

Two methods comparison based on one simulation sample

Xijin

10/24/2019

Motivation

In attempt to figure out the difference of these two modern methods for sample size calculation, we would work on one single simulated data set and then use both methods. Whole process works like:

1. Simulate an original dataset
2. Calculate required sample size by Riley three-criteria method
3. Re-simulate a sample just change the size of the sample to satisfy three-criteria
4. Develop a model use Maximum likelihood and Heuristic shrinkage modeling method
5. Calculate predictive error metrics (*Brier*, *MAPE*, *MSPE*)
6. Based on values of these predictive error metrics, recalculate required sample size by metamodel (VanSmeden method)

Sample size calculation by three criteria (1 and 2)

Table 1: Characteristics of data (Original, n=100)

Name	Values
Number of predictors	4
Outcome proportion	0.44

Table 2: Anticipated values by Riley

Name	Values
Shrinkage factor	0.9
Adjusted Cox-Snell R-squared	0.112
Absolute difference	0.05
Margin of error	0.05

Table 3: Required sample size by Riley three-criteria method

Name	Values
Relative drop	301
Absolute difference	99
Precise estimation	379
Minimumized size	379

Predictive error of model with the sample satisfying Riley's three-criteria method

Calculation of predicitive error and metamodel result for sample size based on Maximum likelihood modeling strategy

$$N_{MSPE} = \exp((-0.59 - \log(MSPE) + 0.36 * \log(Ef) + 0.94 * \log(P))/1.06)$$

$$N_{MAPE} = \exp((-0.48 - \log(MAPE) + 0.31 * \log(Ef) + 0.48 * \log(P))/0.53)$$

$$N_{Brier} = \exp((-0.91 - \log(Brier) + 0.62 * \log(Ef) + 0.04 * \log(P))/0.04)$$

Table 4: Predictive error based on ML

Name	Values
Brier	0.236
MAPE	0.0218
MSPE	7e-04

Table 5: Calculated sample size by metamodel results

Name	Values
Sample size by Brier	7
Sample size by MAPE	1198
Sample size by MSPE	1404

```
## Intercept          X1          X2          X3          X4
## -0.06325362  0.20691319  0.22399564  0.25876583  0.30382034
```

Calculation of predicitive error and metamodel result for sample size based on Heuristic shrinkage modeling strategy

$$N_{MSPE} = \exp((-0.75 - \log(MSPE) + 0.44 * \log(Ef) + 0.74 * \log(P))/0.97)$$

$$N_{MAPE} = \exp((-0.56 - \log(MAPE) + 0.33 * \log(Ef) + 0.39 * \log(P))/0.49)$$

$$N_{Brier} = \exp((-0.93 - \log(Brier) + 0.62 * \log(Ef) + 0.02 * \log(P))/0.03)$$

Table 6: Predictive error based on HS

Name	Values
Brier	0.2367
MAPE	0.0309
MSPE	0.0014

Table 7: Calculated sample size by metamodel results

Name	Values
Sample size by Brier	3
Sample size by MAPE	10
Sample size by MSPE	4

```
## (Intercept)          X1          X2          X3          X4
##  0.03666769  0.16959489  0.18359640  0.21209554  0.24902414
```

Comparison of two sample size computation method (Riley and van Smeden)

Table 8: Riley method

Name	Values
Relative drop	301
Absolute difference	99
Precise estimation	379
Minimized size	379

Table 9: Van Smeden (ML)

Name	Values
Brier	7
MAPE	1198
MSPE	1404

Table 10: Van Smeden (HS)

Name	Values
Brier	3
MAPE	10
MSPE	4

Randomization is introduced during the process of modeling.

