Point Estimation

WHAT Y'ALL CAME HERE FOR

RDS-I

Description

Models sampling as a mixing of a Markov Chain.

- Only available for categorical variables.
- Comes in two versions.
  - Regular: Doesn’t use chain reversibility
  - Data Smoothed (DS): Assumes that the chain is reversible

RDS-II

Description

Weights cases based on the inverse of their network size.

- Assumes that **either:**
  - The Markov chain has reached convergence.
  - Network connections are completely random (Configuration graph), and sampling is done with replacement.

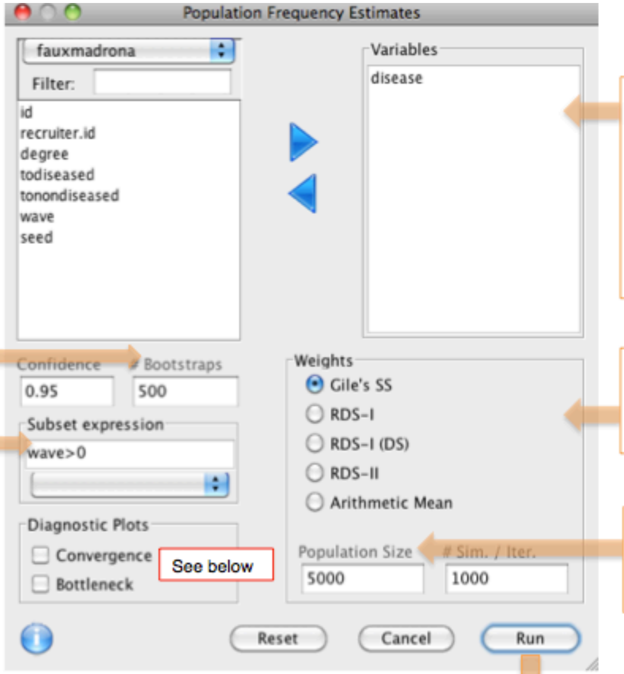
Gile's Sequential Sampler

Description

Adjusts RDS-II for sampling without replacement.

- Best when sample size is a significant fraction of population size.
- Requires the following information:
  - Population Size: An estimate of the population size
  - # Sim. \ Iter: An algorithmic parameter controlling the number of sampling simulations per iteration. Defaults to 1000.

The Dialog: Categorical Variables



fauxmadrona

Filter:

id
recruiter.id
degree
todiseased
tonondiseased
wave
seed

Variables: disease

Confidence: 0.95

Bootstraps: 500

Subset expression: wave>0

Diagnostic Plots

☐ Convergence

☐ Bottleneck

See below

Weights

● Gile's SS

○ RDS-I

○ RDS-I (DS)

○ RDS-II

○ Arithmetic Mean

Population Size: 5000

Sim. / Iter.: 1000

Often an indicator variable for a characteristic of interest—disease in this case. It may also be a categorical variable with multiple possible states.

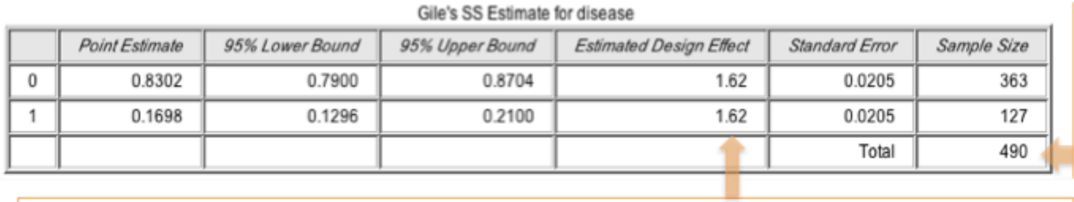
Entering multiple variables here will output multiple tables like the one below.

Number of bootstrap samples for estimating the uncertainty of the estimator.

Restrict the data used for calculating.
Here we remove the initial sampling wave.
Note: Diagnostic plots are not enabled if a subset expression Is used.

Method for calculating sampling weights. See *RDS Analyst Compute Weights*.

An estimate of the target population size is required if weight type is Gile's SS or Arithmetic Mean.



Gile's SS Estimate for disease

	Point Estimate	*95% Lower Bound*	*95% Upper Bound*	*Estimated Design Effect*	*Standard Error*	*Sample Size*
0	0.8302	0.7900	0.8704	1.62	0.0205	363
1	0.1698	0.1296	0.2100	1.62	0.0205	127
					Total	490

Sample size excludes 10 respondents from wave 0 due to the subset in the dialog.

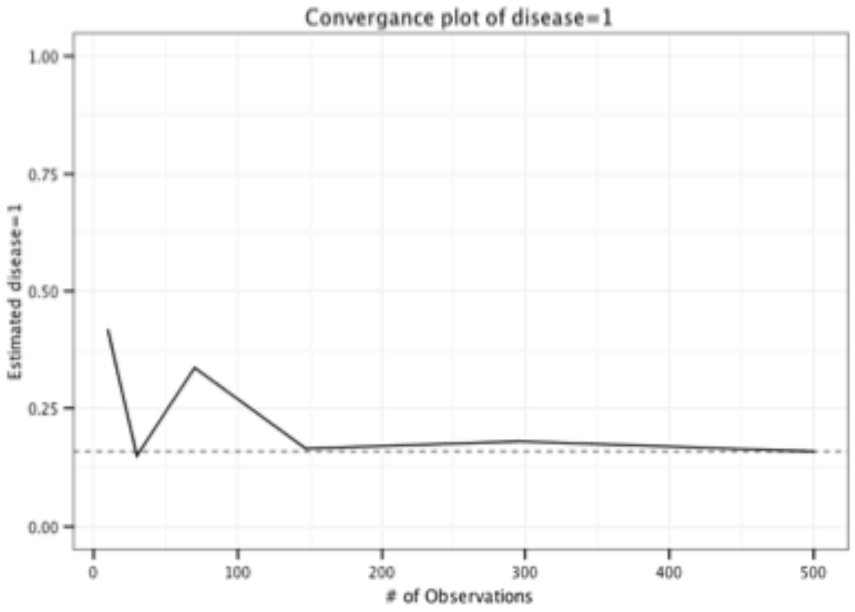
Design effect is the necessary sample size relative to a simple random sample. For example, if the value is two you would have to recruit twice as many subjects as with a simple random sample in order to get the same power.

Diagnostics: The convergence plot

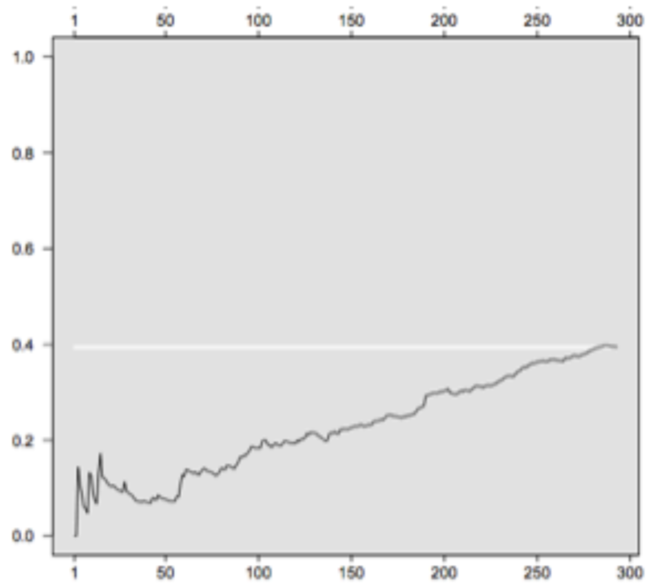
- Measures the systematic trending of the estimate as sampling progresses
- Can indicate that seed bias has not been overcome.

Diagnostics: The Convergence PLot

Good



Bad

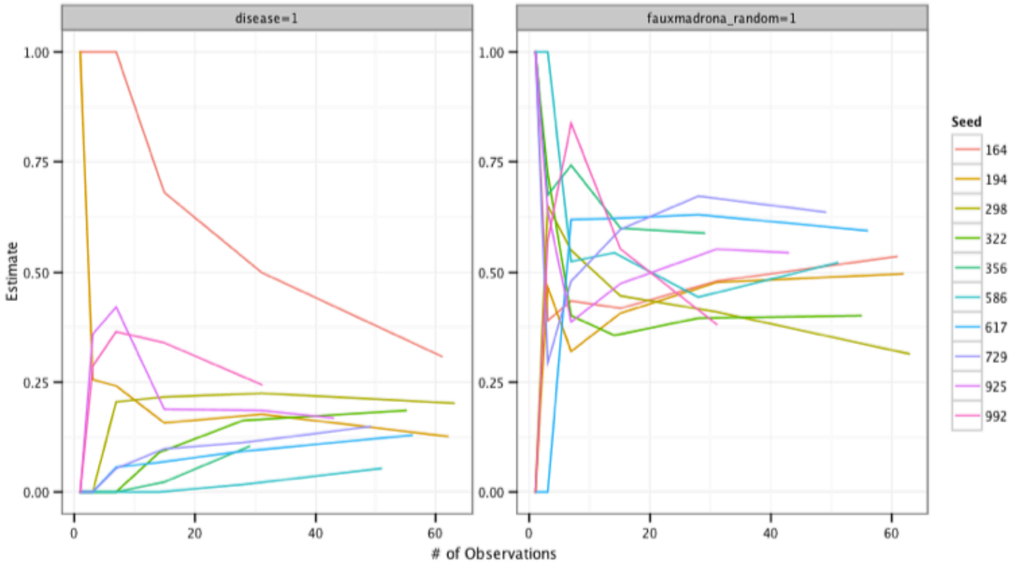


Diagnostics: Bottleneck plot

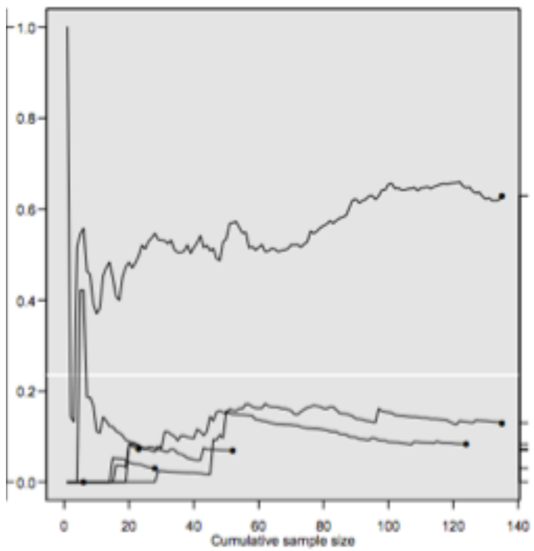
- Measures the trend in the estimate as the sampling process proceeds, broken down by initial seed.
- Suggests if some seeds located in qualitatively different populations.
- Can indicate that the population is actually composed of several disconnected sub-populations.

Diagnostics: Bottleneck Plot

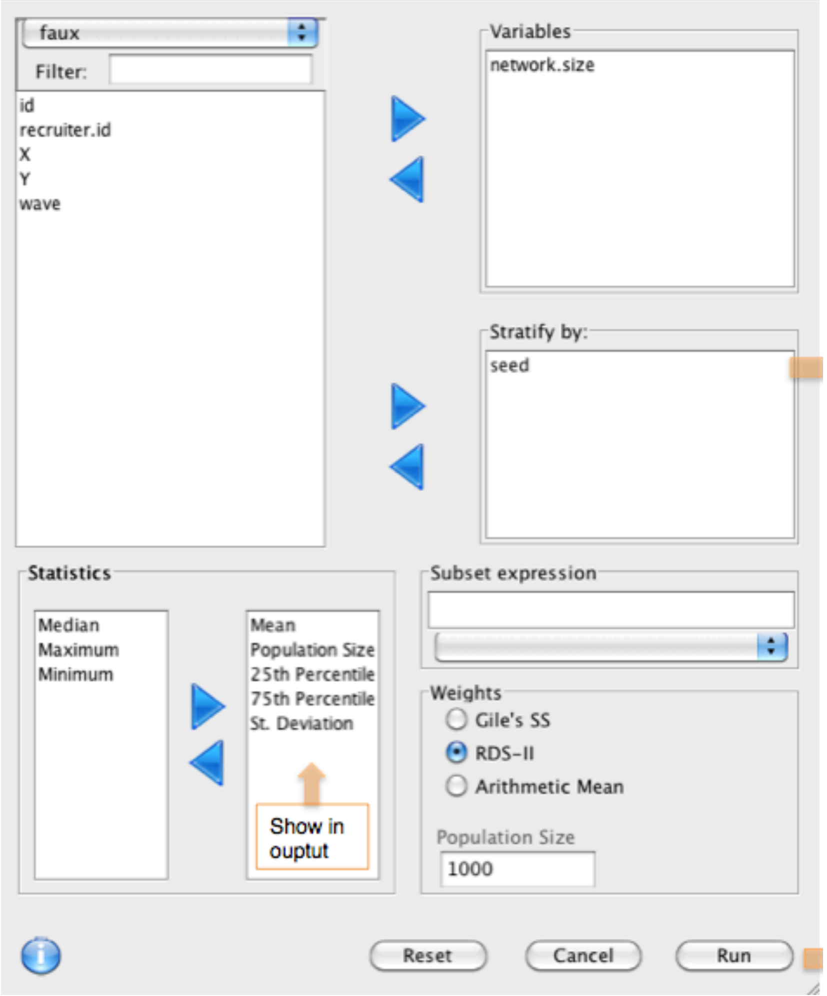
Good



Bad



The Dialog: Continuous Variables



faux

Filter:

id
recruiter.id
X
Y
wave

Variables: network.size

Stratify by: seed

Statistics

Median	Mean
Maximum	Population Size
Minimum	25th Percentile
	75th Percentile
	St. Deviation

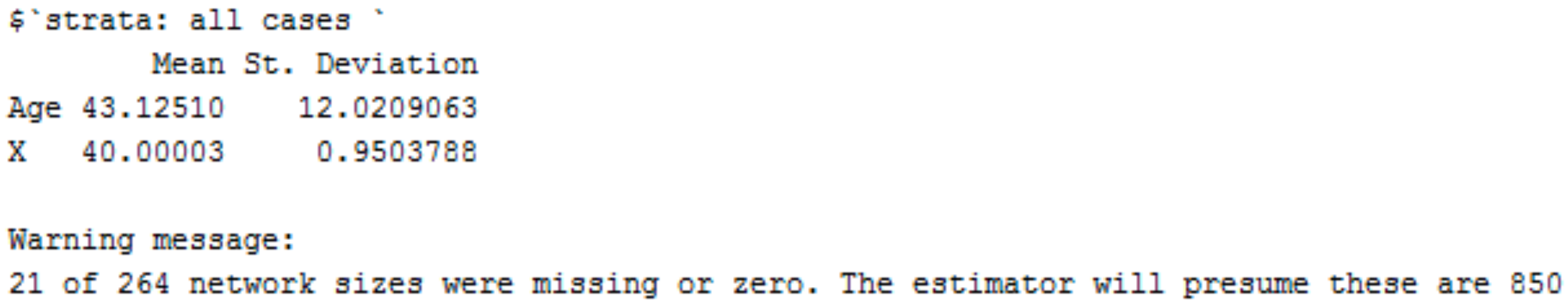
Show in ouptut

Subset expression:

Weights
- ○ Gile's SS
- ● RDS-II
- ○ Arithmetic Mean

Population Size: 1000

Continuous Variables: Output



```
$`strata: all cases `
        Mean St. Deviation
Age 43.12510    12.0209063
X   40.00003     0.9503788

Warning message:
21 of 264 network sizes were missing or zero. The estimator will presume these are 850
```