Optimal Data Analysis
2010, Vol. 1, Release 1 (September 17, 2010), 62-66

2155-0182/10/$3.00

UniODA vs. Chi-Square: Ordinal Data Sometimes Feign Categorical

Paul R. Yarnold, Ph.D. and Robert C. Soltysik, M.S.
Optimal Data Analysis, LLC

Assessed using perhaps the most widely used type of measurement scale in all science, ordinal data are often misidentified as being categorical, and incorrectly analyzed by chi-square analysis. Three examples drawn from the literature are reanalyzed.

Consisting of a relatively small number of graduated levels of the measured attribute, ordinal scales may be the most broadly employed type of measurement scale in all of science. Likert-type scales, typically involving between three and ten levels, are perhaps most common.[1] For example, one's socioeconomic status is often assessed using a three-level ordinal scale, with categories corresponding to low, middle, and upper class. Also widely used, ordinal categorical scales consist of a relatively small number of qualitative categories ordered with respect to some theoretical factor.[2] For example, at the end of a clinical trial patients might be classified as being worse, unchanged, or better: the three qualitative categories are worse, unchanged, and better; the theoretical factor is quality of clinical outcome; and the categories are ordered from lowest (worse) to highest (better) with respect to quality of clinical outcome.

Since the metric underlying the attribute is ordinal, neither chi-square (nominal data) nor *t*-test (interval data) is appropriate to assess if therapies can be discriminated on the basis of clinical outcome. Traditional methods used for analysis of ordinal data include Mann-Whitney *U* test or the log-linear model, but excessive ties compromise *U*, and maximum likelihood-based methods require large samples.[3-5] Assuming neither the absence of ties nor the presence of large samples, univariate optimal discriminant analysis (UniODA) is ideal for such designs.

Plaintiff Gender and Age

Seaman and Hill[6] analyzed data obtained by Cox and Key[7] from court records of an Ohio county, involving the frequency of plaintiffs in divorce actions cross-classified by gender (wife or husband) and age (<25, 25-34, 35-44, >44). "The hypothesis that the proportion of plaintiffs that are husbands is the same, regardless of age" (p. 454) was tested using the traditional model, homogeneity of proportions. All possible *post hoc* pairwise comparisons—involving 6 separate 2-by-2 chi-square tests, were conducted to ascertain the specific reason the omnibus test was statistically significant. Two pairwise comparisons were statistically significant: those comparing the >44 age category with the 25-34 and 35-44 categories (p's<0.05). Analysis via chi-square thus indicated a greater proportion of husband plaintiffs in the >44 age category, and a greater proportion of wife plaintiffs in the 25-34

2155-0182/10/$3.00

and 35-44 age categories. No statistically significant pairwise comparisons involved the <25 age category, so this strata could not be assessed in relation to other strata in the study.

Table 1: Plaintiff Age in a Divorce Action

Age	<25	25-34	35-44	>44
Husband	8	8	6	16
Wife	18	48	22	10

Note: Adapted from Seaman and Hill (1996). Tabled are frequency counts.

After gender (1=Husband, 2=Wife) and age (1='<25', 2='25-34', 3='35-44', 4='>44') were dummy-coded, data were reanalyzed using the following ODA[5] code (commands indicated in red; non-directional exploratory analysis is conducted as no *a priori* hypothesis was made):

```
open data;
output seaman.out;
vars gender age;
data;
2 4 (repeated 10 times)
2 3 (repeated 22 times)
2 2 (repeated 48 times)
2 1 (repeated 18 times)
1 4 (repeated 16 times)
1 3 (repeated 6 times)
1 2 (repeated 8 times)
1 1 (repeated 8 times)
end;
class gender;
attr age;
mcarlo iter 25000;
loo;
go;
```

The resulting UniODA model was: if age≤35-44 then predict class=wife, otherwise predict class=husband. The model achieved a moderate ESS of 31.9 (p<0.0001), and results were stable in jackknife validity analysis. The model classified 88 (90%) of 98 women correctly, versus only 16 (42%) of 38 men. All subjects were classified by the ODA model, including those younger than 25 years of age.

Outcomes of Marital Therapy

Snyder, Wills and Grady-Fletcher[8] reported the following four-year termination outcomes of two different types of therapy for unhappily married couples. The expected value for both entries in the right-most column of the data table is less than five, invalidating the use of chi-square with this sparse table.[9] An omnibus chi-square statistic was given for the 2-by-3 table, then eyeball interpretation of the omnibus effect was rendered: "a significantly higher percentage of (behavior therapy couples) had experienced divorce, p<0.01." Although no explanation was provided—perhaps to defeat the aforementioned minimum expectation assumption violation, the No Change ("distressed") and Improved classes were collapsed and chi-square reported higher divorce rates for behavior therapy, p<0.05.

Table 2: Outcomes of Marital Therapies

Type of Therapy	Divorced	No Change	Improved
Insight	3	22	4
Behavior	12	13	1

Note: Tabled are frequency counts.

These data were analyzed by ODA code paralleling that used in the first example. The model was: if outcome=divorced then predict therapy=behavior, otherwise predict therapy= insight. The model correctly classified 90% in insight therapy, 46% in behavior therapy, and yielded a moderate ESS of 35.9 (p<0.006).

2155-0182/10/$3.00

Strength of Gender Differences

Hyde and Plant[10] reported frequencies of five categories of Cohen's *d* measure of effect strength for representative studies of gender differences, versus studies of other effects in the field of psychology. An omnibus chi-square statistic was provided for the 2-by-5 table ($p<0.0001$): "The difference between the distributions of gender effect sizes and other effect sizes is highly significant." Pairwise comparisons to disentangle the omnibus effect were not reported. Eyeball analysis suggested: "more gender differences fall in the close-to-zero category than other effects in psychology."

Table 3: Cohen's *d* by Type of Study

Type of Study	≤0.1	≤0.35	≤0.65	≤1.0	>1.0
Gender	43	60	46	17	5
Other	17	89	116	60	20

Note: Tabled are frequency counts.

For these data the exploratory hypothesis that type of study could be discriminated on the basis of effect strength was tested using priors-weighted UniODA, via ODA code paralleling that used in the prior examples. The model was: if $d \leq 0.35$ then predict gender study; otherwise, predict non-gender study. Thus, relative to other areas, gender studies have disproportionately more effect sizes in the close-to-zero (≤0.1) and next-to-close-to-zero (0.11-0.35) categories. By correctly classifying 60.2% of the gender difference studies, and 64.9% of other studies, the model yielded a moderate, jackknife-stable ESS=25.1 ($p<0.0001$).

Discussion

Initial study of the congruence between chi-square and UniODA in analysis of real-world data suggests consistent findings may often be achieved, and instances of inconsistent findings may often accompany grossly imbalanced marginals.[11] Distinct advantages versus chi-square include that, for UniODA: *directional* tests of statistical hypotheses may be conducted; the validity of exact *p* is uncompromised by sparse, empty or missing cells, small samples or imbalanced marginal distributions; and use of the normed ESS index allows direct comparison of model performance across analyses differing in number of observations, marginal imbalance, and/or number of levels for categorical class variables and/or attributes.

Optimal ordinal analysis may be generalized to designs involving class variables having more than two categories (Yarnold and Soltysik[5] discuss degenerate designs involving fewer categories for attribute than class). For example, imagine a design involving a three-category class variable—such as therapies A, B, and C, and an ordinal categorical attribute with at least three ordinal improvement categories—such as none, some, and much. A UniODA model for such a design would be of the form: if improvement=none, predict therapy=A; otherwise, if improvement=some, predict therapy=B; otherwise predict therapy=C. As is true for all ODA applications, for three-category designs: exact *p* is obtained for performance achieved by the model; mean sensitivity across therapies is translated into the normed ESS scale of effect strength; and leave-one-out (LOO) "jackknife" validity analysis is used to assess the potential generalizability of the findings were the model used to classify independent random samples.

Generalizing exact ordinal analysis to designs involving more than one assessment dimension is also straightforward, whether by linear or nonlinear methods. Imagine an application having two therapeutic strategies (class variable) and two ordinal categorical outcome scales (attributes)—one assessing degree of recovery (worse, unchanged, better), and the other assessing satisfaction (unhappy, neutral,

2155-0182/10/$3.00

happy). Using an optimal multivariable linear approach with these data, one could obtain a main effects model including an intercept and separate coefficients for recovery and satisfaction; a saturated model additionally including a coefficient for the recovery-by-satisfaction interaction; and a quadratic model additionally including coefficients for the squares (or higher exponents) of each main effect.[12-14] Model coefficients may be real numbers, or may be constrained to any range, even binary.[15] Structurally, these ODA models are similar to models developed via traditional multivariable techniques such as discriminant or logistic regression analysis. Functionally, however—as is constitutionally true of all ODA analyses, these models would explicitly maximize (weighted) classification accuracy achieved for the sample.[5] Using an optimal multivariable nonlinear approach with these data currently entails conducting hierarchically optimal classification tree analysis, or CTA.[16]

Regardless of choice of (non)linear method, to ensure the validity of analytic findings it is recommended that variables which truly are measured using an ordinal scale are treated as though they were in fact measured using an ordinal scale.

References

[1]Nunnally JC. *Psychometric theory*. McGraw-Hill, New York NY, 1967.

[2]Kazdin AE. *Research design in clinical psychology (2nd ed.)*. Allyn & Bacon, New York NY, 1992.

[3]Mann HB, Whitney DR. On a test of whether one of two variables is stochastically larger than the other. *Annals of Mathematical Statistics* 1947, 18:50-60.Snyder, Wills, 1991

[4]Hagenaars JA. *Categorical longitudinal data: log-linear, panel, trend, and cohort analysis.* Sage, Newbury Park, CA, 1990.

[5]Yarnold PR, Soltysik RC. *Optimal data analysis: a guidebook with software for Windows*. APA Books, Washington, DC, 2004.

[6]Seaman MA, Hill CC. Pairwise comparisons for proportions: a note on Cox and Key. *Educational and Psychological Measurement* 1996, 56:452-459.

[7]Cox MK, Key CH. Post hoc pairwise comparisons for the chi-square test of homogeneity of proportions. *Educational and Psychological Measurement* 1993, 53:951-962.

[8]Snyder DK, Wills RM, Grady-Fletcher A. Long-term effectiveness of behavioral versus insight-oriented marital therapy: a four-year follow up study. *Journal of Consulting and Clinical Psychology* 1991, 59:138-141.

[9]Yarnold JK (1970). The minimum expectation of chi-square goodness-of-fit tests and the accuracy of approximations for the null distribution. *Journal of the American Statistical Association*, *65*, 864-886.

[10]Hyde JS, Plant EA. Magnitude of psychological gender differences: another side to the story. *American Psychologist* 1995, 50:159-161.

[11]Yarnold PR, Hart LA, Soltysik RC. Optimizing the classification performance of logistic regression and Fisher's discriminant analyses. *Educational and Psychological Measurement* 1994, 54:73-85.

[12]Yarnold PR, Soltysik RC, Martin GJ. Heart rate variability and susceptibility for sudden cardiac death: an example of multivariable optimal discriminant analysis. *Statistics in Medicine* 1994, 13:1015-1021.

[13]Soltysik RC, Yarnold PR. The Warmack-Gonzalez algorithm for linear two-category multivariable optimal discriminant analysis. *Computers and Operations Research* 1994, 21: 735-745.

2155-0182/10/$3.00

[14]Yarnold PR, Soltysik RC, McCormick WC, Burns R, Lin EHB, Bush T, Martin GJ. Application of multivariable optimal discriminant analysis in general internal medicine. *Journal of General Internal Medicine* 1995, 10:601-606.

[15]Yarnold PR, Soltysik RC, Lefevre F, Martin GJ. Predicting in-hospital mortality of patients receiving cardiopulmonary resuscitation: unit-weighted MultiODA for binary data. *Statistics in Medicine* 1998, 17:2405-2414.

[16]Yarnold PR, Soltysik RC, Bennett CL. Predicting in-hospital mortality of patients with AIDS-related *Pneumocystis carinii* pneumonia: an example of hierarchically optimal classification tree analysis. *Statistics in Medicine* 1997, 16:1451-1463.

Author Notes

Address correspondence to the authors at: Optimal Data Analysis, 1220 Rosecrans St., Suite 330, San Diego, CA 92106. Send E-mail to: Journal@OptimalDataAnalysis.com.