Validate ML/AI via Statistical Sampling

By Arnold Miller, Sr. SDET (SW QA-Tester)

Agile Testing & Test Automation Summit

Denver, CO, USA

19 Dec 2019

Machine Learning / Artificial Intelligence

Examples

- Self-Driving Vehicle; Hazard Warnings
- Disability Vision, Sound Recognition
- Security Physical, Cyber; Threat Detection
- Equipment and Health Diagnostics
- Election Voting Predictions
- Call Center, Customer Care Analysis

Validation via Sampling Methods

Sampling method refers to the way that observations are selected from a **population** to be in the **sample** for a **survey sample**

https://stattrek.com/survey-research/sampling-methods.aspx

The reason for conducting a sample survey is to **estimate** the **value** of some **attribute** of a **population**.

- **Population parameter**. A population parameter is the true value of a population attribute.
- **Sample statistic**. A sample statistic is an estimate, based on sample data, of a population parameter

Non-Probability Sample Methods

Do not know the probability that each population element will be chosen,

Cannot be certain that each population element has a non-zero chance of being chosen.

Two Main types are:

- **Voluntary sample**. Made up of people who self-select into the survey. Often, these folks have a strong interest in the main topic of the survey.
- **Convenience sample**. Made up of people who are easy to reach.

Probability Sample Methods

Each population element has a known (non-zero) chance of being chosen for the sample.

Main Sampling Methods

- **Simple Random**
- **Stratified**
- **Cluster**
- **Multistage**
- **Systematic Random**

Simple Random Sampling

Property

- The population consists of N objects.
- The sample consists of n objects.
- All possible samples of n objects are equally likely to occur

Example: Lottery method.

- Each of the N population members is assigned a unique number.
- The numbers are placed in a bowl and thoroughly mixed.
- A blind-folded researcher selects n numbers.
- Survey only members that have the selected numbers

Stratified Sampling

Property

- Population is divided into groups (strata), based on some characteristic.
- Then, within each group, a probability sample (like: Simple Random Sampling) is selected.
- In stratified sampling, the groups are called **strata**.
- The sample includes **elements** from **each stratum**.

Example: US National Elections

- Divide the population into groups or strata, based on geography like US States
- Within each stratum use simple randomly select survey respondents.

Cluster Sampling

Property

- Every member of the population is assigned to one, and only one, group.
- Each group is called a **cluster**.
- A sample of clusters is chosen, using a probability method (like: Simple Random Sampling).
- Survey only individuals within sampled clusters
- the sample includes **elements** only from **sampled clusters**

Example: Items in Group

- Each Group has 10 times.
- Select Groups via Simple random sampling
- Survey all or simple random items in Selected Group

Multistage Sampling

Property

- Select a sample by using combinations of different sampling methods.

Example

- Stage 1, Use cluster sampling to choose clusters from a population.
- Stage 2, Use simple random sampling to select a subset of elements from each chosen cluster for the final sample.

Systematic Random Sampling

Property

- Create a list of every member of the population.
- From the list, we randomly select the first sample element from the first k elements on the population list.
- Thereafter, we select every kth element on the list.
- Every possible sample of n elements is not equally likely (not a simple random sample)

Example

- Have 24,000 items in some order
- Random select first item from first 50 items
- Second select item is first random plus 50; Third select item is first random plus 100; etc.
- Survey has total 480 items (Which is 24,000 / 50)

Confidence Interval

Best for Simple Random Samples

Plus or Minus error value reported based on

Confidence level: Percentage certain for the interval. Normal 95% or 99%

Sample Size: numbers of items surveyed. The more the better but no linear

Population Size: Total number that could be surveyed

Percentage: Likely outcome via population size

https://www.surveysystem.com/sscalc.htm

Sample	Population	Percent	95% Cnf	99% Cnf
620	6,500	50%	3.74%	4.93%
620	65,000	50%	3.92%	5.16%
620	650,000	50%	3.93%	5.18%
620	6,500,000	50%	3.94%	5.18%

Example: Compare Probability Sample Methods

On-Demand TV Shows (about 6,500 total population)

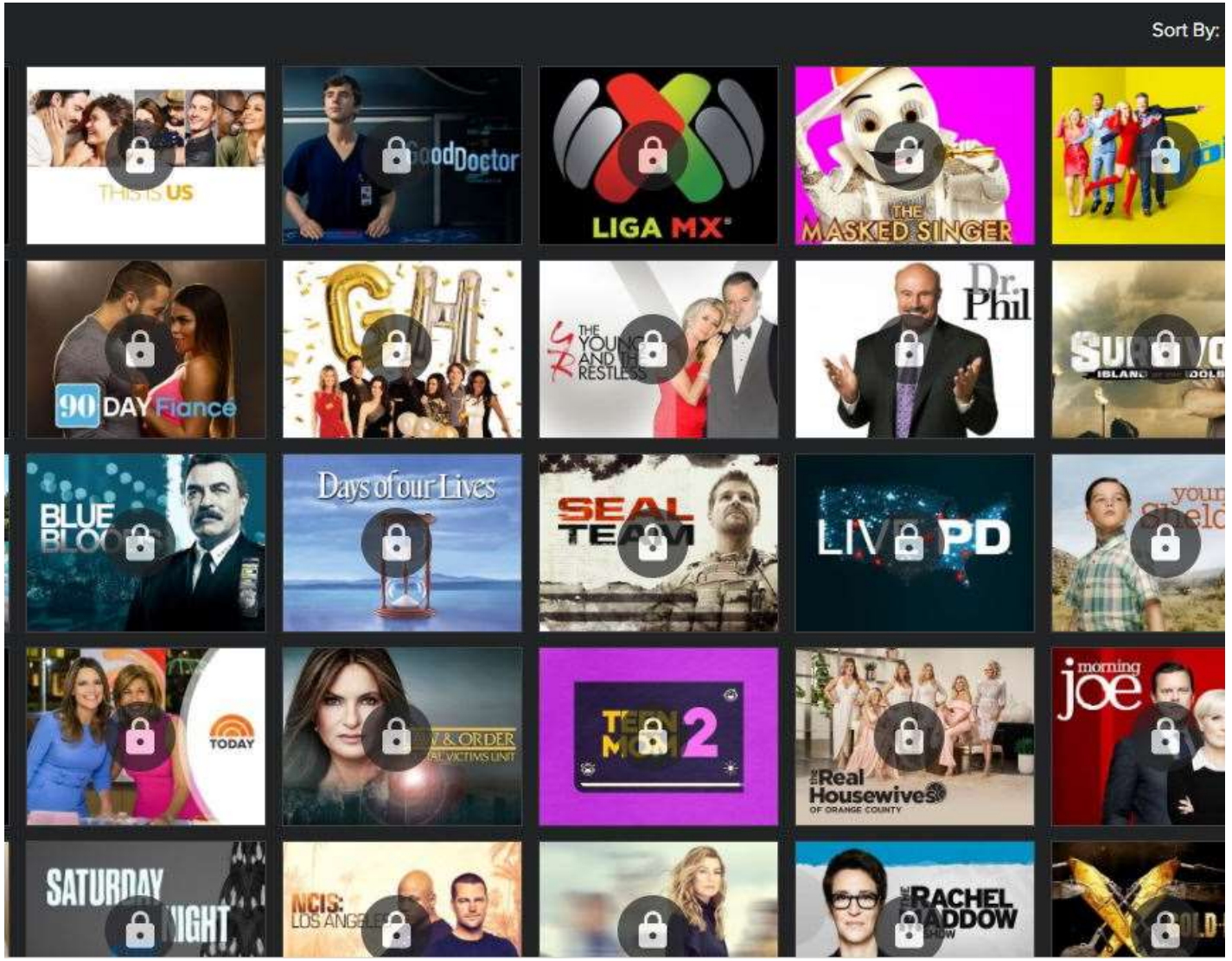
Web Site displays TV Shows in groups of 100 (Popular, Title, Critics, Added Date)

Evaluate TV Shows for attributes: Description, View Rating, Genre, and Network

Percentage has all these attributes

Percentage missing each one attributes

- No Description
- No View Rating
- No Genre
- No Network



Example: Models Evaluation Execution time

Total Population via 65 groups of 100, just like Web site

- Execution time about 7.2 minutes (50% more than Simple Random)

Simple Random 620 individual items

- Execution time about 4.8 minutes (Baseline execution time)

Stratified via Starting Letter Strata with pro-rated Simple Random individual items

- 620 individual items pro-rated via Strata population with minimum 1 per Strata
- Execution time about 5.3 minutes (10% more than Simple Random)

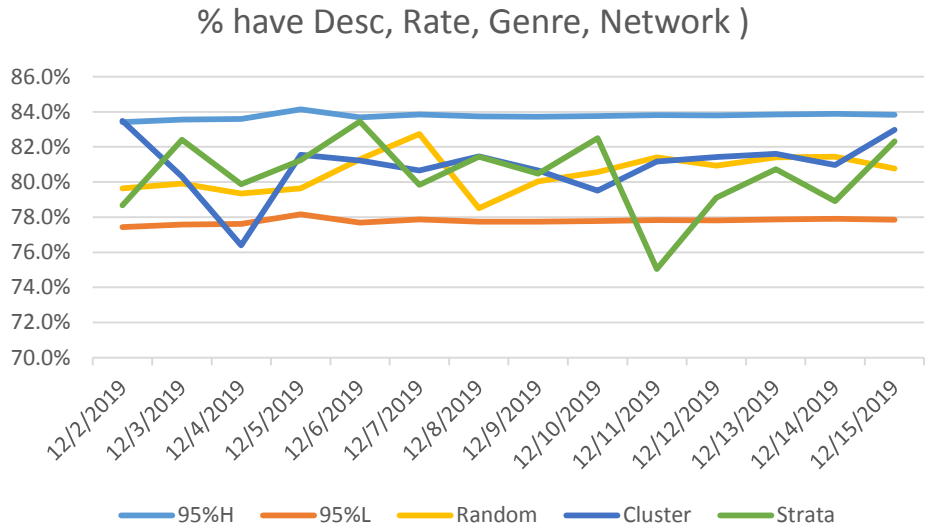
Cluster via 62 Simple Random groups of 10

- Evaluate all items in group
- Execution time about 1.2 minutes (about 75% less than Simple Random)

Example: Attribute – Have All

Historically: 80% have Description, View Rating, Genre and Network

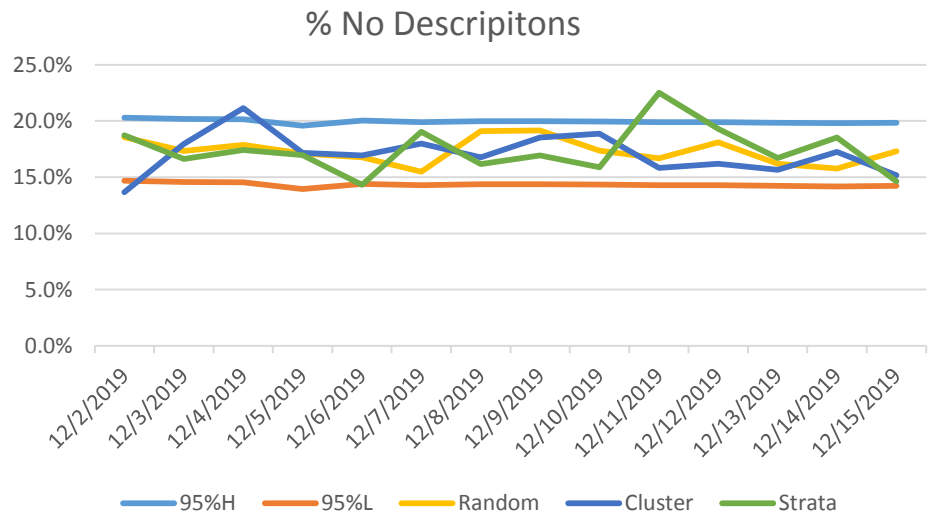
Sample	Population	Percent	95% Cnf	99% Cnf
620	6,500	80%	2.99%	3.94%



Example: Attribute – No Description

Historically: 17% are missing Description

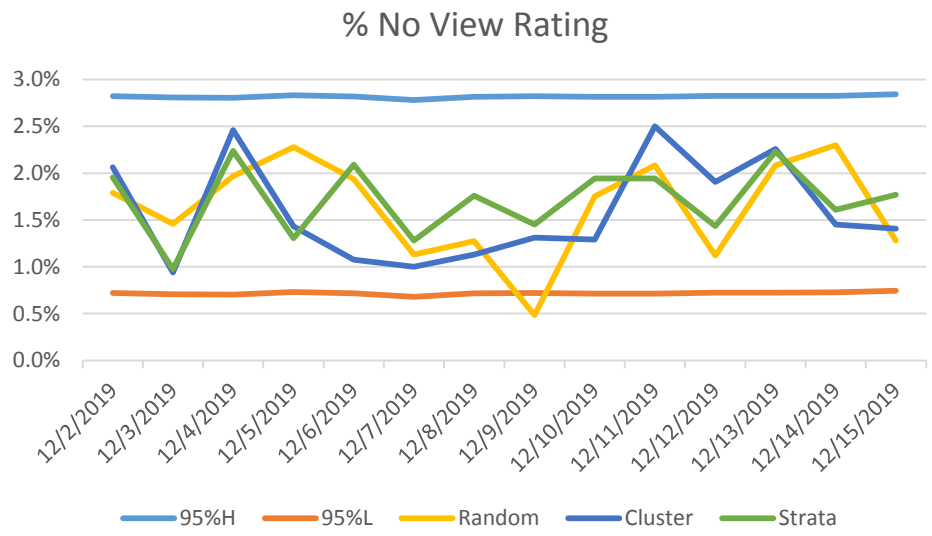
Sample	Population	Percent	95% Cnf	99% Cnf
620	6,500	17%	2.81%	3.70%



Example: Attribute – No View Rating

Historically: 2% are missing View Rating

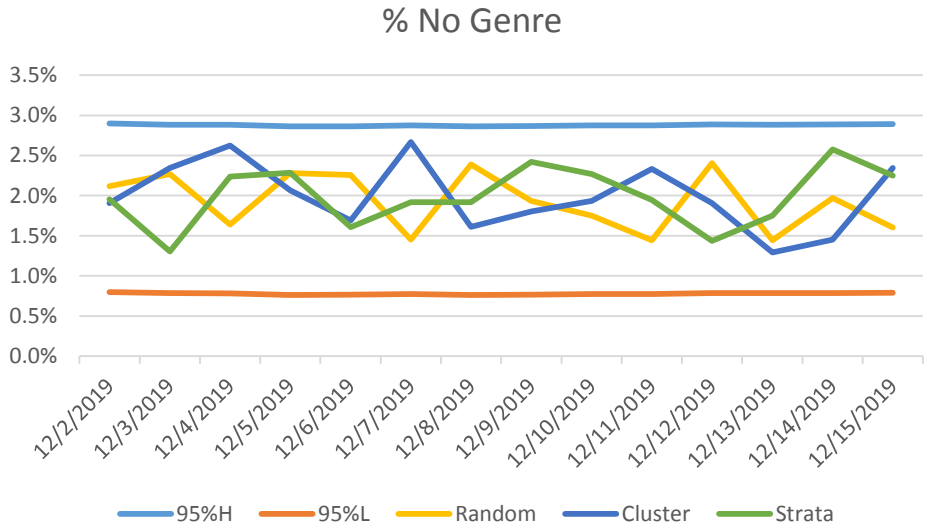
Sample	Population	Percent	95% Cnf	99% Cnf
620	6,500	2%	1.05%	1.38%



Example: Attribute – No Genre

Historically: 2% are missing Genre

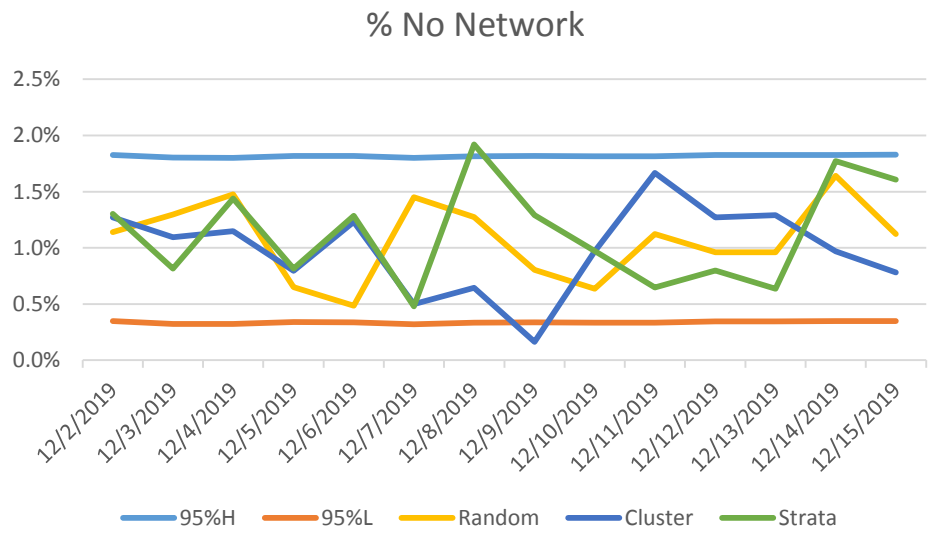
Sample	Population	Percent	95% Cnf	99% Cnf
620	6,500	2%	1.05%	1.38%

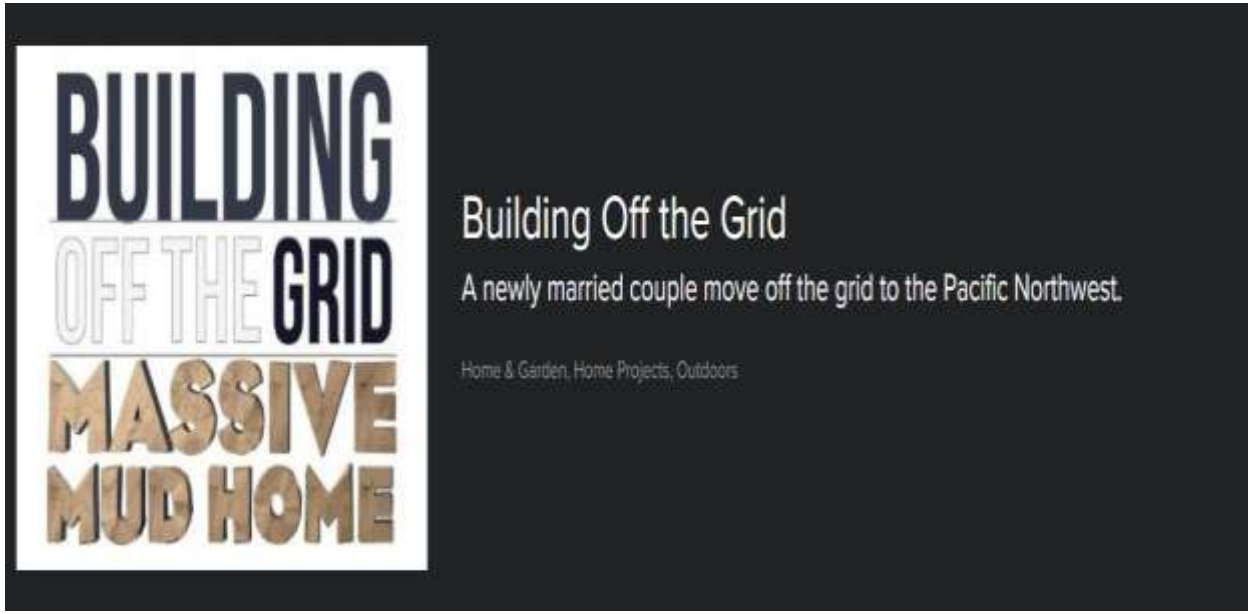


Example: Attribute – No Network

Historically: 2% are missing Network

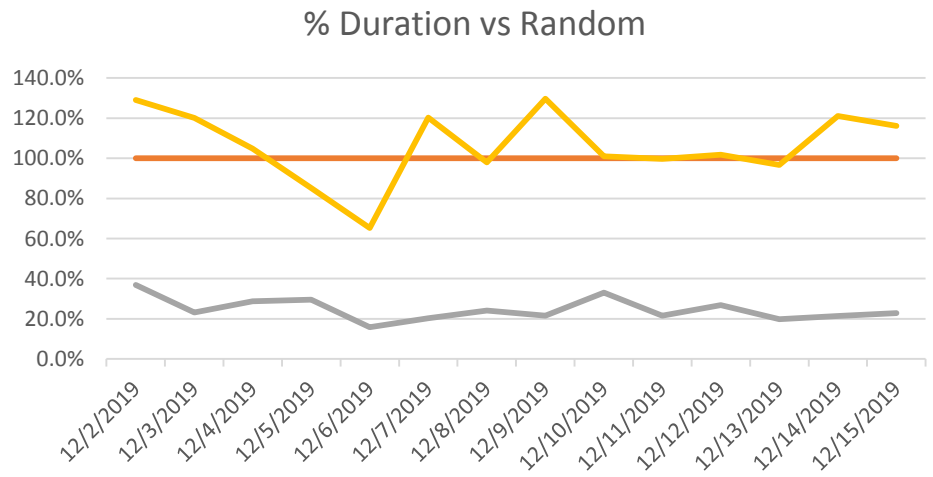
Sample	Population	Percent	95% Cnf	99% Cnf
620	6,500	1%	0.74%	0.98%





Example: Conclusion

- All these Methods seem to evaluate within 95% Confidence Interval
- **Simple Random** individual items
- **Stratified** via Starting Letter with pro-rate simple random
- **Cluster** via Simple Random groups of 10 (all evaluated)
- Choose daily **Cluster** as takes 75% less time than **Simple Random**
- Continue weekly **Simple Random** and **Stratified** methods



Your Follow-up and Thanks

Based on this information

- You should be able to expand the example
- To Validate Machine Learning Algorithms using
- Probability (Statistical) Sample Methods

Thanks

- Arnold Miller, Sr. SDET (Sw. QA-Tester)
- Email: arnold.miller0@gmail.com
- Linkedin: http://www.linkedin.com/pub/arnold-miller/13/47b/a87