Optimal Data Analysis
Vol. 7 (November 17, 2018), 82-84

2155-0182/10/$3.00

Comparing Exact Discrete 95% CIs for Model *vs.* Chance ESS to Evaluate Statistical Significance

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

Satisfaction ratings (1=very dissatisfied; 2=somewhat dissatisfied; 3=neutral; 4=somewhat satisfied; 5=very satisfied) provided by 4,583 hospital patients in three successive cohorts (two consecutive 3-month-long *baseline* cohorts and one 3-month-long *post-intervention* cohort) were compared to evaluate a program which was designed to increase patient-rated satisfaction with in-hospital received care (Table 1).

Table 1: Patient-Rated In-Hospital[1] Satisfaction

	Study Cohort		
Satisfaction Rating	First-Baseline	Second-Baseline	Post-Intervention
5	596	586	716
4	357	480	557
3	432	345	287
2	60	75	32
1	45	15	0

Each of the three pairwise comparisons of the three cohorts was modeled via the novometric algorithm, treating satisfaction rating as an ordered attribute. The confirmatory ("one-tailed") alternative hypothesis is satisfaction ratings are greater post-intervention *vs.* in either baseline cohort. The exploratory ("two-tailed") alternative hypothesis is satisfaction ratings differed between baseline cohorts. If a pairwise comparison involved a descendant family with multiple models, the model having the lowest D statistic was selected as the globally-optimal or GO solution. Exact discrete 95% CIs for model accuracy (ESS and D) were obtained for models via 10,000 bootstrap iterations, and obtained for chance via 10,000 Monte Carlo experiments: if 95% CIs for model and for chance do *not* overlap the modeled effect is statistically significant. Reported models had stable ESS in training and leave-one-out (jackknife) analysis. [2-21]

For the *first-baseline vs. post-intervention* comparison the GO model was: if satisfaction rating≤3 predict first-baseline; otherwise predict post-intervention. This model correctly predicted the actual class status of 537/1,490 (36.0%) first-baseline and 1,293/1,592 (80.0%) post-intervention observations, yielding relatively weak[2,22] ESS=16.0 (95% CI for model= 12.2-19.7; 95% CI for chance= 0.11-3.14): the CIs for model and chance ESS do not overlap, so this model is statistically significant. For this model, D=10.5 (95% CI=8.15-14.4).

For the *second-baseline vs. post-intervention* comparison the GO model was: if satis-

2155-0182/10/$3.00

faction rating≤3 predict second-baseline; otherwise predict post-intervention. The model correctly predicted the actual class status of 435/1,501 (29.0%) second-baseline and 1,273/1,592 (80.0%) post-intervention observations, yielding weak ESS=8.94 (95% CI for model=5.35-12.5; 95% CI for chance= 0.12-2.99): the CIs for model and chance ESS do not overlap, so this model is statistically significant. For this model, D=22.4 (95% CI=14.0-35.4).

Finally, for the *first- vs. second-baseline* comparison the GO model was: if satisfaction rating≤3 then predict first-baseline; otherwise predict second-baseline. The model correctly predicted the actual class status of 537/1,490 (36.0%) first-baseline and 1,066/1,501 (71.0%) second-baseline observations, yielding weak ESS=7.06 (95% CI for model=3.04-11.1; 95% CI for chance=0.11-3.37): the CIs for model and chance ESS overlap, so this model is not statistically significant. For this model, D=26.3 (95% CI=16.0-63.8).

In summary, post-intervention satisfaction ratings were significantly greater than first- or second-baseline ratings (which could not be discriminated from each other). The cohorts studied here were temporally defined, but this is not a requirement—these methods may be used with cohorts assessed at a single point in time.

References

[1]Drawn from a confidential proprietary analysis, these data are presented anonymously with the permission of the owner.

[2]Yarnold PR, Soltysik RC (2014). Discrete 95% confidence intervals for ODA model- and chance-based classifications. *Optimal Data Analysis*, *3*, 110-112.

[3]Yarnold PR, Soltysik RC (2014). Globally optimal statistical classification models, I: Binary class variable, one ordered attribute. *Optimal Data* Analysis, *3*, 55-77.

[4]Yarnold PR, Soltysik RC (2016). *Maximizing predictive accuracy*. Chicago: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

[5]Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice, 22*, 839-847.

[6]Linden A, Yarnold PR (2016). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice, 22*, 848-854.

[7]Linden A, Yarnold PR. Using machine learning to identify structural breaks in single-group interrupted time series designs (2016). *Journal of Evaluation in Clinical Practice, 22*, 855-859.

[8]Linden A, Yarnold PR, Nallomothu BK (2016). Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice, 22*, 860-867.

[9]Yarnold PR, Linden A (2016). Using machine learning to model dose-response relationships via ODA: eliminating response variable baseline variation by ipsative standardization. *Optimal Data Analysis, 5*, 41-52.

[10]Linden A, Yarnold PR (2016). Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice, 22*, 868-874.

[11]Linden A, Yarnold PR (2016). Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice, 22*, 875-885.

[12]Yarnold PR, Linden A (2016). Novometric analysis with ordered class variables: The optimal alternative to linear regression analysis. *Optimal Data Analysis, 5*, 65-73.

2155-0182/10/$3.00

[13]Yarnold PR, Linden A (2016). Theoretical aspects of the D statistic. *Optimal Data Analysis*, *5*, 171-174.

[14]Linden A, Yarnold PR (2017). Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice*, *23*, 703-712.

[15]Yarnold PR, Linden A (2017). Computing propensity score weights for CTA models involving perfectly predicted endpoints. *Optimal Data Analysis*, *6*, 43-46.

[16]Linden A, Yarnold PR (2017). Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *23*, 1299-1308.

[17]Linden A, Yarnold PR (2017). Minimizing imbalances on patient characteristics between treatment groups in randomized trials using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *23*, 1309-1315.

[18]Linden A, Yarnold PR (2018). The Australian gun buy-back program and the rate of suicide by firearm. *Optimal Data Analysis*, *7*, 28-35.

[19]Linden A, Yarnold PR (2018). Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *24*, 353-361.

[20]Linden A, Yarnold PR (2018). Estimating causal effects for survival (time-to-event) outcomes by combining classification tree analysis and propensity score weighting. *Journal of Evaluation in Clinical Practice*, *24*, 380-387.

[21]Linden A, Yarnold PR (2018). Using machine learning to evaluate treatment effects in multiple-group interrupted time series analysis. *Journal of Evaluation in Clinical Practice*, *24*, 740-744.

[22]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*, Washington, DC, APA Books.

Author Notes

No conflict of interest was reported.