Two methods comparison based on one simulation sample

Xijin

10/24/2019

Motivation

In attemplt 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 requried 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 thses 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	0.14

Table 2: Anticipated values by Riley

Name	Values
Shrinkage factor	0.9
Adjusted Cox-Snell R-squared	0.0833
Absolute difference	0.05
Margin of error	0.05

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

Name	Values
Relative drop	412
Absolute difference	136
Precise estimation	185
Minimumized size	412

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

Calculation of predicitve 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	0.1027
MAPE	0.0393
MSPE	0.0035

Table 5: Calculated sample size by metamodel results

Name	Values
Sample size by Brier	158
Sample size by MAPE	202
Sample size by MSPE	208

```
##  Intercept          X1          X2          X3          X4
## -2.1931554   0.7335201   0.3452103   0.2760173   0.5496961
```

Calculation of predicitve 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	0.1027
MAPE	0.0389
MSPE	0.0038

Table 7: Calculated sample size by metamodel results

Name	Values
Sample size by Brier	173
Sample size by MAPE	27
Sample size by MSPE	6

```
## (Intercept)          X1          X2          X3          X4
##  -1.9448359   0.6550690   0.3082895   0.2464969   0.4909053
```

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

Table 8: Riley method

Name	Values
Relative drop	412
Absolute difference	136
Precise estimation	185
Minimized size	412

Table 9: Van Smeden (ML)

Name	Values
Brier	158
MAPE	202
MSPE	208

Table 10: Van Smeden (HS)

Name	Values
Brier	173
MAPE	27
MSPE	6

Randomization is introduced during the process of modeling.