Optimal Data Analysis
Vol. 2, Release 2 (October 22, 2013), 86-93

2155-0182/13/$3.00

Minimum Standards for Reporting UniODA Findings for Class Variables with Three or More Response Categories

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

An incontrovertible advantage of establishing a minimum set of standards for reporting findings obtained using *any method*—is researchers from *all fields* will be able to easily and clearly understand fundamental statistical results of *any study* reporting findings using that method. This note extends minimum standards proposed for reporting of UniODA findings with binary class variables[1] to applications involving class variables with more than two response categories, and testing confirmatory or exploratory hypotheses.

Exposition begins with a *confirmatory* or one-tailed example, for which the data are taken from a survey-based study of patient satisfaction with care received in an Emergency Department (ED).[2] The class variable is patient-rated likelihood of recommending the ED to others (REC), and the attribute is rating of overall satisfaction with ED care (SAT). Both variables are assessed using a categorical ordinal 5-point Likert-type scale: 1=very poor, 2=poor, 3=fair, 4=good, 5= very good. The *a priori* hypothesis tested is that satisfaction ratings are positively predictive of recommendation: the DIR command specifies a linear relationship.[3] Sufficient Monte Carlo experiments were run, for example, to yield near 100% confidence for desired ("target") $p<0.01$, and a leave-one-out (LOO) jackknife validity analysis was conducted. The UniODA code[3] used to run this analysis (control commands are indicated in red) was:

```
VARS REC SAT;
CLASS REC;
ATTR SAT;
DIR < 1 2 3 4 5;
MCARLO ITER 1000;
LOO;
GO;
```

Exposition ends with an *exploratory* or two-tailed example, using data taken from the same source as the confirmatory example. The class variable is a categorical indicator of who accompanied the patient to the ED, with three mutually-exclusive dummy-coded levels: 2 was used to indicate the patient arrived with a friend; 3 indicated the patient arrived with an employer; 4 indicated the patient arrived alone (WHO). The attribute is a 5-point Likert-type patient rating of courtesy of ED staff toward the patient,

2155-0182/13/$3.00

family, and friends: 1=very poor, 2=poor, 3= fair, 4=good, 5=very good (COURTESY). The *post hoc* hypothesis is that patients who arrive to the ED alone, with a friend, or with an employer give different recommendations for the ED (the DIR command is not used in *post hoc* analyses). Sufficient Monte Carlo experiments were run, for example, to yield near 100% confidence for desired $p<0.05$, and a leave-one-out (LOO) validity analysis was performed. UniODA code used to conduct this analysis (control commands indicated in red) was:

```
VARS WHO COURTESY;
CLASS WHO;
ATTR COURTESY;
MCARLO ITER 5000;
LOO;
GO;
```

Analyses were completed for both of the examples in 78 CPU seconds running UniODA[3] on a 3 GHz Intel Pentium D microcomputer.

Results and Discussion

Table 1 gives descriptive statistics for ordered measures (for the exploratory study: 59 patients arrived at the ED with friends; 369 with employers; and 86 patients arrived alone). All of the means exceeded the scale midpoint value of 3 due to negative skew, and the modest levels of variability indicate homogeneous responding.

Table 1: Descriptive Statistics for Ordered Study Measures

Confirmatory Example (*n*=2070)	Mean	SD	Median	Skewness	Kurtosis	CV
Overall satisfaction	4.10	1.15	4	-1.34	1.02	28.0
Likelihood of recommending ED	4.03	1.24	4	-1.27	0.59	30.7
Exploratory Example (*n*=514)						
Courteous treatment of patient/friends	4.01	1.08	4	-1.11	0.64	27.0

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.

Predictive Value. As seen in Table 2, minimal sufficient information that is required to understand the *predictive value* of a model for a given attribute includes the UniODA model; number of observations and percent of correctly classified observations in every predicted class category; and model Type I error (p) and ESS in *training analysis* conducted for the total sample, and if run also in leave-one-out (LOO) jackknife *validity analysis*.[3] In Table 2 the classification results are correctly indicated as being stable in LOO analysis: that is, the same as were obtained in the training analysis. Hypothetically, if LOO performance had been lower than training performance, then this would have been indicated as illustrated in Table 3, in which p and ESS for LOO analysis are given beneath corresponding values obtained in training analysis.

To illustrate where information in Tables 2 and 3 is found within UniODA output, Table 4 reproduces applicable output generated by the program code given earlier for the *confirmatory* example. The UniODA model is indicated in the output as *ODA model*. Note that variable codes were converted to text in Table 2 (generally this should be done).

2155-0182/13/$3.00

Table 2

Critical Information for UniODA Analyses: *Predictive Value*

Attribute	UniODA Model	*n*	% Correct	*p* <	ESS
Overall satisfaction with ED care	If SAT = very poor, predict REC = very poor	122	91.0	0.01	59.9
	If SAT = poor, predict REC = poor	107	42.1		
	If SAT = poor, predict REC = poor	213	62.0		
	If SAT = poor, predict REC = poor	629	72.0		
	If SAT = poor, predict REC = poor	999	90.3		
Courtesy shown to patient/friends in ED	If WHO = employer, predict COURTESY = fair, poor, or very poor	180	12.8	0.05	12.4
	If WHO = friend, predict COURTESY = good	126	77.8		
	If WHO = alone, predict COURTESY = very good	208	24.5		

Note: *p* is desired (target) Type I error rate, and ESS (effect strength for sensitivity) is a normed index of effect strength on which 0 represents the classification accuracy expected by chance, and 100 represents errorless classification.[3] All attributes were stable in LOO validity analysis.[3]

Table 3

Modification to Table 1 Based on Hypothetical LOO-Instability of Ordered Attribute

Attribute	UniODA Model	*n*	% Correct	*p* <	ESS
Overall satisfaction with ED care	If SAT = very poor, predict REC = very poor	122	91.0	0.01 0.02	59.9 57.3

Note: Model *p* and ESS values for training analysis are provided in the first (top) row, and if LOO-unstable, then *p* and ESS values for LOO validity analysis are given in the second (bottom) row (hypothetical results illustrated).

2155-0182/13/$3.00

The next column in Table 2 is *n*, which is located just beneath the output Classification Performance Summary table (Table 4). In Table 2 a total of 122 patients were predicted by the model to have given a "very poor" rating, and this is seen at the bottom of the "Predicted 1" column, on the left-hand-side of the output Classification Performance Summary table. In Table 2, a total of 999 patients are predicted to have given a "very good" rating, as seen at the bottom of the "Predicted 5" column on the right-hand-side of the output summary table.

Table 4

Selected UniODA Output for Confirmatory Example Analysis (Ordered Attribute)

```
ODA model:
----------

IF SAT <= 1.5 THEN REC = 1
IF 1.5 < SAT <= 2.5 THEN REC = 2
IF 2.5 < SAT <= 3.5 THEN REC = 3
IF 3.5 < SAT <= 4.5 THEN REC = 4
IF 4.5 < SAT THEN REC = 5


Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations               Estimated p
----------               -----------
   1000                    .000000

Confidence levels for estimated p:
----------------------------------
Desired p   Confidence      Desired p   Confidence
---------   ----------      ---------   ----------
 p<.001       63.27%         p>.001       36.73%
 p<.01       100.00%         p>.01         0.00%
 p<.05       100.00%         p>.05         0.00%
 p<.10       100.00%         p>.10         0.00%


Classification performance summary:
-----------------------------------    Overall      Mean Sens
  Correct   Incorrect                 accuracy   across classes
    1643       427                    79.37%       67.88%

Class                  Predicted
 REC       1       2       3       4       5
       ---------------------------------------- NA      Sens
       |       |       |       |       |       |
     1 |  111  |  41   |  19   |   3   |   2   | 176    63.07%
A      |       |       |       |       |       |
c    2 |   8   |  45   |  29   |   7   |   0   | 89     50.56%
t      |       |       |       |       |       |
u    3 |   2   |  19   |  132  |  73   |   9   | 235    56.17%
a      |       |       |       |       |       |
l    4 |   0   |   2   |  30   |  453  |  86   | 571    79.33%
       |       |       |       |       |       |
     5 |   1   |   0   |   3   |  93   |  902  | 999    90.29%
       |       |       |       |       |       |
       ----------------------------------------
   NP     122     107     213     629     999
   PV    90.98%  42.06%  61.97%  72.02%  90.29%  Mean PV  71.46%

Effect strength Sens  59.86%      Effect strength PV  64.33%
```

2155-0182/13/$3.00

The next column in Table 2 is % Correct—the percent of correct model predictions for a given target rating, which is located below *n* in the output classification performance summary table. In Table 4, beneath the *n* for predicted class 1 (122), the percent of the predicted class 1 observations who in reality did use a rating of 1 (as the UniODA model predicted) is given for the column, as 90.98%. And, beneath the *n* for predicted class 5 (999), the percent of the predicted class 5 observations who in reality did use a rating of 5 (as predicted by the UniODA model) is given for the column, as 90.29%.

The next-to-final column in Table 2 is *p*, or Type I error, estimated using the MC command.[3] As seen, in this example estimated $p < 0.0001$, but simulation based on 1,000 Monte Carlo experiments indicates a strong level of confidence for $p<0.05$. To obtain acceptable confidence for a lower desired (target) *p* value requires running more Monte Carlo experiments. For example, 5,000 experiments used here yielded 99.3% confidence for $p<0.001$.

Table 5

Selected UniODA Output for Exploratory Example Analysis (Categorical Attribute)

```
ODA model:
----------

IF COURTESY <= 3.5 THEN WHO = 3
IF 3.5 < COURTESY <= 4.5 THEN WHO = 2
IF 4.5 < COURTESY THEN WHO = 4

Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations                Estimated p
----------                -----------
   5000                     .015800

Confidence levels for estimated p:
----------------------------------
Desired p   Confidence      Desired p   Confidence
---------   ----------      ---------   ----------
 p<.001        0.00%         p>.001      100.00%
 p<.01         0.01%         p>.01        99.99%
 p<.05       100.00%         p>.05         0.00%
 p<.10       100.00%         p>.10         0.00%

 Classification performance summary:
-----------------------------------    Overall      Mean Sens
  Correct   Incorrect                 accuracy   across classes
    172        342                     33.46%       41.61%

Class          Predicted
 WHO       2       3       4
       ------------------------- NA      Sens
A      |       |       |       |
c     2|  23   |  13   |  23   | 59     38.98%
t      |       |       |       |
u     3|  137  |  98   |  134  | 369    26.56%
a      |       |       |       |
l     4|  20   |  15   |  51   | 86     59.30%
       |       |       |       |
       -------------------------
   NP     180     126     208
   PV    12.78%  77.78%  24.52%  Mean PV  38.36%

Effect strength Sens  12.42%      Effect strength PV   7.54%
```

2155-0182/13/$3.00

The final column in Table 2 gives model ESS, which is provided just beneath and on the left-hand-side of the output performance summary table. Presently this is 59.86% in Table 4, and rounded scientifically to one decimal place as 59.9% in Table 2. ESS achieved in LOO analysis was the same as obtained in training analysis: this is called a *LOO-stable model*. If ESS obtained in LOO analysis is lower than is achieved in training analysis, the model is called *LOO-unstable* (see Table 3). ESS by LOO analysis is a more conservative estimate of prospective validity than training ESS, and it reflects an upper bound for the expected hold-out validity.[3]

Table 5 reproduces the applicable output generated by the program code given earlier for the *exploratory* example (variable codes have been converted to text in Table 2). The UniODA model is indicated in the output as *ODA model*. As in the prior example, *n* is located beneath the output Classification Performance Summary table (Table 5). In Table 2 a total of 180 patients were predicted by the model to have arrived at ED with a friend, and this is seen at the bottom of the "Predicted 2" column, left-hand-side of the output summary table. In Table 2, a total of 208 patients were predicted to have arrived at ED alone, seen at the bottom of the "Predicted 4" column, right-hand-side of output summary. Similarly, % Correct is located below *n* in the output summary table. In Table 5, beneath the *n* (180) for predicted class 2, the percent of the predicted class 2 observations who in reality did use a rating of "fair" or worse to describe the likelihood of recommending the ED to others (as the UniODA model predicted) is given for the column as 12.78%. Beneath *n* (208) for predicted class 4, the percent of the predicted class 4 observations who in reality did use a rating of "very good" to describe the likelihood of their recommending the ED to others (as the UniODA model predicted) is given for the column as 24.52%. The next-to-final column in Table 2 is *p*: here estimated $p<0.0158$, but simulation using 5,000 Monte Carlo experiments indicates strong confidence for $p<0.05$. The final column in Table 2 gives model ESS, which is provided just beneath and on the left-hand-side of the output performance summary table. Presently this is 12.42% in Table 4, and rounded scientifically to one decimal place as 12.4% in Table 2. ESS achieved in LOO analysis was the same as obtained in training analysis.

Sensitivity. The minimal sufficient information needed to understand the *sensitivity* of a model for any given category of a class variable is the total number of correct predictions of the category divided by total number of instances of the category in the sample (the UniODA model, ESS, and *p* are given in Table 2). Table 6 shows this information for the confirmatory example, and Table 7 for the exploratory example.

Table 6: Predicting Actual Ratings of Likelihood to Recommend the ED to Others (Confirmatory Example)

Actual Rating	Predicted Rating: Very Poor (1)		Poor (2)		Fair (3)		Good (4)		Very Good (5)	
1	28/81	34.6%	51/81	63.0%	12/29	41.4%	17/29	58.6%	32/79	40.5%
2	15/83	18.1%	39/83	47.0%	3/33	9.1%	13/33	39.4%	26/82	31.7%
3	81/192	42.2%	45/192	23.4%	17/50	34.0%	7/50	23.4%	109/203	53.7%
4	377/559	67.4%	350/559	62.6%	108/145	74.5%	101/145	69.7%	382/555	68.8%
5	680/885	76.8%	636/885	71.9%	213/278	76.6%	213/278	76.6%	715/887	80.6%

Note: Tabled for each *Actual Class* category in red is the number of correctly predicted observations (the numerator); total number of times each class category was present in the sample (the denominator); and the percentage accuracy or *sensitivity* obtained for each class category. Correct predictions are attributable to the *a priori* hypothesis.

2155-0182/13/$3.00

Table 7: Predicting Actual Class Membership: Exploratory Example

Actual Class	Predicted Class 2 (Friend)		3 (Employer)		4 (Alone)	
2	28/81	34.6%	51/81	63.0%	12/29	41.4%
3	15/83	18.1%	39/83	47.0%	3/33	9.1%
4	680/885	76.8%	636/885	71.9%	213/278	76.6%

Note: See Note to Table 6. A two-tailed hypothesis was tested in this application. Correct predictions made by the UniODA model are indicated using red.

Highest accuracy for the confirmatory model (Table 6) was achieved for prediction of actual class categories 4 and 5—good and very good likelihood of recommending ED to others. Best accuracy for the exploratory model (Table 7) was achieved for prediction of actual class category 4 (arrive at ED alone). Overall accuracy (ESS) of the confirmatory model represented a relatively strong effect (defined as ESS≥50), while ESS for the exploratory model represented a weak effect (ESS<25).[3]

Aggregated Confusion Table (ACT). A clear conceptual understanding of the findings can sometimes be facilitated by examination of ACTs for UniODA models involving an ordered attribute.[4] ACTs have 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: the midpoint rating (2 on a 3-point scale; 3 on a 5-point scale) forms the second (middle) row and column of the ACT, and represents a "neutral" or "undecided" response. For rating scales with an even number of response options the midpoint falls between two ratings, so there is no middle value in the ACT. To standardize ACTs 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 midpoint are ignored.

An ACT is given for the confirmatory example, that involves a class variable measured on an ordered scale, in Table 8. For example, the intersection of actual and predicted ratings of 1 or 2 (Table 4) is 111+41+8+45=205; and the intersection of actual and predicted ratings of 4 or 5 is 453+ 86+93+902=1534 (Table 8).

Table 8: Aggregated Confusion Table for the Confirmatory Example

Actual Rating	Predicted Rating 1 or 2	4 or 5
1 or 2	205	12
4 or 5	3	1534

For this ACT, ESS=94.3, a very strong effect (ESS≥75).[3] For all ratings except those reflecting ambivalence (midpoint ratings of 3), the *a priori* UniODA model was almost perfect in classifying people as being on one side of the midpoint, or the other.

References

[1]Yarnold PR (2013). Assessing technician, nurse and doctor ratings as predictors of overall satisfaction ratings of Emergency Room patients: A maximum-accuracy multiple regression analysis. *Optimal Data Analysis*, 2, 76-85.

[2]Yarnold PR, Michelson EA, Thompson DA, Adams SL (1998). Predicting patient satisfac-

2155-0182/13/$3.00

tion: A study of two emergency departments. *Journal of Behavioral Medicine, 21*: 545-563.

[3]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books.

[4]Yarnold PR (2013). Assessing technician, nurse, and doctor ratings as predictors of overall satisfaction ratings of Emergency Room patients: A maximum-accuracy multiple regression analysis. *Optimal Data Analysis,* 2, 76-85.

Author Notes

E-mail: Journal@OptimalDataAnalysis.com.

Mail: Optimal Data Analysis, LLC
1220 Rosecrans St., #330
San Diego, CA 9210

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