Optimal Data Analysis
Vol. 2, Release 2 (October 29, 2013), 106-119

2155-0182/10/$3.00

Standards for Reporting UniODA Findings Expanded to Include ESP and All Possible Aggregated Confusion Tables

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

UniODA models maximize Effect Strength for Sensitivity (ESS), a normed measure of classification accuracy (0=chance, 100=perfect classification) that indexes the models ability to accurately identify the members of different class categories in the sample. In a study discriminating genders, for example, percent of each gender accurately classified by the model is indexed using ESS. Unlike ESS, the Effect Strength for Predictive Value (ESP) varies across base-rate. Measured using the identical scale as ESS, ESP indexes the models ability to produce accurate classifications. In the study discriminating genders, for example, the percent of the time the model made an accurate prediction that an observation was either male or female is indexed using ESP. While ESS is important in helping to guide the development and testing of theory, ESP is important in translating theory from laboratory to real-world applications, and is thus added to the recommended minimum standards[1] for reporting of *all* UniODA findings. In addition, the evaluation of *all possible aggregated confusion tables* aids in interpreting UniODA findings, and evaluating the potential for increasing classification accuracy by improving measurement of ordered class variables and/or attributes, and so was also added as a recommended minimum standard. Current standards are demonstrated using three examples: (1) using income to discriminate gender in a sample of 416 general internal medicine (GIM) patients, testing the *a priori* hypothesis that men have higher income than women; (2) using body mass index (BMI) to discriminate income in a sample of 411 GIM patients, testing the *a priori* hypothesis that BMI and income are positively related; and (3) discriminating mental focus using GHA (a measure of barometric pressure) in a *post hoc* analysis of 297 sequential daily entries of a fibromyalgia patient using an intelligent health diary, that were separated into training and hold-out validity samples.

2155-0182/10/$3.00

Example 1

Discriminating Sex using Income: Confirmatory Analysis with Binary Class Variable and Ordered Attribute

Data were obtained from a convenience sample of *n*=416 adult ambulatory patients waiting to be seen in general internal medicine clinic at a private hospital in Chicago, Illinois. The binary class variable SEX indicated whether the patient was female (dummy-coded 2; *n*=324) or male (dummy-coded 1; *n*=92). INCOME was an ordered 7-point scale with 1 used to indicate up to $10,000 per year, 2 was used for ≤$20,000, 3 for ≤$30,000; 4 for ≤$40,000; 5 for ≤$50,000; 6 for ≤$60,000; and 7 was used to indicate more than $60,000 per year. This scale was developed when data input was accomplished by scanning "bubble forms." Annual income is preferred as a more accurate measure than the 7-point Likert-type scale that was used. Descriptive statistics for INCOME were as follows. For men: mean= 3.24; standard deviation (SD)=1.98; median=3; skewness=0.55; kurtosis=-0.68; and coefficient of variation (CV)=61.1. For women: mean= 2.88; standard deviation (SD)=1.64; median=2; skewness=0.76; kurtosis=0; and CV=56.9.

The first *confirmatory* analysis tested the *a priori* hypothesis that men have higher income than women by running the following UniODA[2] code (control commands indicated using red):

```
VARS SEX INCOME;
CLASS SEX;
ATTR INCOME;
DIR < 2 1;
MCARLO ITER 10000;
GO;
```

The DIRECTIONAL or DIR command specifies the *a priori* hypothesis that women (2) will have lower (<) INCOME than men (1).[2] A total of 10,000 Monte Carlo experiments were used to estimate *p*, which is the default setting used in the ODA laboratory for exploratory data analysis involving UniODA and a binary class variable.[3] The analysis was completed in 3 CPU seconds by UniODA[2] run on a 3 GHz Intel Pentium D microcomputer.

Findings support the *a priori* hypothesis that men have greater INCOME: the model met the generalized criterion[2] (i.e., per-comparison $p < 0.05$) for achieving statistical significance, and was stable in LOO analysis (Table 1). However, ESS and ESP were both very weak, calling the theoretical and pragmatic efficacy of the finding into question. The model performed weakly in classifying the men but strongly in classifying the women in the sample (28% versus 84% sensitivity, respectively): it was accurate 80% of the times a patient was predicted to be female, versus 33% of the times a patient was predicted to be male. Such results are efficiently presented using a confusion table, as seen in Table 2.

Table 2: Confusion Table for Confirmatory UniODA Model Discriminating Gender using INCOME for Total Sample

		Patient Predicted Status		
		Male	Female	
Patient Actual Status	Male	26	66	28.3%
	Female	53	271	83.6%
		32.9%	80.4%	

The aggregated confusion table, or *ACT*, was developed as a tool for enhancing conceptual understanding of UniODA models involving class variables with more than two response categories.[1,4,5] This procedure involves computing subsets of confusion tables including observations scoring successively further from the model decision threshold, thereby increasing the reliability of and discriminability between class categories. ACTs were defined based upon class variables, but extending this idea to applications with a binary class variable is straightforward.

2155-0182/10/$3.00

Table 1: UniODA Model *Confirmatory* Performance: Discriminating Gender using Income

Predictive Value

Income (UniODA Model)	Predicted Gender	*n*	% Correct	ESP	*p* <	Confidence
≤40	2 (Female)	337	80.4	13.3	0.05	99.999%
>40	1 (Male)	79	32.9			

Sensitivity

Actual Gender	*n*	Number Correctly Predicted	Sensitivity (% Accuracy)	ESS
1 (Male)	92	26	28.3	11.7
2 (Female)	324	271	83.6	

Note: UniODA results are provided for a directional (confirmatory, *a priori*) hypothesis specifying that men have greater INCOME than women (the value "40" signifies $40,000 per year); *n* for predictive value is number of times the model predicted an observation was a member of the indicated class category; % Correct is percent correct predictions of observations' actual class category; ESP is effect strength for predictive value, where 0=chance and 100=errorless prediction; *p* is the desired (target) Type I error rate (any target *p* may be used); Confidence for target *p* is based on 10,000 Monte Carlo experiments (estimated $p<0.0397$); *n* for sensitivity is the number of observations in the sample that are members of the indicated class category; Number Correctly Predicted is the number of observations in each indicated class category that were correctly classified by the UniODA model; Sensitivity is the percent accurate classification of each indicated class category for the sample; and ESS is effect strength for sensitivity (0=chance; 100=perfect classification).[2] Model performance was stable in LOO validity analysis.[2]

A conceptually related, commonly used method in personality research is to limit the sample to individuals having relatively extreme scores on a measured factor, thereby increasing the reliability of group designations based on the measured factor.[6-8] The analogue to the ACT performed on the attribute (not the class variable) in this case begins by dropping patients in the two *attribute* levels in the middle of the scale (<$40,000, <$50,000) from the sample, and then computing the resulting ACT.

This was done here and results are summarized in Table 3. As seen, findings were similar to findings for the total sample, although the model sensitivity and predictive value for classification and prediction of females both increased. Here ESS=13.3 and ESP=19.7, both still representing very weak effects. A total of

2155-0182/10/$3.00

288 patients were included in the ACT, 69.2% of the total sample.

Table 3: Aggregated Confusion Table for Confirmatory UniODA Model Discriminating Gender using INCOME for Total Sample *Excluding* Patients Scoring at <$40,000 and <$50,000

		Patient Predicted Status		
		Male	Female	
Patient Actual Status	Male	13	44	22.8%
	Female	22	209	90.5%
		37.1%	82.6%	

The next ACT increases the reliability of group membership further by dropping patients scoring at the remaining closest response levels to the midpoint: here, <$30,000 and <$60,000. As seen, there has been little change: ESS=9.5, ESP=22.2. A total of 222 patients were included in the ACT, 53.4% of the total sample.

Table 4: Aggregated Confusion Table for Confirmatory UniODA Model Discriminating Gender using INCOME for Total Sample, *Additionally Excluding* Patients Scoring at <$30,000 and <$60,000

		Patient Predicted Status		
		Male	Female	
Patient Actual Status	Male	7	39	15.2%
	Female	10	166	94.3%
		41.2%	81.0%	

The final ACT increases the reliability of group membership to the most extreme possible level, by including only the two most extreme response categories (<$10,000 and >$60,000), with results shown in Table 5. As seen, things became a little worse: ESS=3.6, ESP=15.0. A total of 105 patients were included in the ACT, 25.2% of the total sample.

Table 5: Final Aggregated Confusion Table for Confirmatory UniODA Model Discriminating Gender using INCOME for Total Sample, *Only Including* Patients Scoring at <$10,000 and >$60,000

		Patient Predicted Status		
		Male	Female	
Patient Actual Status	Male	2	25	7.4%
	Female	3	75	96.2%
		40.0%	75.0%	

Thus, although there is statistical support for the *a priori* hypotheses that men have a greater INCOME than women the effect is very weak, and there is no evidence that increasing precision of the INCOME measure will be able to improve ESS or ESP to a clinically meaningful level, because model performance actually began to degrade when the most extreme scores were used in analysis,

For expository purposes an exploratory analysis was conducted for these data that tested the *post hoc* hypothesis that men and women have different INCOME by commenting-out the DIR command and then rerunning the program:

```
*DIR < 2 1;
GO;
```

For this analysis confidence for target $p<0.10$ was 99.999% (estimated $p<0.0831$).

2155-0182/10/$3.00

Example 2

Predicting Income using Body Mass Index:
Confirmatory Analysis of Ordered
Class Variable and Attribute

While INCOME may not bear much of a relationship to gender, perhaps it can predict one of the pandemic consequences of opulence—obesity? This study tested the *a priori* hypothesis that INCOME is positively related to body mass index (BMI, measured in kg/m^2). Data were obtained from a convenience sample of n=411 adult ambulatory patients waiting to be seen in general internal medicine clinic at a private hospital in Chicago, Illinois.

INCOME was treated as an ordered class variable consisting of seven response levels (see Example 1): descriptive statistics were mean= 2.97; SD=1.69; median=3; skewness=.62; kurtosis=-0.50; CV = 61.1. BMI was treated as an ordered attribute measured on an interval scale: mean=28.6; SD=6.2; median=27.98; skewness= 1.25; kurtosis=3.37; CV=21.6. The confirmatory hypothesis was tested by running the following UniODA[2] code (control commands indicated using red):

```
VARS INCOME BMI;
CLASS INCOME;
ATTR BMI;
DIR < 1 2 3 4 5 6 7;
MCARLO ITER 500;
GO;
```

The DIR command specifies the *a priori* hypothesis that poorer patients will have lower (<) BMI than wealthier patients.[2] A total of 500 Monte Carlo experiments were used to estimate p because a test run with 100 experiments indicated p was statistically marginal, and that 500 experiments should render near maximum level of confidence for target p<0.10. LOO analysis was not performed because of significant computational effort required, considered in conjunction with the finding that training ESS and ESP values were very low, so that a possible finding of diminished jackknife performance would be redundant. Analysis was completed in 1.91 CPU hours by UniODA[2] run on a 3 GHz Intel Pentium D microcomputer.

Results offer marginal statistical support (p<0.10) for the *a priori* hypothesis that greater INCOME predicts greater BMI (see Table 6). However, as for the prior example ESS and ESP are both very weak. Model predictive values are in the moderate range for lower income patients having BMI<29.5 kg/m^2. Model sensitivities are comparable in magnitude to predictive values, but unlike predictive values, model sensitivity does not present an obvious pattern with respect to actual income level.

All possible ACTs were examined next. As is seen in Table 7, similar to the findings in Example 1—in which INCOME was used as an *attribute*, the results here using INCOME as a *class variable* offer little evidence that there is a relationship with BMI, and offer little hope that improving measurement precision will improve either ESS or ESP. Even in the most reliable of circumstances, ESS would be moderate at best, and ESP would remain very low.

Example 3

Predicting Mental Focus via GHA: Exploratory
Analysis of Ordered Class Variable and
Attribute in Case Series with
Hold-Out Sample

This study tested the *post hoc* hypothesis that mental focus (FOCUS)—one of the primary negative symptoms in fibromyalgia (FM), is related to atmospheric pressure. Data were abstracted with permission from a computer log containing 297 sequential entries made by an anonymous patient with FM using an intelligent health diary.[9,10]

2155-0182/10/$3.00

Table 6: UniODA Model Performance: Predicting Income using BMI

Predictive Value

UniODA Model						
BMI	Predicted Income	*n*	% Correct	ESP	*p* <	Confidence
≤22.6	1 (≤$10,000/yr)	61	42.6	11.3	0.10	99.83%
≤26.6	2 (≤$20,000/yr)	107	31.8			
≤27.4	3 (≤$30,000/yr)	25	32.0			
≤29.5	4 (≤$40,000/yr)	67	25.4			
≤32.1	5 (≤$50,000/yr)	51	19.6			
≤37.0	6 (≤$60,000/yr)	66	10.6			
>37.0	7 (>$60,000/yr)	34	5.9			

Sensitivity

Actual Income Level	*n*	Number Correctly Predicted	Sensitivity (% Accuracy)	ESS
1 (≤$10,000/yr)	98	26	26.5	11.7
2 (≤$20,000/yr)	102	34	33.3	
3 (≤$30,000/yr)	58	8	13.8	
4 (≤$40,000/yr)	74	17	23.0	
5 (≤$50,000/yr)	44	10	22.7	
6 (≤$60,000/yr)	18	7	38.9	
7 (>$60,000/yr)	17	2	11.8	

Note: See Note to Table 1. UniODA results are given for a directional hypothesis specifying a linear relationship between income measured on the ordinal scale in the Table and BMI (kg/m^2). Confidence for target *p* is based on 500 Monte Carlo experiments (estimated $p<0.064$). LOO analysis was not performed due to the significant computational effort required, considered in conjunction with the fact that training ESS and ESP values were very low, so the possible finding of diminished jackknife performance would be redundant.

2155-0182/10/$3.00

Table 7: All Possible Confusion Tables for Income and BMI Example Training Analysis (*n*=411)

Root Confusion Table (ESS=11.67, ESP=11.31)

Actual	Predicted Income						
Income	≤10	≤20	≤30	≤40	≤50	≤60	>60
≤10	26	16	5	17	11	14	9
≤20	11	34	4	13	14	14	12
≤30	6	14	8	7	5	15	3
≤40	7	22	4	17	9	9	6
≤50	8	12	1	6	10	5	2
≤60	2	4	0	4	1	7	0
>60	1	5	3	3	1	2	2

All Possible Aggregated Confusion Tables

Actual	Predicted Income ≤30	≥50
≤30	124	97
≥50	36	30

ESS=10.66
ESP=1.12

Actual	Predicted Income ≤20	≥60
≤20	87	49
≥60	12	11

ESS=11.80
ESP=6.21

Actual	Predicted Income ≤10	>60
≤10	26	9
>60	1	2

ESS=40.96
ESP=14.48

Note: ACT excluding midpoint category ≤40 classified *n*=287 (69.8% of sample); ACT excluding categories ≤30 to ≤50 classified *n*=159 (38.7% of sample); and ACT including categories ≤10 and >60 classified *n*=38 (9.2% of sample).

2155-0182/10/$3.00

The attribute was atmospheric pressure assessed as 500 mb GHA, measured using an interval scale (GHA).[11] The class variable was the patient's 10-point Likert-type rating of percent of maximum possible mental focus available in the prior 24 hours (FOCUS).[9]

Recent research demonstrates how to use such information in real-time to provide patients with FM prospective alerts about upcoming bad and good symptom periods.[11] As experience increases patients learn to interact with the dairy in an individually-tailored manner, and ODA-generated alerts consequently are increasingly sensitive as more patient data are obtained.[9,10] Classifications are most accurate and may use the fewest attributes when the class variable and antecedent attributes are stable over some period of time, as compared with situations when data change rapidly—sometimes more rapidly than measuring instruments are able to capture due to discrete implementation (for example, real-time GHA measures are not available). Lowest levels of accuracy are thus expected under conditions in which antecedent attributes (weather) change randomly. This study thus investigated accuracy of a UniODA[2] model obtained after data were randomly assigned into either training (n=164) or hold-out validity (n=133) samples.[12]

Because data were split into training and hold-out samples it was necessary to ensure that there were sufficient responses in every rating category in both samples. The total of 297 responses included two ratings of 10% and 18 of 20% that were combined with 24 ratings of 30% to construct the new lower-end category, ≤30%. And, 7 ratings of 90% were combined with 22 ratings of 80% to construct the new higher-end category, ≥80%.

For GHA *training* data: mean=5575; SD =164; median=5565; skewness=-0.11; kurtosis= -0.46; CV=3.0. For *hold-out* data: mean=5575; SD=151; median=5568; skewness=0.13; kurtosis=-0.82; CV=2.7.

For FOCUS *training* data: mean=6.45; SD=1.47; median=6; skewness=0.10; kurtosis= -0.88; CV=22.7. For *hold-out* data: mean=6.32; SD=1.33; median=6; skewness=0.03; kurtosis= -0.66; CV=21.1.

FOCUS and GHA were synchronized by recording date. The exploratory hypothesis was tested with the following UniODA code (control commands shown in red):

```
VARS FOCUS GHA;
CLASS FOCUS;
ATTR GHA;
MCARLO ITER 300;
GO;
```

As seen in Table 8, the results offer statistically significant support (p<0.05) for the *post hoc* hypothesis that GHA predicts FOCUS, and ESS and ESP both fall within the domain of moderate classification performance.[2]

Model predictive values are weak for predicted ratings of ≤80% of maximum; strong for predicted ratings of ≤30% of maximum; and moderate for other class categories. And, model sensitivities are weak for actual ratings of ≤50% and ≤70% of maximum; moderate for actual ratings of ≤40% of maximum; and strong for the other actual class categories—notably, the scale extremes.

Notice the exploratory UniODA model is not perfectly linear, which would be indicated if the class category codes occurred in the model in a perfectly linear order (e.g., the classes were ordered as 8, 7, 6, 5, 4, 3, with respect to GHA). As seen, the classes were ordered as 6, 8, 7, 5, 4, 3: the code 6 is moved two spaces to the left as compared with the perfectly linear model. Yet, the three highest-valued categories (6-8) are on the lower portion of the GHA domain, and the three lowest-values categories (3-5) are on the higher portion of the GHA domain. This is an example of a *Type C nonlinear reliability model* described elsewhere[2] (ps. 136-138), in which the attribute shows local regression through some or all of its range.

2155-0182/10/$3.00

Table 8: UniODA Model *Training* Performance: Predicting Mental Focus using GHA

Predictive Value

UniODA Model						
GHA	Predicted Mental Focus	*n*	% Correct	ESP	*p* <	Confidence
≤5496	6 (60% of Maximum)	51	39.2	28.7	0.05	99.999%
≤5636	8 (80% of Maximum)	53	22.6			
≤5668	7 (70% of Maximum)	9	44.4			
≤5692	5 (50% of Maximum)	9	44.4			
≤5760	4 (40% of Maximum)	16	31.2			
>5760	3 (30% of Maximum)	26	61.5			

Sensitivity

Actual Symptom Level	*n*	Number Correctly Predicted	Sensitivity (% Accuracy)	ESS
3 (30% of Maximum)	24	16	66.7	27.0
4 (40% of Maximum)	18	5	27.8	
5 (50% of Maximum)	38	4	10.5	
6 (60% of Maximum)	39	20	51.3	
7 (70% of Maximum)	26	4	15.4	
8 (80% of Maximum)	19	12	63.2	

Note: UniODA model based on non-directional (*post hoc*) hypothesis. See Note to Table 1. Confidence for target *p* based on 500 Monte Carlo experiments (estimated $p<0.0001$). Monte Carlo analysis was completed in 1.21 CPU hours by UniODA[2] running on a 3 GHz Intel Pentium D microcomputer.

.

Selected UniODA output for this analysis is presented in Table 9 to illustrate how the entries in Table 8 are found.

In Table 8, under Predictive Value, the first element of the UniODA model is GHA≤ 5496, Predicted Mental Focus=6, *n*=51, % Correct=39.2. In Table 9 this is found under the Classification performance summary table, beneath the Predicted Class 6 column.

And, in Table 8, for Sensitivity, the first Actual Symptom Level is 3, with *n*=24, Number Correctly Predicted=16, Sensitivity (% Accuracy)=66.7. In Table 9, the number correctly predicted is found *in* the Classification performance summary table, as the intersection of Predicted and Actual Class 3, and the remaining values are found *to the right* of the table for the row representing actual Class category 3.

2155-0182/10/$3.00

Table 9: Output Showing UniODA Model and Classification Performance Summary for *Training* Analysis: GHA and Mental Focus Example

```
ODA model:
----------
IF GHA <= 5495.5 THEN FOCUS = 6
IF 5495.5 < GHA <= 5636.0 THEN FOCUS = 8
IF 5636.0 < GHA <= 5668.5 THEN FOCUS = 7
IF 5668.5 < GHA <= 5691.5 THEN FOCUS = 5
IF 5691.5 < GHA <= 5760.0 THEN FOCUS = 4
IF 5760.0 < GHA THEN FOCUS = 3


Classification performance summary:
-----------------------------------    Overall      Mean Sens
  Correct   Incorrect                  accuracy   across classes
     61        103                      37.20%       39.13%

Class                      Predicted
 V4         3       4       5       6       7       8
        ------------------------------------------------- NA      Sens
         |       |       |       |       |       |       |
       3|  16   |   1   |   2   |   0   |   1   |   4   | 24     66.67%
         |       |       |       |       |       |       |
A      4|   4   |   5   |   0   |   4   |   0   |   5   | 18     27.78%
c        |       |       |       |       |       |       |
t      5|   6   |   6   |   4   |  12   |   3   |   7   | 38     10.53%
u        |       |       |       |       |       |       |
a      6|   0   |   2   |   2   |  20   |   1   |  14   | 39     51.28%
l        |       |       |       |       |       |       |
       7|   0   |   0   |   1   |  10   |   4   |  11   | 26     15.38%
         |       |       |       |       |       |       |
       8|   0   |   2   |   0   |   5   |   0   |  12   | 19     63.16%
         |       |       |       |       |       |       |
        -------------------------------------------------
 NP      26      16       9      51       9      53
 PV    61.54%  31.25%  44.44%  39.22%  44.44%  22.64%  Mean PV  40.59%

Effect strength Sens  26.96%      Effect strength PV  28.71%
```

Table 10 summarizes findings when the UniODA model identified for the training sample was used to classify the observations in the hold-out validity sample. As seen, results offer statistically significant support (p<0.01) for the *a priori* hypothesis that the training model using GHA to predict FOCUS will generalize to hold-out sample, however ESS and ESP were both in the domain of weak classification performance.[2]

Model predictive values are weak for all predicted ratings except for the strong predicted ratings of 30% and 50% of maximum. Model sensitivities are strong for extreme actual ratings of 30% and 80% of maximum; moderate for actual ratings of 60% of maximum; and weak for the other actual symptom ratings.

The following UniODA[2] code was used to estimate confirmatory p for hold-out results (control commands are indicated in red):

2155-0182/10/$3.00

Table 10: UniODA Model *Hold-Out* Performance: Predicting Mental Focus using GHA

Predictive Value

GHA	UniODA Model Predicted Mental Focus	*n*	% Correct	ESP	*p* <	Confidence
≤5496	6 (60% of Maximum)	45	24.4	19.8	0.01	99.999%
≤5636	8 (80% of Maximum)	45	11.1			
≤5668	7 (70% of Maximum)	5	0			
≤5692	5 (50% of Maximum)	4	75.0			
≤5760	4 (40% of Maximum)	12	25.0			
>5760	3 (30% of Maximum)	22	63.6			

Sensitivity

Actual Symptom Level	*n*	Number Correctly Predicted	Sensitivity (% Accuracy)	ESS
3 (30% of Maximum)	20	14	70.0	16.6
4 (40% of Maximum)	18	3	16.7	
5 (50% of Maximum)	31	3	9.7	
6 (60% of Maximum)	30	11	36.7	
7 (70% of Maximum)	24	0	0	
8 (80% of Maximum)	10	5	50.0	

Note: See Note to Table 8. TheUniODA model identified for training sample was applied to the hold-out sample in a confirmatory (*a priori*, one-tailed) analysis.[2] UniODA was used to estimate hold-out *p* with 5,000 Monte Carlo experiments (estimated $p<0.0002$), and yielded 99.999% (due to memory limit, UniODA reported 100%) Confidence for target $p<0.01$. Model performance indices declining in hold-out analysis are shown in red.

```
OPEN DATA;
OUTPUT holdout.out;
CATEGORICAL ON;
TABLE 6;
CLASS ROW;
DIRECTIONAL < 1 2 3 4 5 6;
MCARLO ITER 100000 TARGET .0003;
DATA;
14 3 0 0 1 2
4 3 0 6 0 5
4 1 3 11 2 10
0 4 0 11 2 13
0 1 1 12 0 10
0 0 0 5 0 5
END DATA;
GO;
```

2155-0182/10/$3.00

Used to analyze data organized in square tables, TABLE requires a parameter indicating the number of rows (and columns) in the table to be analyzed. The CLASS variable is indicated as ROW because actual class category forms the *rows* of the confusion table that is used as data for the analysis. The DIRECTIONAL command uses the training model to classify observations in the hold-out sample.[2] A test run indicated that simulation proceeded quickly and p was small, so 100,000 MONTE CARLO experiments were run to obtain confidence for target $p<0.0003$ for expository purposes (analysis was completed in 3 CPU seconds by UniODA[2] running on a 3 GHz Intel Pentium D microcomputer). Selected output from this analysis is given in Table 11.

Table 11: UniODA Output Showing Estimated Confirmatory *p*
for Hold-Out Results Given in Table 10: GHA and Mental Focus Example

```
ODA model:
----------
IF COLUMN = 1 THEN ROW = 1
IF COLUMN = 2 THEN ROW = 2
IF COLUMN = 3 THEN ROW = 3
IF COLUMN = 4 THEN ROW = 4
IF COLUMN = 5 THEN ROW = 5
IF COLUMN = 6 THEN ROW = 6

Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations                Estimated p
----------                -----------
 100000                    .000070

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

Target p    Confidence      Target p    Confidence
---------   ----------      ---------   ----------
p<.000300   100.00%         p>.000300     0.00%
```

All possible ACTs were examined next. As seen in Table 12, the more extreme the class categories, the stronger the sensitivity and the predictive value of the model. This procedure is motivated because, as was discussed earlier, the three lowest FOCUS ratings are on one side of the UniODA models predicted class category ordering, and the three highest FOCUS ratings are on the other side of the predicted ordering. A strong ESS is obtained for the least extreme division between class categories (essentially a "median split" performed on the categories) for the training sample. For the next-most extreme division eliminating the two intermediate class categories, training and hold-out ESS and ESP metrics all fall in the domain of a strong effect. And, for the most extreme, highest reliability division involving only the two most extreme class categories, three of the four measures fall in the domain of a very strong effect.[2]

2155-0182/10/$3.00

Table 12: All Possible Confusion Tables for GHA and Mental Focus Example: Training (*n*=164) and Hold-Out (*n*=133) Analyses

Actual Focus	Predicted Mental Focus 30%	40%	50%	60%	70%	80%
30%	16	1	2	0	1	4
	14	3	0	0	1	2
40%	4	5	0	4	0	5
	4	3	0	6	0	5
50%	6	6	4	12	3	7
	4	1	3	11	2	10
60%	0	2	2	20	1	14
	0	4	0	11	2	13
70%	0	0	1	10	4	11
	0	1	1	12	0	10
80%	0	2	0	5	0	12
	0	0	0	5	0	5

Actual	Predicted Mental Focus 30-50	60-80
30-50	44	36
	32	37
60-80	7	77
	6	58
	ESP=54.4	ESP=45.3
	ESS=46.7	ESS=37.0

Actual	Predicted Mental Focus 30-40	70-80
30-40	26	10
	24	8
70-80	2	15
	2	24
	ESP=63.4	ESP=57.5
	ESS=64.5	ESS=63.2

Actual	Predicted Mental Focus 30	80
30	16	4
	14	2
80	0	12
	0	5
	ESP=75.0	ESP=71.3
	ESS=80.0	ESS=87.5

Note: *Red entries* are for *training* model, *black entries* are for *hold-out* model. ESS, ESP, and *p* given for root confusion tables in Tables 8 and 10. Training ACT excluding categories 50 and 60 classified *n*=53 (32.3% of sample), and excluding categories 40-70 classified *n*=32 (38.7% of sample); hold-out ACT excluding categories 50 and 60 classified *n*=58 (43.6% of sample), and excluding categories 40-70 classified *n*=32 (38.7% of sample).

2155-0182/10/$3.00

Findings suggest improved measurement of GHA (e.g., obtaining a precise value for the patient's location when the symptom rating was made, rather than for general geographic area as done presently due to availability limitations), and of patient symptoms (e.g., perhaps by using a more reliable categorical ordinal scale[2]), offers promise of increasing model ESS and ESP.

References

[1]Yarnold PR (2013). Minimum standards for reporting UniODA findings for class variables with three or more response categories. *Optimal Data Analysis,* 2, 86-93.

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

[3]Yarnold PR, Soltysik RC (2010). Precision and convergence of Monte Carlo Estimation of two-category UniODA two-tailed *p*. *Optimal Data Analysis, 1*, 43-45.

[4]Yarnold PR (2013). Maximum-accuracy multiple regression analysis: Influence of registration on overall satisfaction ratings of emergency room patients. *Optimal Data Analysis*, *2*, 72-75.

[5]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.

[6]Yarnold PR (1987). Norms for the Glass model of the short student version of the Jenkins Activity Survey. *Social and Behavioral Science Documents, 16,* 60. MS# 2777.

[7]Yarnold PR Lyons JS (1987). Norms for college undergraduates on the Bem Sex-Role Inventory and the Wiggins Interpersonal Behavior Circle. *Journal of Personality Assessment, 51*, 595-599.

[8]Yarnold PR, Bryant FB (1988). A note on measurement issues in Type A research: Let's not throw out the baby with the bath water. *Journal of Personality Assessment*, *52*, 410-419.

[9]Collinge WC, Soltysik RC, Yarnold PR (2010). An internet-based intervention for fibromyalgia self-management: Initial design and alpha test. *Optimal Data Analysis, 1*, 163-175.

[10]Collinge W, Yarnold PR, Soltysik, RC (2013). Fibromyalgia symptom reduction by online behavioral self-monitoring, longitudinal single subject analysis and automated delivery of individualized guidance. *North American Journal of Medical Sciences, 5*, 546-553.

[11]Yarnold PR, Soltysik RC, Collinge W (2013). Modeling individual reactivity in serial designs: An example involving changes in weather and physical symptoms in fibromyalgia. *Optimal Data Analysis, 2*, 37-42.

[12]This was accomplished using the following SAS™ code:

```
data q;infile 'c:\total.dat';
input FOCUS GHA;
if FOCUS<3 then FOCUS=3;
if FOCUS=9 then FOCUS=8;
data randomlist;
do i=1 to 297;
if uniform(453233)<0.5 then holdout=0;
else holdout=1;
output;
end;
data q2;merge q randomlist;
file 'c:\train.dat';
if holdout=0;put FOCUS GHA;
data q3;merge q randomlist;
file 'c:\holdout.dat';
if holdout=1;put FOCUS GHA;
run;
```

Author Notes

E-mail: Journal@OptimalDataAnalysis.com.

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