Optimal Data Analysis
Vol. 7 (September 29, 2018), 66-71

2155-0182/10/$3.00

Comparative Accuracy of a Diagnostic Index Modeled Using (Optimized) Regression *vs.* Novometrics

Ariel Linden, Dr.P.H. and Paul R. Yarnold, Ph.D.
Linden Consulting Group, LLC Optimal Data Analysis LLC

Diagnostic screening tests are used to predict an individual's graduated disease status which is measured on an ordered scale assessing disease progression (severity of illness). Maximizing the predictive accuracy of the diagnostic or screening test is paramount to correctly identifying an individual's actual score along the ordered continuum. The present study compares two approaches for mapping a statistical model to a diagnostic index in order to make accurate outcome predictions for individuals. The application involves a dataset composed of multiple biomedical voice measurements for 42 individuals with early-stage Parkinson's disease, who completed a six-month trial of a device for remote symptom progression telemonitoring. For 16 voice measures, each treated as a main effect, ordinary least-squares regression is used to predict baseline motor impairment component score. ODA is used to maximize accuracy of the regression model when it is mapped to the diagnostic index, and results are compared with accuracy achieved by the novometric solution.

Many diagnostic screening tests are employed to predict an individual's disease status, measured on a graduated, ordered ("continuous") scale of disease progression or severity of illness. The conventional modelling approach used in such analytic problems is least squares regression, in which the disease status score is regressed on an array of covariates treated as main effects. The predictive accuracy of such models is assessed by the R^2 statistic and root mean squared error (RMSE). Such models assume the relationships between disease status and covariates are linear, and are unlikely to produce accurate predictions along the entire continuum of values.[1]

Our recent series of papers demonstrated the use of ODA, CTA and novometric methods in analysis of observational data—and data from randomized controlled trials, in making causal inferences about treatment effects.[2-20] Although we unequivocally advocate using novometry to identify maximum-accuracy (optimal) solutions, most research currently uses regression-based models to generate diagnostic models. Thus, for exposition, here we demonstrate the use of ODA

2155-0182/10/$3.00

to maximize the accuracy of regression-based predictions mapped to an ordered diagnostic index, and compare the most accurate regression model possible with the novometric solution.

Methods

Data[21] were obtained from 42 people with early-stage Parkinson's disease, recruited to a 6-month trial of a remote telemonitoring symptom progression monitoring device. Data were individual's scores on the Unified Parkinson's Disease Rating Scale (UPDRS) which reflects both the presence and severity of symptoms (but not their underlying causes), as well as the predicted UPDRS score (P_{UPDRS}) obtained by regression analysis using UPDRS as the dependent variable and the 16 voice measurement variables as main-effect independent variables. Only the baseline measurements were used in order to obviate concerns over autocorrelated data.

The regression model ESS was assessed by treating P_{UPDRS} as an ordered ("continuous") attribute with 42 levels. Model accuracy was assessed by evaluating the fit between predicted and actual P_{UPDRS} scores based on class intervals computed for *integers*.[22-23] For exposition we also illustrate the effect of metric granularity on the ESS of the regression model by subjecting P_{UPDRS} to an *ordinal* transformation into an attribute having *7 levels* (sequential blocks of 6 ordered P_{UPDRS} scores), *6 levels* (sequential blocks of 7 ordered P_{UPDRS} scores), and *2 levels* (sequential blocks of 21 ordered P_{UPDRS} scores).

Novometric analysis was conducted treating UPDRS as an ordered class variable and P_{UPDRS} as an ordered attribute: no directional hypothesis was specified.[2,17,24]

Results

The regression analysis modeling UPDRS scores as a simple linear function of P_{UPDRS} scores was: UPDRS = 0.33 + 0.91* P_{UPDRS}. The model intercept did not differ significantly from zero (t=0.03, P< 0.98), however the P_{UPDRS} coefficient satisfied the generalized (per-comparison) criterion for statistical significance (t=2.03, P<0.049). For this model, used in training (full sample) analysis, R^2=0.094, indicating that UPDRS and P_{UPDRS} scores shared 9.4% of their variance (LOO analysis for regression modeling is not supported by most statistical software).

Table 1 presents UPDRS and P_{UPDRS} scores for all 42 observations, ordered from lowest to highest UPDRS Score. Assessed using one-unit class intervals the regression model correctly identified 5/42 (11.9%) observations. However, total percent accurate classification (PAC) is not normed against chance.[24-25]

Next, scores were splined to create three lower-granularity ordinal scales. The 7-class scale breaks 42 sorted UPDRS scores into 7 ordered groups (class levels) consisting of six scores apiece, and the 6-class scale breaks the 42 sorted UPDRS scores into 6 ordered groups (class levels) of seven scores apiece. The third, lowest-possible-granularity scale creates a two-class-level ("binary") class variable by breaking the 42 sorted UPDRS scores into two ordered groups of 21 scores apiece.

Consider first the 7-class spline. As seen in Table 1, the domain of the UPDRS values in Class 1 was 6 to 10.737, and the domain of the P_{UPDRS} values was 16.05 to 24.48—therefore all of the observations in this UPDRS score segment (class level 1 of the 7-category class variable) were misclassified, so the model sensitivity for class level 1 was 0%. Similarly, all of the observations in classes 2, 3, 4, 6, and 7 were misclassified (sensitivity= 0%). In class level 5, observations 37, 33, 25, and 39 were correctly classified (class 5 sensitivity = 4/6 = 66.67%). For this model ESS= -5.56, or 5.56% worse than expected by chance.

Consider next the 6-class spline. As seen in Table 1, the UPDRS domain in Class 1 was 6 to 11.078, and the domain of P_{UPDRS} values was 16.05 to 24.48—all observations in this segment of the UPDRS score scale (class level 1 of the

2155-0182/10/$3.00

Table 1

Subject ID, UPDRS Score, and P_{UPDRS} Score Computed using the Regression Model

18	6	21.795298
14	6.5651002	19.918833
27	7.3449001	16.048582
16	8.9390001	17.827396
15	9.3273001	19.485985
4	10.737	24.479759
2	11.078	17.683687
22	11.293	15.529094
10	12	21.960381
24	12.224	23.01252
23	12.288	19.752378
20	12.362	17.74873
7	15.234	23.928732
40	15.255	24.129951
19	15.991	19.789335
13	16.072001	24.171127
32	16.487	20.696474
9	17	21.093864
11	17.466	20.493387
8	18.093	21.049313
38	18.256001	21.334663
17	19.093	19.771648
36	19.656	14.085366
42	19.725	20.530104
12	20.896	20.852905
37	22.962	23.42835
33	23.326	24.678873
3	23.437	25.269072
25	24.205999	24.396626
39	25.033001	21.342693
26	25.264	20.291403
29	27.549	24.420732
21	27.612	19.35153
31	27.681	22.533978
6	27.882999	22.682585
1	28.198999	23.072725
41	29.211	24.891562
34	29.291	22.203928
5	31	21.816959
28	31.93	22.685665
30	32.535	17.21484
35	36.073002	20.675291

6-category class variable) were misclassified, and the sensitivity of the model for class level 1 was 0%. Similarly, all observations in class levels 2, 3, and 6 were misclassified. Four observations in class 4 (17, 42, 12, 37) were correctly classified: sensitivity for class 4= 57.14%. And, two observations in class 5 (25, 29) were correctly classified: sensitivity for class 4=28.57%. Here ESS= -2.9, or 2.9% worse than expected by chance.

Finally, consider the two-class spline. Seen in Table 1, the domain of UPDRS values for the first 21 observations was 6 to 18.256, and the P_{UPDRS} score of five observations (27, 16, 2, 22, 20) fell in this domain and so were correctly classified (sensitivity for class 1 = 23.81%). For class=2 (UPDRS domain = 19.093 to 36.073) two observations (36, 30) were misclassified (sensitivity for class 2 = 90.48%): $P<0.42$, ESS=14.29—a relatively weak effect.

Novometric Analysis

Analysis identified eight statistically viable ODA models predicting UPDRS as a function of P_{UPDRS} (both treated as ordered variables). Two identical models emerged with greatest ESS in LOO analysis—these were the only models having stable ESS in both total sample (training) and LOO (generalizability) analysis, involving the 15th and 16th largest UPDRS values in the sample. The model having the larger UPDRS value was selected on the basis of providing greatest statistical power.[24] The "globally optimal" (GO) model was:

If $P_{UPDRS} \leq 20.105$ Predict UPDRS$\leq$16.072,

and

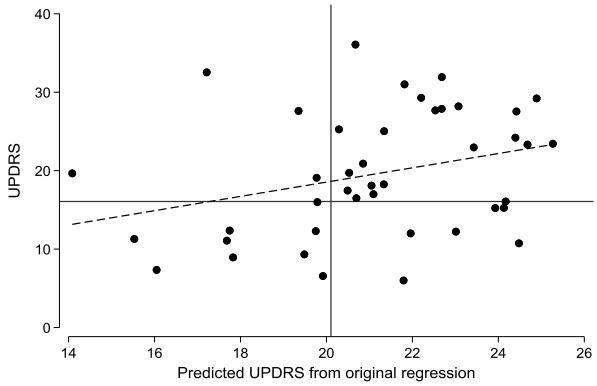
If $P_{UPDRS} > 20.105$ Predict UPDRS>16.072.

For this sample, a UPDRS score of 16.072 corresponds to normative $z<-0.39$, and a P_{UPDRS} score of 20.105 corresponds to normative $z<-0.44$.[24] This model correctly classified 9 of 15 (60%) observations with UPDRS$\leq$16.072,

2155-0182/10/$3.00

and 23 of 27 (85.2%) observations with UPDRS >16.072, in both training (P<0.027) and LOO (P<0.0039) analysis. For this model ESS=45.2, a moderate effect.[24,25] This GO model is shown in Figure 1, in which both UPDRS (16.072) and P_{UPDRS} (20.105) cut-points are plotted.

Figure 1: Scatterplot of UPDRS and P_{UPDRS} Scores, Illustrating Novometric and Regression Models



As seen, 23 observations falling in the upper right-hand quadrant created by the intersection of these cut-point-based axes are *correctly* classified by the ODA model as having relatively high UPDRS scores; 4 observations in the upper left-hand quadrant are *incorrectly* classified as having relatively high UPDRS scores; 6 observations in the lower right-hand quadrant are *incorrectly* classified as having relatively low UPDRS scores; and 9 observations in the lower left-hand quadrant are *correctly* classified as having relatively low UPDRS scores.

Discussion

For regression analysis only the 2-class-category ordinal transformation of P_{UPDRS} yielded a level of predictive accuracy (ESS=14.29) which exceeded what is expected by chance. However, this maximum-accuracy regression solution was not statistically reliable (P<0.41). In contrast, the novometric GO model identified a 2-class-category solution yielding ESS=45.20 (the minimum criterion for a relatively strong effect is ESS=50), which was statistically significant in both total sample "training" (P<0.027) and LOO (P<0.0039) analysis.[24-25]

In this paper we have demonstrated how ODA can be used to improve predictions of ordered (continuous) outcomes derived using conventional regression models. As seen, such models assume linearity over the entire continuum of values, which can result in highly inaccurate predictions at points along the continuum where the data are indeed non-linear. Given that most research currently uses regression-based models to generate diagnostic models, the issue of non-linearity is not trivial. As such we unequivocally advocate using ODA and CTA modeling approaches to identify maximum-accuracy (optimal) solutions that inherently identify maximum-accuracy, reproducible solutions for linear as well as non-linear phenomena.

References

1. Yarnold PR, Bryant FB, Soltysik RC (2013). Maximizing the accuracy of multiple regression models via UniODA: Regression *away from* the mean. *Optimal Data Analysis, 2*, 19-25.

2. Yarnold PR, Linden A. Novometric analysis with ordered class variables: The optimal alternative to linear regression analysis. *Optimal Data Analysis* 2016;22:65-73.

3. Linden A, Yarnold PR. Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice* 2016;22:839-847.

4. Linden A, Yarnold PR. Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice* 2016;22:848-854.

2155-0182/10/$3.00

5. Linden A, Yarnold PR. Using machine learning to identify structural breaks in single-group interrupted time series designs. *Journal of Evaluation in Clinical Practice* 2016;22:855-859.

6. Linden A, Yarnold PR, Nallomothu BK. Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice* 2016;22:860-867.

7. Yarnold PR, Linden A. Using machine learning to model dose-response relationships via ODA: eliminating response variable baseline variation by ipsative standardization. *Optimal Data Analysis* 2016;22:41-52.

8. Linden A, Yarnold PR. Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice* 2016;22:868-874.

9. Linden A, Yarnold PR. Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice* 2016;22:875-885.

10. Yarnold PR, Linden A. Theoretical aspects of the D statistic. *Optimal Data Analysis* 2016;22:171-174.

11. Linden A, Yarnold PR. Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice* 2017;23:703-712.

12. Yarnold PR, Linden A. Computing propensity score weights for CTA models involving perfectly predicted endpoints. *Optimal Data Analysis* 2017;6:43-46.

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

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

15. Linden A, Yarnold PR. The Australian gun buy-back program and the rate of suicide by firearm. *Optimal Data Analysis* 2018;7:28-35.

16. Linden A, Yarnold PR. Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice* 2018;24:353-361.

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

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

2155-0182/10/$3.00

19. Linden A, Yarnold PR. Using ODA in the evaluation of randomized controlled trials. *Optimal Data Analysis* 2018;7:46-49.

20. Linden A, Yarnold PR. Using ODA in the evaluation of randomized controlled trials: application to survival outcomes. *Optimal Data Analysis* 2018;7:50-53.

21. Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE transactions on Biomedical Engineering*. 2010;57:884-93. See also: https://archive.ics.uci.edu/ml/datasets/Parkinsons.

22. Yarnold PR. Alternative prediction-interval scaling strategies for regression models. *Optimal Data Analysis* 2018;7:44-45.

23. Akai TJ. *Applied numerical methods for engineers*. New York, NY: John Wiley & Sons, 1994, pp. 156-161.

24. Yarnold PR, Soltysik RC. *Maximizing Predictive Accuracy*. Chicago, IL: ODA Books, 2016. DOI: 10.13140/RG.2.1.1368.3286

25. Yarnold PR, Soltysik RC. *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books, 2005.

Author Notes

No conflict of interest was reported by either author.