Sample weight:

The variable FINALWT in the data file represents the final sampling weight, accounting for different probabilities of selection and rates of participation in the survey.

For example, if FINALWT for a building is 1,000, that building represents itself and 999 other non-sampled and non-responding

Sampling error: RSE (?)

Building characteristics (WWR) – building energy use

Methods: include regression [46–48], artificial neural networks (ANN) [49,50], random forest [51], clustering[52] and principal component analysis (PCA) [53–55].

The manual selection process of stepwise regression can provide insight to individual variables and the influence of different levels

1. Observation & variable selection

Numerical or categorical

Decompose into several aspects

1. Data formatting and filtering

Categorical

Numerical: binning

Missing value

1. Statistical methods

Illustrate sample weights

Feasible research question:

Predictor: incorporation of renewable energy sources

Control: 各种需要考虑的factor

Response: energy efficiency (energy consumption per square foot), different kinds of energy (electricity, gas, oil, heat)

需要验证independence?

<https://energycompliance.com.au/impact-of-ceiling-height-on-energy-demand/>

imputation:

BRAINSTORMING:

明天检查一下Derived variable: Pct heated by xx的数据分布

Heat: electricity, natural gas, fuel oil, propane, district, wood, solar, coal, other

Cool: electricity, natural gas, fuel oil, propane, district

想法：探索各种能源类型与consumption之间的关系？

探索energy-effective strategy 对consumption的影响

行动（light during off/on）

Smart thermostat