Optimal Data Analysis
Vol. 6 (January 5, 2017), 18-24

2155-0182/10/$3.00

Novometric Pairwise Comparisons in Consolidated Temporal Series

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

A total of 6,005 hospital patients rated their satisfaction with time taken for a nurse to respond to the call button (1=very dissatisfied; 2=somewhat dissatisfied; 3=neutral; 4=somewhat satisfied; 5=very satisfied).[1] Responses were consolidated into four successive quarters (3rd and 4th Quarter of 1989, 1st and 2nd Quarter of 1990). Novometric analysis[2-11] is used to evaluate statistical and ecological significance of all pairwise comparisons between the four quarters within this consolidated temporal series.

Data (N for each rating level and fiscal quarter) analyzed herein are given in Table 1: patient satisfaction rating was treated as an ordered attribute, and quarter as the class variable.[2]

Table 1: Patient Satisfaction with Nurse Response Time to Call Button[1]

	Fiscal Quarter			
Satisfaction Rating	Q3 (1989)	Q4 (1989)	Q1 (1990)	Q2 (1990)
5	596	586	626	716
4	357	480	412	557
3	432	345	313	287
2	60	75	57	32
1	45	15	14	0

All pairwise comparisons between the four quarters involved binary class variables (i.e., Q_x *vs*. Q_y) modeled by enumerated-optimal classification tree analysis (EO-CTA). All of the reported effects satisfied the Sidak criterion for experimentwise $p<0.05$, all associated models had identical accuracy in training and leave-one-out jackknife analysis, and exact discrete 95% confidence intervals (CIs) for predictive accuracy normed for chance (ESS) as well as for parsimony (D) were obtained for models using 10,000 bootstrap iterations, and for chance using 10,000 Monte Carlo experiments: if the 95% CIs for model and chance overlap then the model is judged statistically unreliable.[12--17]

A single optimal model emerged for three pairwise comparisons—all involving Q2. In these applications (an application is defined as the intersection of hypothesis, measures, and sample), the single emergent model was therefore the globally-optimal (GO) model.[2]

For the Q1 *vs*. Q2 pairwise comparison the GO model was: if satisfaction score≤3 then predict Q1; otherwise predict Q2 (i.e., Q1 had lower satisfaction scores than Q2). Table 2 is the model confusion matrix (Sens=model sensitivity). As seen, the model correctly predicted the actual class status of 5 of 9 Q1 observations

2155-0182/10/$3.00

(50% sensitivity is expected by chance), and of 4 of 5 Q2 observations: the effect is dominated by the preponderance of scores >3 in Q2 *vs*. Q1. The model yielded moderate ESS=35.9 (95% CI for model=32.0-39.8; 95% CI for chance=0.13-3.45). The 95% CIs for model and chance ESS do not overlap, so the model is statistically reliable. For this model D=3.57 (95% CI=3.03-4.25).

Table 2: Confusion Matrix for EO-CTA Model: Q1 *vs*. Q2

		Predicted Quarter		
		Q1	Q2	Sens
Actual	Q1	796	626	56.0
Quarter	Q2	319	1273	80.0

For the Q4 *vs*. Q2 pairwise comparison the GO model was: if satisfaction score≤3 then predict Q4; otherwise predict Q2. Table 3 is the confusion matrix for this model, which correctly predicted the actual class status of 3 of 10 of the Q4 observations (markedly less than expected by chance), and 4 of 5 Q2 observations. The model had weak ESS=8.94 (95% CI for model= 5.35-12.5; 95% CI for chance=0.12-2.99): since the 95% CIs for model and chance ESS do not overlap, the model is statistically reliable. For this model D=22.4 (95% CI=14.0-35.4).

Table 3: Confusion Matrix for EO-CTA Model: Q4 *vs*. Q2

		Predicted Quarter		
		Q4	Q2	Sens
Actual	Q4	435	1066	29.0
Quarter	Q2	319	1273	80.0

Finally, for the Q3 *vs*. Q2 pairwise comparison the GO model was: if satisfaction score ≤3 then predict Q3; otherwise predict Q2. Table 4 gives the confusion matrix for this model that correctly predicted the actual class status of 3 of 8 Q3 observations, and 4 of 5 Q2 observations. The model yielded relatively weak ESS=16.0 (95% CI for model=12.2-19.7; 95% CI for chance=0.11-3.14): 95% CIs for model and chance ESS do not overlap, so this model is statistically reliable. For this model D=10.5 (95% CI=8.15-14.4).

Table 4: Confusion Matrix for EO-CTA Model: Q3 *vs*. Q2

		Predicted Quarter		
		Q3	Q2	Sens
Actual	Q3	537	953	36.0
Quarter	Q2	319	1273	80.0

For the remaining three pairwise models more than one optimal model was identified.[18-20]

Two optimal models emerged for the Q3 *vs*. Q4 pairwise comparison. The first, and most complex (non-linear, three-strata) model is seen in Figure 1 (given beneath endpoints, numerator is number correctly classified and denominator is total number of observations in the endpoint).

Figure 1: Q3 *vs*. Q4 Three-Strata Model

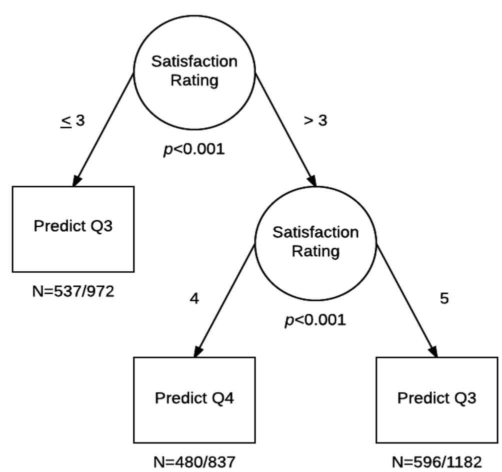

Table 5 is the confusion matrix for this model, which correctly predicted actual class status of 3 of 4 Q3 observations, and 1 of 3 Q4 observations. The model had weak ESS=8.02

2155-0182/10/$3.00

(95% CI for model=4.23-11.9; 95% CI for chance=0.13-3.20): 95% CIs for model and chance ESS do not overlap, so this model is statistically reliable. For this model D=34.4 (95% CI=22.2-67.9).

Table 5: Confusion Matrix for EO-CTA Model: Q3 *vs*. Q4, Three-Strata Model

		Predicted Quarter		
		Q3	Q4	Sens
Actual	Q3	1133	357	76.0
Quarter	Q4	1021	480	32.0

The second (two-strata) model was: if satisfaction score≤3 then predict Q3; otherwise predict Q4. Table 6 is the confusion matrix for this model, which correctly predicted the actual class status of 3 of 8 of the Q3 observations, and 7 of 10 of the Q4 observations. The model had weak ESS=7.06 (95% CI for model=3.04-11.1; 95% CI for chance=0.11-3.37): 95% CIs for model and chance ESS overlap, so the model is statistically unreliable. For this model D=26.3 (95% CI=16.0-63.8).

Table 6: Confusion Matrix for EO-CTA Model: Q3 *vs*. Q4, Two-Strata Model

		Predicted Quarter		
		Q3	Q4	Sens
Actual	Q3	537	953	36.0
Quarter	Q4	435	1066	71.0

For this pairwise comparison the less complex two-strata model is not statistically reliable, whereas the more complex three-strata model is statistically reliable. The three-strata model is thus the GO model in this application.

Three optimal models emerged for the Q4 *vs*. Q1 pairwise comparison. The first, most complex (non-linear, four-strata) model is given in Figure 2. Table 7 is the confusion matrix for this model, which correctly predicted the actual class status of 3 of 8 Q4 observations, and 19 of 20 Q1 observations. The model had moderate ESS=33.0 (95% CI for model=29.7-36.2; 95% CI for chance=0.11-3.04): 95% CIs for model and chance ESS do not overlap, so the model is statistically reliable. For this model D=8.13 (95% CI=7.05-9.47).

Figure 2: Q4 *vs*. Q1 Four-Strata Model

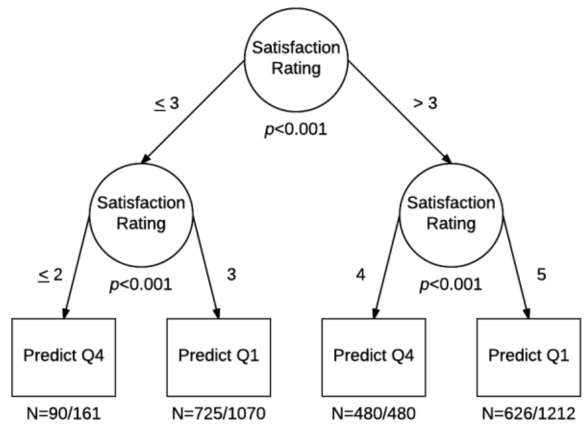

Table 7: Confusion Matrix for EO-CTA Model: Q4 *vs*. Q1, Four-Strata Model

		Predicted Quarter		
		Q4	Q1	Sens
Actual	Q4	570	931	38.0
Quarter	Q1	71	1351	95.0

The second (non-linear, three-strata) model is presented in Figure 3. Table 8 presents the confusion matrix for this model, which correctly predicted the actual class status of 1 of 3 of Q4 observations, and of all Q1 observations. The model had moderate ESS=32.0 (95% CI for model=29.2-34.8; 95% CI for chance=0.07-2.67): 95% CIs for model and chance ESS do not overlap, the model is statistically reliable. For this model D=26.3 (95% CI=5.62-7.27).

The third and final (two-strata) model was: if satisfaction score≤3 then predict Q1; otherwise predict Q4. Table 9 is the confusion matrix for this model, which correctly predicted the actual class status of 7 of 10 Q4 observa-

2155-0182/10/$3.00

tions, and 5 of 9 Q1 observations. The model had weak-moderate ESS=27.0 (95% CI for model=22.9-31.1; 95% CI for chance=0.12-3.58): 95% CIs for model and chance ESS do not overlap, the model is statistically reliable. For this model D=5.41 (95% CI=4.43-6.73).

Figure 3: Q4 *vs*. Q1 Three-Strata Model

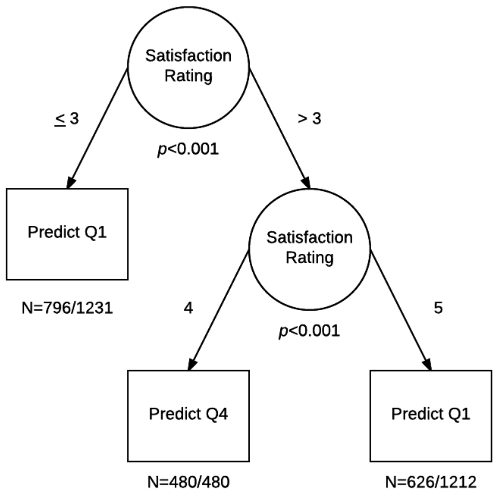

Table 8: Confusion Matrix for EO-CTA Model: Q4 *vs*. Q1, Three-Strata Model

		Predicted Quarter		
		Q4	Q1	Sens
Actual	Q4	480	1021	32.0
Quarter	Q1	0	1422	100.0

Table 9: Confusion Matrix for EO-CTA Model: Q4 *vs*. Q1, Two-Strata Model

		Predicted Quarter		
		Q4	Q1	Sens
Actual	Q4	1066	435	71.0
Quarter	Q1	626	796	56.0

For this pairwise comparison the least complex two-strata model had the lowest D statistic (by point estimate), and thus is the GO model in this application. However, 95% CIs for the more complex three- and four-strata models overlap the 95% CI for the two-strata model. The decision regarding the appropriate model to use in a future application is thus a function of factors such as *a priori* theory, sample size (the first axiom of novometric theory requires adequate statistical power for all hypothesis tests), and objective (e.g., sometimes models with a small D statistic are sought, and sometimes models with a large D statistic are sought[8]).

Finally, three optimal models emerged for the Q3 *vs*. Q1 pairwise comparison. The first, most complex (non-linear, four-strata) model is presented in Figure 4.

Figure 4: Q3 *vs*. Q1 Four-Strata Model

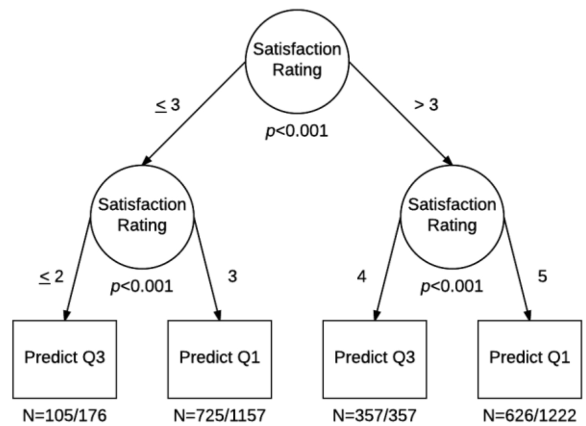

Table 10 presents the confusion matrix for this model, which correctly predicted the actual class status of 3 of 10 Q3 observations, and 19 of 20 Q1 observations.

Table 10: Confusion Matrix for EO-CTA Model: Q3 *vs*. Q1, Four-Strata Model

		Predicted Quarter		
		Q3	Q1	Sens
Actual	Q3	462	1028	31.0
Quarter	Q1	71	1351	95.0

The model had moderate ESS=26.0 (95% CI for model=22.9-29.1; 95% CI for

2155-0182/10/$3.00

chance=0.10-2.85): 95% CIs for model and chance ESS do not overlap, so the model is statistically reliable. For this model D=11.4 (95% CI=9.75-13.5).

The second (non-linear, three-strata) model is presented in Figure 5.

Figure 5: Q3 *vs*. Q1 Three-Strata Model

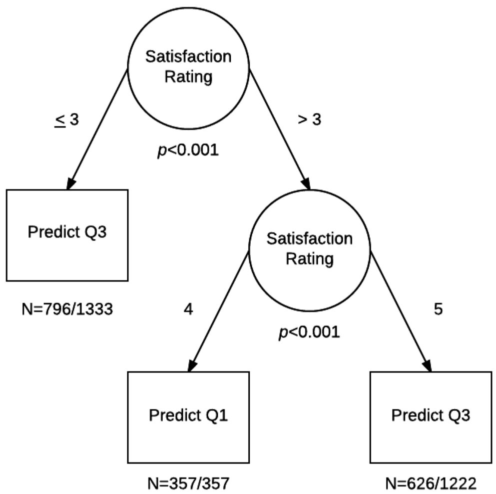

Table 11 is the confusion matrix for this model, which correctly predicted the actual class status of 1 of 4 of Q3 observations, and of all Q1 observations. The model had moderate ESS=24.0 (95% CI for model=21.4-26.6; 95% CI for chance=0.09-2.38): 95% CIs for model and chance ESS do not overlap, so the model is statistically reliable. For this model D=9.52 (95% CI=8.28-11.0).

Table 11: Confusion Matrix for EO-CTA Model: Q3 *vs*. Q1, Three-Strata Model

		Predicted Quarter		
		Q3	Q1	Sens
Actual	Q3	357	1133	24.0
Quarter	Q1	0	1422	100.0

The third and final (two-strata) model was: if satisfaction score≤3 then predict Q1; otherwise predict Q3. Table 12 is the confusion matrix for this model, which correctly predicted the actual class status of 5 of 8 Q3 observations, and 5 of 9 Q1 observations. The model yielded relatively weak ESS=19.9 (95% CI for model= 15.8-24.2; 95% CI for chance=0.13-3.70): 95% CIs for model and chance ESS do not overlap, so the model is statistically reliable. For this model D=8.03 (95% CI=6.26-9.38).

Table 12: Confusion Matrix for EO-CTA Model: Q3 *vs*. Q1, Two-Strata Model

		Predicted Quarter		
		Q3	Q1	Sens
Actual	Q3	953	537	64.0
Quarter	Q1	626	796	56.0

As was the case for the preceding (Q4 *vs*. Q1) analysis, for this (Q3 *vs*. Q1) pairwise comparison the least complex two-strata model had the lowest D statistic (by point estimate), and is therefore the GO model in this application. However, 95% CIs for the more complex three- and four-strata models overlap the 95% CI for the two-strata model.

Comments

Satisfaction dominance of Q2 1990 *vs*. the preceding three Quarters is revealed in the first three analyses—in which only one identical optimal model emerged: scores indicating satisfaction (4,5) were predicted to have been from Q2 of 1990, and scores reflecting ambivalence or dissatisfaction (1-3) were predicted to have been from earlier Quarters. Symbolic notation is used to summarize the findings graphically[21,22] (differences that may exist between the Quarters in brackets are not yet elucidated):

$$[Q3,Q4,Q1] < Q2 \quad . \qquad (1)$$

Including the two GO models involving binary parses (Q1<Q3, Q1<Q4), the findings are further summarized:

Q1 < [Q3,Q4] < Q2 . (2)

If the binary model for the Q3 *vs*. Q4 comparison *had been* statistically reliable, then a simple one-dimensional representation of the findings *would have been* possible. That is, since the binary comparison found Q3<Q4, the solution *would* be:

Q1 < Q3 < Q4 < Q2 . (3)

However, since the 95% CIs for model and chance for the binary Q3 *vs*. Q4 model overlapped—and the model was thus judged to be statistically unreliable, the three-strata model (Figure 1) was identified as the GO model in this comparison.

The integrated findings are summarized using symbolic notation, with parentheses used to indicate that a non-linear model discriminates the included Quarters.

Q1 < (Q3,Q4) < Q2 . (4)

Using novometric analysis with the present data, prior research[11] tested the directional hypothesis that satisfaction increased in quarter-to-quarter pairwise comparisons moving forward: that is, satisfaction in Q3 < Q4 (1989); Q4 (1989) < Q1 (1990); and Q1 < Q2 (1990). None of these confirmatory comparisons was statistically reliable, and the explanation why is clearly evident in the present exploratory findings: the opposite of each of these hypothesized effects emerged. Interestingly, in the confirmatory analysis[11] only one statistically reliable effect was identified (braces indicate a combined group that was evaluated in a pairwise comparison):

{Q3,Q4,Q1} < Q2 . (5)

The present study compared satisfaction ratings between independent, primarily independent temporally-consolidated groups. However, these methods may adapted to evaluate temporal data for single sample designs, and for designs involving multiple class categories that are examined across time.[23,24]

References

[1]Steiber SR, Krowinski WJ (1990). *Measuring and managing patient satisfaction*. Chicago, IL: American Hospital Publishing, Inc. (pp. 156-157).

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

[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 (2014). Globally optimal statistical classification models, II: Unrestricted class variable, two or more attributes. *Optimal Data Analysis*, *3*, 78-84.

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

[6]Yarnold PR, Bennett CL (2016). Novometrics *vs*. correlation: Age and clinical measures of PCP survivors, *Optimal Data Analysis*, *5*, 74-78.

[7]Yarnold PR, Bennett CL (2016). Novometrics *vs*. multiple regression analysis: Age and clinical measures of PCP survivors, *Optimal Data Analysis*, *5*, 79-82.

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

[9]Yarnold PR (2016). Using novometric analysis of transition matrices to ascertain Markovian order. *Optimal Data Analysis*, *6*, 4-7.

2155-0182/10/$3.00

[10] Yarnold PR (2016). Novometric comparison of Markov transition matrices for heterogeneous populations. *Optimal Data Analysis*, *6*, 8-11.

[11]Yarnold PR (2017). Novometric comparison of patient satisfaction with nurse responsiveness over successive quarters. *Optimal Data Analysis*, *6*, 12-16.

[12]Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*, *6*, 839-847.

[13]Yarnold PR (1996). Discriminating geriatric and non-geriatric patients using functional status information: An example of classification tree analysis via UniODA. *Educational and Psychological Measurement*, *56*, 656-667.

[14]Yarnold PR, Soltysik RC, Bennett CL (1997). Predicting in-hospital mortality of patients with AIDS-related *Pneumocystis carinii* pneumonia: An example of hierarchically optimal classification tree analysis. *Statistics in Medicine*, *16*, 1451-1463.

[15]Soltysik RC, Yarnold PR (2010). Automated CTA software: Fundamental concepts and control commands. *Optimal Data Analysis*, *1*, 144-160.

[16]Yarnold PR, Bryant FB (2015). Obtaining an enumerated CTA model via automated CTA software. *Optimal Data Analysis*, *4*, 54-60.

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

[18]Yarnold PR (2016). How many EO-CTA models exist in my sample and which is the best model? *Optimal Data Analysis*, *5*, 62-64.

[19]Yarnold PR (2016). Novometric theorem generalized to unrestricted class variables. *Optimal Data Analysis*, *5*, 102-103.

[20]Yarnold PR (2016). Refined definition of the descendant family in novometric theory. *Optimal Data Analysis*, *5*, 153.

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

[22]Yarnold PR (2013). Ascertaining an individual patient's *symptom dominance hierarchy*: Analysis of raw longitudinal data induces Simpson's Paradox. *Optimal Data Analysis*, *2*, 159-171.

[23]Linden A, Yarnold PR, Nallamothu BK (2016). Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice*, *6*, 860-867.

[24]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*, *6*, 875-885.

Author's Notes

Analyzed data are publically available, and no conflict of interest was reported.

Mail: Optimal Data Analysis, LLC
6348 N. Milwaukee Ave., Suite 163
Chicago, IL 60646