sqrt{\sum_{j=1}^{7} (\beta_j SF_{j,i})^{2\gamma}}$, then stratifying by z_i such that the sum of the z_i in each strata are equal will be the most efficient way to stratify the population and designate the sample.

The Sample Design

Based on this stratification variable and the university's budgetary constraint to a sample size of 20 buildings, we used a four-strata MBSS sample design. The largest laboratory ended up in a certainty stratum while the rest of the sample was spread across the four strata as shown in Table 5.

Table 5: Strata bounds for a sample size of 20

Stratum	Pop Buildings	Sample Buildings	Min Z(i)	Max Z(i)	Mean Z(i)	Inclusion Probaility
1	43	5	5,859	13,769	9,731	0.12
2	22	5	15,411	25,227	18,302	0.23
3	12	5	29,006	45,785	39,627	0.42
4	8	4	51,159	99,452	74,602	0.50
5	1	1	146,679	146,679	146,679	1.00

Analysis Approach

Savings Ratios

When weighting the sample back up to the population, we need the weights to reflect the sample design under which the sample was drawn. Defining π_i as the inclusion probability of each building *i* as listed by stratum in Table 5 above, the weight of each building is the inverse of its inclusion probability $w_i = \pi_i^{-1}$. We then estimate the savings ratio of each space type *j* as the sum of the projected savings for that space type in each building i in the sample, divided by the total floor area of that space type in the sample. That is, $b_j = \frac{\sum_{i=1}^{n} \pi^{-1} savings_{i,j}}{\sum_{i=1}^{n} \pi^{-1} FloorArea_{i,j}}$, where $savings_{i,j}$ is the estimated savings of space type j in building i, $FloorArea_{i,j}$ is the floor area of space type j in building i, and b_j is the estimated potential savings per square foot of space type j.

This ratio estimate has an error associated with it for each building/space combination in the sample, which we call $e_{i,j}$. It is the difference between the modeled savings of the building/space and the expected savings given the calculated b_j; $e_{i,j} = savings_{i,j} - b_j \cdot FloorArea_{i,j}$.

Total Savings

Total savings in the population can then be calculated by multiplying the savings ratios by the total floor area of each space type. The total savings of each space type is the sum of the floor areas of that space type in all buildings in the population, multiplied by the estimated savings ratios, $FASavings_j = b_j \cdot \sum_{i=1}^{N} FloorArea_{i,j}$. Total savings is the sum of the savings of all of the space types, $TotalSavings = \sum_{j=1}^{7} FASavings_j$. This is equivalent to the total we would get if we totaled the savings by building first and then summed those savings to produce the estimate of total savings.

Overall Precision

Calculating overall relative precision requires that we aggregate the estimates and the error of the estimates up to the building level. This is the level at which we selected sites, and thus the level at which we have to apply our weights under this sample design. First, we calculate the estimated savings for each building in the sample using the space type ratios, $BldgSavings_i = \sum_{j=1}^{7} b_j \cdot FloorArea_{i,j}$. This is subtracted from the modeled savings of the buildings to give the building-level error term $e_i = \sum_{j=1}^{7} savings_{i,j} - \sum_{j=1}^{7} b_j \cdot FloorArea_{i,j}$. Disaggregating the two summations and pairing terms with the same j, this can be rearranged to $e_i = \sum_{j=1}^{7} (savings_{i,j} - b_j \cdot FloorArea_{i,j}) = \sum_{j=1}^{7} e_{i,j}$. That is to say, the error of each building is the sum of the errors of its constituent space types.

Relative precision can then be calculated using the equation reported in the California Evaluation Framework: $rp = z \cdot \frac{\sqrt{\sum_{i=1}^{n} \pi_i^{-1} \left(\pi_i^{-1} - 1\right) e_i^2}}{\sum_{i=1}^{n} (\pi_i^{-1} \sum_{j=1}^{7} savings_{i,j})}$ where z = 1.282 for an 80% level of confidence and 1.645 for a 90% level of confidence.

Building Modeling

Each building in the sample was modeled in DOE-2, and adjusted to self-reported occupancy schedules and, where applicable, thermostat and building control settings. The models were further calibrated against billing and usage data provided by the university. For the study, 26 Energy Conservation Measures (ECMs) were identified as applicable to the campus buildings. These measures were grouped into three categories: low-cost measures, less-than-three-years payback measures, and long-term payback measures.

For each building, the modeling audit was used to propose measures from each of these categories. The auditor/modelers were careful to choose those measures that they would recommend for the building even if it was not in this study. In this way, the potential measures chosen for modeling were reflective of the potential measures that would have been identified by a similarly-qualified auditing professional. Each building then had a simulation run with each of the measure sets included: low-cost measures, medium payback, and long-term payback measures. The projected energy usage of the building under each measure regime was recorded. The difference between each regime and the successive regime (from baseline to low-cost, or from low-cost to medium-term payback) was deemed to be the incremental savings of that measure group. Thus, for each building in the sample, numbers were produced that represented the baseline usage, the potential savings from low-cost measures, the potential savings of medium-term payback measures, and the potential savings of long-term payback measures.

Analysis Results

Sadly, at the time of publication, models and runs had been completed for only half of the 20 sample sites. For the 10 completed sites, taken as a partial sample, sampling error is less than expected, but we cannot draw any conclusions from it at this time. The full results will be presented at the conference session.

Generalizing the Apples-to-Apples Sample Approach

Step 1: Identify a disaggregation dimension that is known in the population

Step 2: Identify or create a stratification variable. Choose one that is expected to highly correlate with the expected estimate errors. If using ratio analysis, create one that is the square-rooted sum of the squares of the X-variables that will be used in estimating the disaggregated ratios.

Step 3: Stratify at the aggregate level and select sites.

Step 4: Analyze at the disaggregate level, calculating sample ratio estimates and errors for each disaggregated unit.

Step 5: Roll up the disaggregated units into the aggregated units, adding the errors together.

Step 6: Roll the aggregate unit estimates and errors up into the population estimates.

References

Cochran, William G (1977). *Sampling Techniques*. 3rd Edition. John Wiley & Sons: New York, NY.

Sarndal, Carl-Eric, Bengt Swensson and Jan Wretman (1992). *Model Assisted Survey Sampling*. Springer-Verlag: New York, NY.

TecMarket Works (2004). *The California Evaluation Framework*. California Public Utilities Commission: California.