Optimal Data Analysis
Vol. 5 (October 1, 2016), 94-101

2155-0182/10/$3.00

Matrix Display of Pairwise Novometric Associations for Ordered Variables

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

Correlation matrices are commonly used to summarize the linear associations for all possible pairings in a set of ordered variables. This paper exports this methodology, creating the conceptual analogue of a correlation matrix in the optimal data analysis (ODA) statistical paradigm.

Correlation matrices are an efficient means of summarizing linear relationships between all possible pairings in a set of ordered variables. For clarity this paper first illustrates matrix presentation of all possible Pearson[1] correlation (*r*) coefficients for a set of variables. Construction of a conceptually parallel pairwise information display for use in novometric[2] analysis is then demonstrated. These methods are illustrated for a sample of 2,011 patients who received care in an Emergency Room (ER), and later returned a completed mailed satisfaction survey.[3] Patient ratings of their ER nurse were made on ordered (categorical ordinal) Likert scales: 1=very poor, 2=poor, 3=fair, 4= good, 5=very good (Table 1).

Table 1: Descriptive Summary of Variables

Variable	N	Mean	SD
Courtesy	2,011	4.37	0.87
Problem Focused	2,009	4.31	0.94
Attentive	2,004	4.11	1.02
Informative	1,986	4.04	1.09
Protective of Privacy	1,965	4.12	1.01
Technical Skill	1,848	4.33	0.88

Correlation Analysis

Table 2 is the correlation matrix for the six nurse rating scores. Every cell in the matrix represents the pair of variables indicated by the column and row variables that intersect to form the cell. Five statistics given in every cell are: *r* (all $p<0.0001$); R^2 (the proportion of variance in common between the two variables); ESS (the effect strength for sensitivity, an index of the predictive accuracy of a model, on which 0=the level of accuracy expected by chance, and 100= errorless prediction[2]); D (the normed distance between an empirical and an ideal model with perfect accuracy and maximum parsimony[2]); and the number of patients with complete data on both variables. Unity values in the matrix diagonal indicate a perfect correlation between each variable and itself: for errorless models R^2=1, ESS= 100 and D=0.

In cells above the diagonal the variables heading columns were treated as independent variables (attributes), and in cells beneath the diagonal variables heading rows were treated as dependent (class) variables.[2] For example, in row 2 (problem focused), column 1 (courtesy),

Optimal Data Analysis
Vol. 5 (October 1, 2016), 94-101

2155-0182/10/$3.00

Table 2: Correlation Matrix for Six Nurse Rating Variables

Dependent or Class Variable		Independent Variable or Attribute: Courtesy	Problem Focused	Attentive	Informative	Protective of Privacy	Technical Skill
Courtesy	*r*	1.00	0.856	0.793	0.741	0.721	0.774
	R^2	1.00	0.733	0.629	0.548	0.520	0.599
	ESS	100	47.32	33.25	27.22	24.64	36.24
	D	0	5.57	10.04	13.37	15.30	8.80
	N	2,011	2,009	2,004	1,986	1,965	1,848
Problem Focused	*r*	0.856	1.00	0.847	0.775	0.739	0.769
	R^2	0.733	1.00	0.717	0.601	0.546	0.592
	ESS	53.90	100	38.53	29.38	23.74	33.68
	D	4.28	0	7.98	12.02	16.06	9.85
	N	2,009	2,009	2,002	1,984	1,963	1,846
Attentive	*r*	0.793	0.847	1.00	0.860	0.787	0.749
	R^2	0.629	0.717	1.00	0.739	0.619	0.561
	ESS	39.42	53.64	100	48.46	38.24	40.61
	D	7.68	4.32	0	5.32	8.08	7.31
	N	2,004	2,002	2,004	1,979	1,960	1,842
Informative	*r*	0.741	0.775	0.860	1.00	0.807	0.759
	R^2	0.548	0.601	0.739	1.00	0.651	0.576
	ESS	30.56	40.39	58.50	100	52.51	38.58
	D	11.36	7.38	3.55	0	4.52	7.96
	N	1,986	1,984	1,979	1,986	1,944	1,829
Protective of Privacy	*r*	0.721	0.739	0.787	0.807	1.00	0.764
	R^2	0.520	0.546	0.619	0.651	1.00	0.584
	ESS	36.12	43.36	39.59	34.52	100	43.67
	D	8.84	6.53	7.63	9.48	0	6.45
	N	1,965	1,963	1,960	1,944	1,965	1,821
Technical Skill	*r*	0.774	0.769	0.749	0.759	0.764	1.00
	R^2	0.599	0.592	0.561	0.576	0.584	1.00
	ESS	36.54	34.32	23.72	25.84	24.32	100
	D	8.69	9.57	16.08	14.35	15.56	0
	N	1,848	1,846	1,842	1,829	1,821	1,848

2155-0182/10/$3.00

courtesy is the independent variable (attribute), and problem solving is the dependent (class) variable. However, in row 1 (courtesy), column 2 (problem focused), the variable roles are reversed: courtesy is the dependent (class) variable, and problem solving the independent variable (attribute).

With respect to *r* and R^2, matrix cells are symmetric about the diagonal: findings for row A and column B are isomorphic with findings for row B and column A. Thus, in correlation analysis either variable may assume the role of independent or dependent variable without affecting the model (*r*), or the associated estimated *p* value or effect strength index (R^2).

Using *r* to make point predictions of the value of the dependent variable for individual observations is straightforward.[4] For example, when data are expressed in raw-score form, the regression model (the equivalent to *r* used with normatively standardized data) for predicting courtesy scores as a linear function of problem focused scores, where Y* is an observation's predicted courtesy score, is: Y*=1.001+0.784* problem focused. This model maximizes the proportion of variance in the dependent measure that is explained by (i.e., "is in common with") the independent variable.

Using the model to classify the sample yields the confusion matrix in Table 3: model sensitivity in correctly classifying each category of actual courtesy score is given at the end of each row; model predictive value in making correct point predictions into each rating category (courtesy score) is given under each column. If the data represent uniform random numbers (all values are equally likely), then a sensitivity of 100%/5=20% is expected for each actual courtesy level.[2] As seen, the *r* model was much less accurate than chance in predicting courtesy scores of 1, slightly more accurate than chance in predicting scores of 2, and much more accurate than chance in predicting scores of 3-5.

Predictive accuracy of any classification algorithm, summarized in the confusion matrix, is computed as the effect strength for sensitivity (ESS) index.[2,5] First compute C*=100/C, where C=number of levels of the dependent (class) variable. Here 5 levels (1-5) are being predicted, so C*=100/5=20. Next, compute ESS=[(mean sensitivity across all predicted categories–C*)/(100-C*)]*100%. In Table 3, C*=100/5=20, and mean sensitivity=(0+34.4+80.7+75.0+99.2)/5= 57.86. Thus, ESS=[(57.86-20)/(100-20)]x100% =(37.86/80)*100%=47.3%. In every application, ESS=0 is the level of predictive accuracy that is expected by chance, and ESS=100 is errorless prediction. Simulation research produced the rule-of-thumb used for qualitative interpretation of this index: ESS<25 is a relatively weak effect; ESS<50 is a moderate effect; ESS<75 is a relatively strong effect; ESS<90 is a strong effect; and ESS≥90 is a very strong effect.[2] For the predictive performance summarized in Table 3, ESS=47.3 (exact *p*<0.0001), indicating that the effect is of moderate strength (see Table 2).

Table 3: Confusion Matrix for *r* Model: Courtesy as Dependent Variable

	Predicted Courtesy					
Actual	1	2	3	4	5	
1	0	37	7	1	0	0%
2	0	11	21	0	0	34.4%
3	0	4	138	28	1	80.7%
4	0	3	79	513	89	75.0%
5	0	1	11	115	950	88.2%
	0%	19.6%	53.9%	78.1%	91.4%	

Although ESS is chance- and maximum-corrected (chance ESS=0; maximum ESS=100) for every application, ESS is not normed for complexity (the opposite of parsimony)—this is done using the D (distance) statistic.[2] According to novometric theory, for any application a theoretically ideal statistical model would achieve perfect accuracy using a minimum number of model endpoints—each representing a unique cluster of observations that are homogeneous within the endpoint and heterogeneous between different endpoints. Clusters of homogeneous observations are called strata—which is exactly

2155-0182/10/$3.00

what all optimal methods identify.[2] In any given application and model, D is the number of additional equivalent (i.e., having the same ESS as the model being assessed) effects needed to achieve a theoretically ideal model.

Thus D is an index of the distance of an empirical model from the theoretically ideal model for the application. ESS and D are both normed over chance and accuracy, but only D is additionally normed over parsimony. ESS may thus be used to compare the predictive accuracy of models having the same complexity (number of strata), however D must be used to compare models of different complexity levels. Nevertheless, while D is used to compute model effect strength quantitatively—*relative to a theoretical ideal*, ESS is used to qualitatively assess model effect strength *relative to chance* by the rule-of-thumb discussed earlier. To compute D, if S is the number of strata (endpoints) in the model, D=[100/(ESS/S)-S. Here ESS/S=44.57/5=8.914, and D=[100/8.914]-5=6.22 (see Table 2). In conclusion, using problem solving ratings (independent variable) to predict courtesy (dependent variable) yielded a moderate level of predictive accuracy in training (total sample) analysis: ESS=44.57, p<0.0001, D=6.218.

We next assess if measures of predictive accuracy (ESS, exact p, D) for the r model predicting courtesy using problem solving—and for the r model predicting problem solving using courtesy, are symmetric about the diagonal as are indices of shared variation (r, estimated p, R^2). Expressed in raw-score form, the regression model predicting problem focused scores as a linear function of courtesy scores is: Y*=0.204+ 0.936*courtesy, where Y* is the predicted problem solving score for an observation. Using this model to classify the data yields the confusion matrix in Table 4. As seen in Table 2, ESS= 53.90 (a relatively strong effect), p< 0.0001, and D=4.277. In contrast to measures of explained variation that are symmetric about the diagonal of the correlation matrix, measures of predictive accuracy are not similarly symmetric.

The factor underlying these asymmetric results is relative range restriction in the lowest rating level (score=1) of Table 3 *vs*. Table 4 (as seen, ratings of 1 were respectively available for 45 *vs*. 56 patients). In the "sample geometry" of the former problem compared to the latter there were insufficient N in score level=1 to sacrifice stronger results obtained for greater ratings, and thus zero patients were predicted to provide ratings of 1 (models failing to classify observations into all the class category levels are known as degenerate models[2,6]). Flatter, wider geometry underlying the latter problem led to 37 actual class=1 patients–who were misclassified as predicted class=2 patients in Table 3 (sensitivity= 0% for class=1)—being correctly classified in Table 4 (sensitivity=66.1% for class=1). All optimal models *must* conform to the geometry underlying applications to maximize ESS.[2]

Table 4: Confusion Matrix for r Model: Problem Focused as Dependent Variable

	Predicted Problem Focused					
Actual	1	2	3	4	5	
1	37	11	4	3	1	66.1%
2	6	16	33	5	1	26.2%
3	1	5	105	74	10	53.8%
4	1	0	28	513	115	78.1%
5	0	0	1	89	950	91.4%
	82.2%	50.0%	61.4%	75.0%	88.2%	

Inspection of Table 2 reveals that when variables were treated as being class variables, relatively strong models emerged for problem focused (predicted by courtesy, ESS=53.90, D= 4.28); attentive (predicted by problem focused, ESS=53.64, D=4.32); and informative (predicted by attentive, ESS=58.50, D=3.55): Figure 1 integrates the pattern of these findings.[2,6]

Figure 1: Integrated Relatively Strong Models: Predicting Dependent (Class) Variables

Courtesy → Problem Focused → Attentive

In contrast, when variables were treated as attributes, a relatively strong model emerged only for protective of patient privacy (predictive of informative, ESS=52.51, D=4.52). Treating technical skill as a class variable returned two relatively weak ESS values (with attentive and protective of patient privacy used as attributes), and treating protective of privacy as an attribute returned two relatively weak ESS values (with courtesy and problem focused used as class variables). All of the remaining ESS values were of moderate strength.

ODA[2,6] was used to compare ESS values (D would be compared if models in the matrix differed in complexity) above *vs*. below the diagonal: ESS=20.0, $p<0.92$, D=8.0.

Novometric Analysis

Prior to the development of novometric theory, predictive accuracy (ESS) of *r* models was explicitly maximized using ODA to adjust (i.e., to optimize) the decision thresholds *r* uses to make point predictions in an application.[2,4,6] Novometric analysis now enables identification of globally-optimal (GO) relationships involving thresholds for both the class variable and the attribute (subject to limitations associated with specific metrics used in the application)—that identify the model with the smallest D that it is possible to attain in the application.[2,7-11]

Statistical power analysis is required in novometric analysis[2] and indicated that $N\geq32$ patients should be classified into every model endpoint (i.e., all strata) for 90% power to detect a moderate effect with a generalized exploratory $p\leq0.05$. All rating categories of all variables surpassed this criterion, so four class dummy variables were constructed for each variable (i.e., 1 *vs*. 2-5; 1-2 *vs*. 3-5; 1-3 *vs*. 4-5; and 1-4 *vs*. 5).[2,7]

Novometric models were constrained to identify effects having identical ESS in training and LOO (single-sample jackknife) analysis in an effort to identify cross-generalizable models.[2]

Table 5 is the *novometric association matrix* for the six nurse rating scores. Matrix cells represent the pair of variables indicated by the column and row variables that intersect to form the cell. Four statistics given every cell are: model cutpoints for attribute and then class variable (see below); ESS (all *p*'s<0.0001); D; and number of patients with complete data on both variables. The diagonal cells are empty because in the ODA paradigm the same variable cannot be the attribute and the class variable. In cells above the diagonal variables were treated as attributes, and in cells beneath the diagonal variables were treated as class variables.

For example in row 2 (problem focused), column 1 (courtesy), the novometric model is: if courtesy $\leq$ 2 then predict problem focused $\leq$ 1; otherwise predict problem focused > 1. Note that in each cell of the matrix the first of the two cuts (cutpoints) given is for the attribute, and the second cutpoint is for the class variable.

Table 6 is the confusion matrix for this model applied to the data (50% sensitivity is expected by chance for each actual satisfaction class category). Using novometric analysis a statistically significant, cross-generalizable, and strong effect emerged: 85.7% (approximately 17 in 20) of the patients rating physician courtesy as poor (2) or worse rated physician problem focus as very poor (1), as compared to 100%-98.5% or 1.5% (1 in 67) patients rating courtesy as fair (3) or better.

Table 6: Confusion Matrix for Novometric Model: Courtesy Rating$\leq$2

		Predicted Problem Focus		
		1	>1	
Actual	1	48	8	85.7%
Focus	>1	29	1,924	98.5%
		62.3%	99.6%	

Next, for row 1 (courtesy), column 2 (problem focused), the novometric model is: if problem focused $\leq$ 2 then predict courtesy $\leq$ 1; otherwise predict courtesy > 1. Table 7 presents the confusion matrix for this model.

Table 5: Novometric Association Matrix for Six Nurse Rating Variables

Class Variable		Attribute: Courtesy	Problem Focused	Attentive	Informative	Privacy Concerned	Technical Skill
Courtesy	Cuts	---	2, 1	2, 1	2, 1	2, 1	3, 2
	ESS		91.79	86.70	81.68	85.13	81.64
	D		0.18	0.37	0.45	0.35	0.45
	N		2,009	2,004	1,986	1,965	1,848
Problem Focused	Cuts	2, 1	---	2, 1	1, 1	2, 1	3, 1
	ESS	84.23		86.69	81.49	76.14	77.65
	D	0.37		0.31	0.45	0.63	0.58
	N	2,009		2,002	1,984	1,963	1,846
Attentive	Cuts	3, 1	3, 1	---	1, 1	4, 4	3, 1
	ESS	80.24	82.93		85.23	77.48	78.36
	D	0.49	0.41		0.35	0.58	0.55
	N	2,004	2,002		1,979	1,960	1,842
Informative	Cuts	3, 1	4, 4	4, 4	---	4, 4	4, 4
	ESS	69.74	74.76	82.40		80.21	75.17
	D	0.87	0.68	0.43		0.49	0.66
	N	1,986	1,984	1,979		1,944	1,829
Privacy Concerned	Cuts	3, 1	3, 1	4, 4	2, 1	---	4, 4
	ESS	69.55	73.53	77.22	83.76		77.97
	D	0.88	0.72	0.59	0.39		0.56
	N	1,965	1,963	1,960	1,944		1,821
Technical Skill	Cuts	3, 1	2, 1	2, 1	2, 1	2, 1	---
	ESS	76.32	76.60	74.42	82.27	80.74	
	D	0.62	0.61	0.69	0.43	0.48	
	N	1,848	1,846	1,842	1,829	1,821	

Table 7: Confusion Matrix for Novometric Model: Problem Focused Rating≤2

		Predicted Courtesy: 1	>1	
Actual Focus	1	43	2	95.6%
	>1	74	1,890	96.2%
		36.8%	99.9%	

Using novometric analysis a statistically significant, cross-generalizable, and very strong effect emerged: 95.6% (approximately 19 in 20) of the patients rating physician problem focus as poor (2) or worse rated physician courtesy as very poor (1), as compared with 100%-96.2% or 3.8% (1 in 26) patients rating physician courtesy as fair (3) or better.

2155-0182/10/$3.00

Comments

Novometric results further demonstrate that findings regarding predictive accuracy in bivariate designs involving ordered variables are not symmetric in variable interrelationship matrices. Comparing ESS yielded by *r* versus novometric models for corresponding analyses (i.e., involving the identical class variable and attribute) revealed the novometric performance was always substantially stronger than obtained via *r*. For novometrics no ESS values fell into the moderate range: five were relatively strong, one was very strong, and the rest were strong effects. Comparing the obtained ESS values between regression and novometric models over all values in Table 2 and Table 5 using ODA, a perfect model emerged: 100% of the *r* models had ESS$\leq$64.05, and 100% of the novometric models had ESS>64.05 (ESS=100, D=0, $p<$ 0.0001).

Using *r* to assess the linear relationship that maximizes *explained variation* among two ordered variables, it is irrelevant which variable is independent and which is dependent: $r_{xy} = r_{yx}$ and *r*, *p*, and R^2 are symmetric about the correlation matrix diagonal. In contrast, using novometrics to identify the relationship that maximizes *predictive accuracy* normed over chance (ESS)—and additionally over complexity (D), selection of a variable to serve as the attribute (independent variable), and to serve as the class (dependent) variable, clearly influences model performance indices as a function of discrete distributional aspects such as domain and skew. One can envision a scenario in which results are statistically significant when assigning variables in one manner, but not in another manner. Conducting both sets of analyses and selecting one model over another after the fact simply because it yielded a statistically or ecologically stronger result is inappropriate. Rather, in an effort to minimize capitalization on chance, researchers should focus on constructing, testing, and reporting *a priori* hypotheses derived based on substantive and/or statistical theory.

Finally, while the present example used all ordered variables, novometric relationship matrices involving all binary attributes, all multicategorical attributes, or a mixture of categorical and ordered variables, all are equivalently straightforward.

References

[1]Pearson K (1895). Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*, 240-242.

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

[3]Yarnold PR, Michelson EA, Thompson DA, Adams SL (1998). Predicting patient satisfaction: A study of two emergency departments. *Journal of Behavioral Medicine*, *21*, 545-563.

[4]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.

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

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

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

2155-0182/10/$3.00

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

[10]Yarnold PR (2016). Novometrics *vs*. regression analysis: Literacy, and age and income, of ambulatory geriatric patients. *Optimal Data Analysis, 5*, 83-85.

[11]Yarnold PR (2016). Novometrics *vs*. regression analysis: Modeling patient satisfaction in the Emergency Room. *Optimal Data Analysis*, *5*, 86-93.

Author Notes

The study analyzed de-individuated data and was exempt from Institutional Review Board review. No conflict of interest was reported. This article serves as an inadequate tribute to the memory of my beloved mother.

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