Optimal Data Analysis
2013, Vol. 2, Release 1 (October 21, 2013), 76-85

2155-0182/10/$3.00

Assessing Technician, Nurse, and Doctor Ratings as Predictors of Overall Satisfaction of Emergency Room Patients: A Maximum-Accuracy Multiple Regression Analysis

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

This study extends recent quality-of-care research undertaken to enhance understanding of ratings of overall satisfaction with care received as an Emergency Room (ER) patient. Multiple regression analysis, optimized by UniODA to maximize predictive accuracy, was used to separately evaluate the ability of ratings of technicians (*n*=535), nurses (*n*=1,800) and doctors (*n*=1,806) seen in the ER to predict overall satisfaction. Results showed that the classification accuracy, as assessed using ESS, which was achieved by optimized MRA models of overall satisfaction was *moderate* for ratings of technicians (39.6) and nurses (42.0), and *relatively strong* for ratings of doctors (50.9). Illustrations show how the method works.

In an effort to better understand patient ratings of overall satisfaction with care they received in an Emergency Room (ER) visit, this study uses the manufacturing quality control methodology of evaluating how to maximize quality (patient satisfaction) at every separate step (every stage of care delivery) in the production process (in a patient's journey through the ER). This study extends recent research[1] investigating influence of the registration process on overall satisfaction ratings, by additionally examining the influence of patient interactions involving nurses, doctors, and technicians seen during the ER visit.

Method, Results, and Discussion

Data were obtained from patients who received care in the ER of a private Midwestern hospital, and were then mailed, completed, and returned a satisfaction survey. The survey obtained ratings of the *technicians*, *nurses*, and *doctors* seen by the patient during the ER visit. Ratings on all study variables were made using categorical ordinal Likert-type scales: 1=very poor, 2=poor, 3=fair, 4=good, and 5=very good.

Separately for each of the three different types of service providers, the *decision thresh-*

Optimal Data Analysis
2013, Vol. 2, Release 1 (October 21, 2013), 76-85

2155-0182/10/$3.00

olds (or "cut-points") used to predict a patient's overall satisfaction score (based on *Y**, computed using a multiple regression analysis or MRA-based model), were individually optimized via UniODA to explicitly maximize predictive accuracy.[2] There were no missing data because MRA uses case-wise deletion: observations with missing values on any variables in the model are omitted from analysis.[3] The outcome (dependent measure) in all three analyses was overall patient satisfaction with care received in the ER. Independent variables available for analysis for *technicians* (*n*=535) were phlebotomist skill, the courtesy of X-Ray technician and phlebotomist, and X-Ray waiting time. Independent variables available for *nurses* (*n*=1,800) were courtesy, if the patient's problem was taken seriously, attention paid to the patient, concern to keep the patient informed about the ER visit, technical skill, and concern for privacy. And, independent variables available for *doctors* (*n*=1,806) were wait time, courtesy, if the patient's problem was taken seriously, concern for comfort, explanation of test/treatment and of illness/injury, and home self-care advice. Descriptive statistics for all study measures are presented in Table 1. As seen, all means exceeded scale midpoint value of 3 due to negatively skewed distributions, and modest levels of variability also indicate relatively homogeneous responding.

Table 1: Descriptive Statistics: All Study Measures

Technicians (*n*=535)	Mean	SD	Median	Skewness	Kurtosis	CV
Overall satisfaction	4.14	1.15	5	-1.37	0.99	27.9
How well blood was taken	4.21	1.10	5	-1.60	1.96	26.2
Courtesy of person taking blood	4.36	0.94	5	-1.88	3.75	21.5
Waiting time in X-Ray	4.21	1.06	5	-1.49	1.70	25.2
Courtesy of X-Ray technologist	4.43	0.89	5	-1.92	4.01	20.0
Nurses (*n*=1,800)						
Overall satisfaction	4.16	1.08	4	-1.40	1.38	25.9
Courtesy	4.37	0.87	5	-1.78	3.70	19.9
Took your problem seriously	4.31	0.94	5	-1.63	2.69	21.8
Attention paid to you	4.11	1.02	4	-1.20	1.07	24.9
Concern to keep you informed	4.04	1.09	4	-1.11	0.61	27.1
Concern for your privacy	4.12	1.01	4	-1.24	1.26	24.5
Technical skill	4.33	0.88	5	-1.70	3.50	20.4
Doctors (*n*=1,806)						
Overall satisfaction	4.16	1.07	4	-1.38	1.34	25.9
Wait time to see doctor	3.74	1.21	4	-0.80	-0.23	32.3
Courtesy	4.43	0.84	5	-1.85	3.92	19.0
Took your problem seriously	4.39	0.91	5	-1.87	3.65	20.8
Concern for your comfort	4.30	0.93	5	-1.59	2.60	21.7
Explanation of test and treatment	4.30	0.98	5	-1.48	1.90	22.9
Explanation of illness or injury	4.18	1.02	4	-1.34	1.36	24.4
Advice about self-care	4.22	1.00	5	-1.43	1.77	23.8

Note: Ratings obtained using 5-point Likert-type scales: 1=very poor; 5=very good. SD=standard deviation; CV=coefficient of variation. Distributional moments are provided as data for future meta-analytic research.

2155-0182/10/$3.00

Before conducting planned multivariable analysis, UniODA[4] was first employed to obtain a basic understanding of the ability of ratings of care-providers to predict overall satisfaction.

Table 2 summarizes findings of analyses assessing the ability of the rated technician attributes to predict overall satisfaction ratings. All four technician ratings were statistically reliable predictors of overall satisfaction, and achieved moderate accuracy in training (total sample) analysis. However, with the exception of blood-taking skill, the strength of this relationship decreased in jackknife validity analysis, suggesting the effect may decline if the model is used to classify an independent random sample.[3] This is interesting because unlike the doctors—to whom patients are assigned and for whom familiarity and trust are maximized (and all ratings were stable in jackknife analysis), or the nurses—assigned to doctors and also scoring high in familiarity (most ratings were jackknife stable), neither phlebotomists nor X-Ray technicians are assigned to, or routinely seen, by ER patients.

Table 2: UniODA Accuracy Predicting Overall Satisfaction: Technicians

Attribute	**Predicted Rating**	***n***	**Accuracy (%)**	***p*<**	**ESS**
Phlobotomist skill	1	31	51.6	0.001	32.7
	2	46	19.6		
	3	16	50.0		
	4	160	50.6		
	5	282	75.5		
Phlobotomist courtesy	1	18	77.8	0.001	32.1
	2	8	25.0	0.001	30.6
	3	40	37.5		
	4	165	52.7		
	5	304	77.0		
X-Ray wait time	1	23	60.9	0.001	26.7
	2	56	14.3	0.001	19.6
	3	21	9.5		
	4	155	49.0		
	5	280	77.1		
X-Ray technician courtesy	1	13	92.3	0.001	27.6
	2	8	12.5	0.001	26.9
	3	42	28.6		
	4	147	55.1		
	5	325	73.8		

Note: The UniODA models that produced the predicted scores were non-directional, and tested the 2-tailed exploratory hypotheis that ratings of the attribute are related to ratings of overall satisfaction.[4] *n* is the number of observations having the indicated predicted rating, and Accuracy, also known as *predictive value*[5], is the percent of time that the predicted and actual ratings were identical for the indicated *n*. The Type I error rate (*p*) was estimated using 1,000 Monte Carlo experiments, yielding greater than 99.9% confidence for $p<$ 0.01. This level of Type I error rate was selected to ensure an experimentwise Type I error rate of $p<0.05$ within *each* of the three sets of analysis herein (nurses, labs, doctors).[4] If the classification accuracy of the model declined in “leave-one-out” or LOO (jackknife) validity analysis, LOO accuracy (ESS) and Type I error rate is reported beneath the results for the training (full sample) analysis.

2155-0182/10/$3.00

Table 3 summarizes findings of parallel analyses assessing the ability of the rated nurse attributes to predict overall satisfaction ratings, and Table 4 summarizes findings of rated doctor attributes to predict overall satisfaction. Lowest jackknife validity occurred for ratings of X-Ray wait time: this is not surprising, as wait times for procedures, and to be seen by the doctor, are known to be unreliable in the ER due to the need to triage real-time cases.[6]

Table 3: UniODA Accuracy Predicting Overall Satisfaction: Nurses

Attribute	Predicted Rating	*n*	Accuracy (%)	*p*<	ESS
Courtesy	1	42	71.4	0.001	28.7
	2	25	20.0		
	3	144	39.6		
	4	599	53.8		
	5	990	75.4		
Took your problem seriously	1	50	76.0	0.001	34.2
	2	45	28.9		
	3	168	37.5		
	4	578	56.1		
	5	959	77.2		
Attention paid to you	1	60	65.0	0.001	38.2
	2	78	25.6		
	3	263	33.5		
	4	608	54.8		
	5	791	84.1		
Kept you informed	1	77	55.8	0.001	36.4
	2	98	22.4		
	3	286	27.6		
	4	555	51.7		
	5	784	82.8		
Concern for your privacy	1	62	58.1	0.001	32.8
	2	61	23.0		
	3	270	28.2		
	4	612	51.8		
	5	795	82.0		
Technical skill	1	48	62.5	0.001	28.5
	2	17	29.4	0.001	27.1
	3	164	33.5		
	4	640	52.7		
	5	931	78.0		

Note: See Note to Table 2.

2155-0182/10/$3.00

Table 4: UniODA Accuracy Predicting Overall Satisfaction: Doctors

Attribute	**Predicted Rating**	***n***	**Accuracy (%)**	***p*<**	**ESS**
Wait time	1	136	34.6	0.001	30.1
	2	152	12.5		
	3	341	20.5		
	4	595	43.4		
	5	582	86.2		
Courtesy	1	31	77.4	0.001	27.9
	2	122	23.8		
	3	36	8.3		
	4	545	55.8		
	5	1072	74.1		
Took your problem seriously	1	48	70.8	0.001	33.3
	2	41	29.3		
	3	127	44.9		
	4	528	59.5		
	5	1062	75.9		
Concern for your comfort	1	47	76.6	0.001	34.4
	2	173	19.6		
	3	46	32.6		
	4	599	55.1		
	5	941	78.8		
Test/treatment information	1	47	76.6	0.001	34.6
	2	76	26.3		
	3	177	31.6		
	4	568	56.0		
	5	938	79.2		
Illness/injury information	1	57	61.4	0.001	33.8
	2	79	20.2		
	3	224	32.6		
	4	562	54.8		
	5	884	80.1		
Self-care information	1	61	59.0	0.001	33.9
	2	220	15.4		
	3	57	33.3		
	4	560	56.4		
	5	908	80.3		

Note: See Note to Table 2.

2155-0182/10/$3.00

The next step of the analysis involved obtaining *Y** for each of the three different types of service providers. For *technicians* the final MRA model was: *Y**=0.092+0.26*(phlebotomist skill)+0.37*(phlebotomist courtesy)+0.14*(X-Ray waiting time) + 0.17*(X-Ray technician courtesy). This model was statistically significant [$F(4,530)=127.1$, $p<0.0001$, $R^2=0.49$], and all independent variables made a statistically significant independent contribution to overall R^2 (p's< 0.03). For *nurses* the final MRA model was: *Y**=0.45+0.35*(took problem seriously)+ 0.23*(attention paid to patient)+0.20*(kept patient informed)+0.11*(technical skill). This model was statistically significant [$F(4,1795)=589.1$, $p<0.0001$, $R^2=0.49$], and all independent variables made a statistically significant independent contribution to overall R^2 (p's< 0.002). Finally, for *doctors* the final MRA model was: *Y**= -0.23 +0.25*(wait time) + 0.24*(took problem seriously)+0.21*(concern for comfort)+0.12 *(test/treatment explanation)+0.23*(home self-care advice). The model was statistically significant [$F(5,1800)=789.2$, $p<0.0001$, $R^2=0.69$], and all the independent variables made a statistically significant independent contribution to overall R^2 (p's< 0.0001).

Table 5 presents the cutpoints used on *Y** to obtain predicted ratings for all five values of the dependent measure. The first set of cutpoints provided is appropriate for general MRA, and is used for all MRA classification decisions based on *Y** for 5-point Likert-type scales. Also shown are cutpoints for optimized MRA models for all three service types, explicitly optimized for each sample via UniODA. For example, for *Y**=3.02 the dependent measure is predicted to be 3 by the general and optimized doctor MRA models, and is predicted to be 2 by the optimized nurse and technician MRA models.

Table 5: Model *Y** Cutpoints used to Obtain Predicted Overall Satisfaction Rating

Predicted Rating	General MRA	Optimized MRA: Nurses	Optimized MRA: Technicians	Optimized MRA: Doctors
1	$Y^*<1.5$	$Y^*<2.48$	$Y^*<2.65$	$Y^*<2.02$
2	$1.5 \le Y^*<2.5$	$2.48 \le Y^*<3.34$	$2.65 \le Y^*<3.39$	$2.02 \le Y^*<2.71$
3	$2.5 \le Y^*<3.5$	$3.34 \le Y^*<3.77$	$3.39 \le Y^*<3.68$	$2.71 \le Y^*<3.73$
4	$3.5 \le Y^*<4.5$	$3.77 \le Y^*<4.66$	$3.68 \le Y^*<4.49$	$3.73 \le Y^*<4.45$
5	$Y^*\ge 4.5$	$Y^*\ge 4.66$	$Y^*\ge 4.49$	$Y^*\ge 4.45$

Seen in Figure 1, an illustration of these four sets of cutpoints aides in conceptual clarity concerning how this methodology influences the model accuracy. Rather than use a "one-set-fits-all" template to predict overall satisfaction ratings on the basis of *Y** as is done in the general MRA approach, ODA optimizes the set of cutpoints to explicitly yield maximum classification accuracy for the sample. The optimized doctor model is closest to the general MRA template, sacrificing some of the *Y** domain used by the general model to predict ratings of 2 (see the thinner yellow band), in order to use more of the *Y** domain to predict ratings of 1 (wider green band). In contrast, the optimized technician and general MRA models are least similar, with the technician model devoting a small portion of the *Y** domain for predicting neutral ratings of 3.

2155-0182/10/$3.00

Figure 1

Model-Specific Cutpoint-Based ***Y** Domains** Yielding Predicted Overall Satisfaction Ratings of *1* (Green), *2* (Yellow), *3* (Blue), *4* (Red), and *5* (Black)

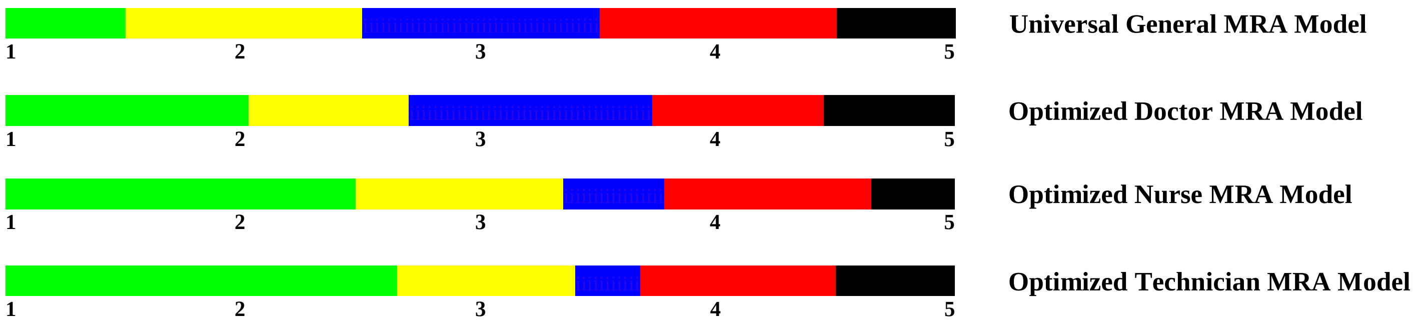

Table 6: Predicting Actual Overall Satisfaction Ratings by General versus Optimized MRA

Actual	Nurses				Technicians				Doctors			
Rating	General MRA		Optimized MRA		General MRA		Optimized MRA		General MRA		Optimized MRA	
1	28/81	34.6%	51/81	63.0%	12/29	41.4%	17/29	58.6%	32/79	40.5%	47/79	59.5%
2	15/83	18.1%	39/83	47.0%	3/33	9.1%	13/33	39.4%	26/82	31.7%	34/82	41.5%
3	81/192	42.2%	45/192	23.4%	17/50	34.0%	7/50	23.4%	109/203	53.7%	122/203	60.1%
4	377/559	67.4%	350/559	62.6%	108/145	74.5%	101/145	69.7%	382/555	68.8%	330/555	59.5%
5	680/885	76.8%	636/885	71.9%	213/278	76.6%	213/278	76.6%	715/887	80.6%	736/887	83.0%
ESS	34.6		42.0		33.9		39.6		43.8		50.9	
LOO			34.7				35.0					

Note: Tabled for each *Actual* (overall satisfaction) *Rating* is the number of correctly predicted ratings (numerator); the total number of times each rating was used in the sample (denominator); and the percentage accuracy or *sensitivity* obtained in classifying each rating category. Greatest sensitivity obtained for each category is indicated in red. LOO, or leave-one-out (jackknife) validity (unavailable for this problem with general MRA), is an estimate of prospective cross-generalizability of the model.[4] The LOO analysis for the optimized doctors model was abandoned after failing to solve in two CPU days, running UniODA[4] software on a 3 GHz Intel Pentium D microcomputer.

2155-0182/10/$3.00

A trifecta of explanations concerning why general MRA fails to explicitly achieve the maximum possible ESS for a sample of data comes to light. *Theoretically*, general MRA is formulated to maximize the proportion of variance in overall satisfaction ratings explained by a linear model, and the proportion of variance explained is not classification accuracy—which *is* the "objective function" that all ODA models explicitly maximize.[3] *Arithmetically*, attributable to the combination of a distributional skew and the use of discrete measurement levels—both of which constitute violations of assumptions underlying MRA[3], the general MRA template is overly conservative for estimates of the overall satisfaction ratings used by a minority of the sample. This characteristic of correlation was demonstrated experimentally, and further simulation research is needed to understand the factors creating suboptimal predicted ratings by general MRA.[2] *Paradoxically*, it is unlikely that the restrictive assumption underlying all linear models—that all independent variables apply equally to all the observations in the sample, are satisfied in the present research.[7] Indeed, there is evidence that use of a general linear model induces *Simpson's Paradox* presently, because ESS achieved *using a single rating*—amount of attention paid to the patient—to predict overall satisfaction via UniODA (38.2, in both training and LOO analysis) was greater than ESS for the general MRA model (34.6 in training, LOO not available) *involving four ratings*—one of which was amount of attention paid to the patient.[8,9]

Table 6 summarizes classification results obtained using general versus optimized MRA models to predict patient overall satisfaction ratings. Shown for general and optimized models, for all three service areas, and for all five overall satisfaction ratings, is the total number of correct predictions of the given rating divided by total number of instances of the given rating in the sample, and the associated sensitivity (accuracy) of the MRA model for each actual rating.

Considered across all three service areas, general MRA always achieved greater accuracy than optimized MRA when predicting responses of 3 (fair), and optimized MRA always achieved greatest accuracy in predicting the dissatisfied responses of 1 (very poor), or 2 (poor). Overall, accuracy yielded by all training models reflected moderate effects except for the optimized doctor MRA model, which met the minimal criterion for a relatively strong effect (ESS≥50).[4] Nevertheless, ESS of the optimized model was greater than ESS of the general model by 21.4% for the nurse model, 16.8% for the technician model, and 16.2% for the doctor model.

Table 7 presents confusion tables for the standard and optimized MRA models, which clearly show the dominant negative skew: both actual and predicted overall satisfaction ratings are primarily either 4 or 5.

An enhanced conceptual understanding of the findings is achieved by examination of *aggregated confusion tables* for the models, that consist of two columns and rows if the rating scale has an *even* number of rating categories, and three columns and rows for an *odd* number of rating categories—in which case the midpoint rating (e.g., 2 on a 3-point scale; 3 on a 5-point scale; 4 on a 7-point scale, etc.) constitutes the second row and second column (Table 7). For rating scales having an even number of response options the midpoint falls between two ratings, so the middle value is omitted from the aggregated table. To standardize the aggregated confusion table across scale range, entries *lower* than midpoint are summed and entered on the left-hand side of the aggregated confusion table; entries *higher* than midpoint are summed and entered on the right-hand side; and entries equal to the midpoint are ignored. For the nurse model, for example, the intersection of actual and predicted ratings of 1 or 2 for the optimized model in Table 7 is 51+16+15+39=121; and for the general model the corresponding value is 28+23+1+15= 67.

2155-0182/10/$3.00

Table 7: Confusion and Aggregated-Confusion Tables for Optimized and General MRA Models of Overall Satisfaction Ratings

	Nurses					Technicians					Doctors				
Actual	Predicted Rating					Predicted Rating					Predicted Rating				
Rating	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	51	16	4	5	5	17	3	2	4	3	47	16	12	4	0
	28	23	19	6	5	12	4	5	5	3	32	27	16	4	0
2	15	39	12	17	0	3	13	1	13	3	5	34	29	13	1
	1	15	42	24	1	0	3	13	14	3	4	26	33	18	1
3	18	65	45	55	9	9	11	7	18	5	6	22	122	40	13
	3	15	81	81	12	0	5	17	23	5	2	12	109	68	12
4	4	45	76	350	84	0	7	12	101	25	2	4	115	330	104
	0	4	62	377	116	0	0	12	108	25	1	3	73	382	96
5	2	9	15	223	636	1	4	4	56	213	0	1	8	142	736
	2	0	13	190	680	0	0	6	59	213	0	0	5	167	715

	Nurses		Technicians		Doctors	
	1 or 2	4 or 5	1 or 2	4 or 5	1 or 2	4 or 5
1 or 2	121	27	36	23	102	18
	67	36	19	25	89	23
4 or 5	60	1293	12	395	7	1312
	6	1363	0	405	4	1360
	ESS=77.3	ESS=64.6	ESS=58.1	ESS=43.2	ESS=84.5	ESS=79.2

Note: Red entries are for the optimized model, black entries are for the general model. For ESS of the 5x5 models see Table 6.

2155-0182/10/$3.00

Table 7 clearly reveals that performance of the standard and optimized MRA models was comparable for patients predicted to score 4 or 5 (satisfied) for all service areas. In contrast, the optimized nurse (181 versus 73) and technician (48 versus 19) models were much more likely to predict scores of 1 or 2 (dissatisfied), although the difference was marginal for the optimized doctor model (109 versus 93)—which was the most similar to the general model (Figure 1). If the phenomena assessed by items in the scale were actionable in real-time, and if a potent intervention existed to address such phenomena in real-time, then the optimized nurse and technician MRA models could be used to prevent many more cases of dissatisfaction, as compared with the standard model.

A total of 1,000 Monte Carlo experiments were used to estimate Type I error for all analyses, which were completed in 4.98 CPU hours running UniODA[4] on a 3 GHz Intel Pentium D microcomputer.

References

[1]Yarnold PR (2013). Maximum-accuracy multiple regression analysis: Influence of registration on overall satisfaction ratings of Emergency Room patients. *Optimal Data Analysis, 2*, 72-74.

[2]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.

[3]Licht MH (1995). Multiple regression and correlation. In: Grimm LG, Yarnold PR (Eds.), *Reading and Understanding Multivariate Statistics*. Washington, DC: APA Books, 1995, pp. 19-64.

[4]Yarnold PR, Soltysik RC. *Optimal data analysis: a guidebook with software for Windows*. APA Books, Washington, DC, 2005.

[5]Ostrander R, Weinfurt KP, Yarnold PR, August G (1998). Diagnosing attention deficit disorders using the BASC and the CBCL: Test and construct validity analyses using optimal discriminant classification trees. *Journal of Consulting and Clinical Psychology, 66*, 660-672.

[6]Tanabe P, Gimbel R, Yarnold PR, Kyriacou D, Adams J (2004). Reliability and validity of scores on Emergency Severity Index Version 3. *Academic Emergency Medicine*, *11*, 59-65.

[7]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.

[8]Yarnold PR (1996). Characterizing and circumventing Simpson's paradox for ordered bivariate data. *Educational and Psychological Measurement*, *56*, 430-442.

[9]Soltysik RC, Yarnold PR (2010). The use of unconfounded climatic data improves atmospheric prediction. *Optimal Data Analysis*, *1*, 67-100.

Author Notes

E-mail: Journal@OptimalDataAnalysis.com.

Mail: Optimal Data Analysis, LLC
1220 Rosecrans St., #330
San Diego, CA 92106

ODA Blog: http://odajournal.wordpress.com/