Optimal Data Analysis
Vol. 5 (October 9, 2016), 121-123

2155-0182/10/$3.00

ODA *vs*. Undocumented Chi-Square: Clarity *vs*. Confusion

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

A longitudinal smoking cessation study followed three patient groups: group 1=40 patients attending at least one group session; group 2=62 interviewed patients who did not attend group sessions; group 3=group 1+group 2. Data were collected four times: at the interview, and two-weeks and one- and two-months post-discharge.[1] Each time patients rated their smoking behavior using a six-point scale: 1=quit smoking; 2=reduced smoking ≥50%; 3=reduced smoking <50%; 4=switched to pipe; 5=no change; 6=smoking increased. The scale is linear—assessing monotonically decreasing smoking—except for response option 4 (with N≤2 at all four testings). Option 4 made the response scale nonlinear so the author used chi-square analysis to compare groups within and across testings (the latter test is a violation of the chi-square assumption that all observations appear once in the design matrix). Nevertheless, N≤2 for Option 4 causes violation of the minimum expectation assumption.[2] Furthermore, were an omnibus effect to be identified then the pairwise comparisons needed to identify the differences that produced the effect would also violate the minimum expectation assumption. The author offered two general statements regarding undocumented chi-square-based findings that violated two crucial underlying assumptions, and offered qualitative discussion. This is a surprisingly common practice in studies involving multicategorical variables that are analyzed using chi-square analysis: ODA clarifies the findings in such applications.[3-8]

Data drawn from this study are given in Table 1. The original report did not discuss the statistical analysis strategy nor provide information about test statistics. However, two qualitative statistical summaries were provided. First, the author concluded: "For the two-week, one-month and two-month follow-up periods, both groups maintained or increased the degree of withdrawal from cigarette smoking" (p. 108). Second: "The discrepancy between the number (of cigarettes) smoked while in the hospital and one month after discharge is *not statistically significant* (emphasis in original, pp. 108-109).

Omission of discussion about statistical methods, and of summary statistics, constitutes an incomplete accounting of the study findings: what is needed is clear, documented statistical evidence that effects did or did not occur.

2155-0182/10/$3.00

Table 1: Data for Smoking Study (N is tabled for each response scale category)

Testing	Group	1	2	3	5	6
At the	1	11	5	2	20	0
Discharge	2	16	10	2	30	4
Interview	3	27	15	4	50	4
Two-	1	12	11	6	6	4
Weeks	2	17	11	10	19	5
Later	3	29	22	15	25	9
One-	1	10	7	7	4	4
Month	2	14	11	9	15	5
Later	3	24	18	16	19	9
Two-	1	7	4	3	2	1
Months	2	9	9	9	11	4
Later	3	16	13	12	13	5

The original study combined groups 1 and 2, perhaps in an effort to increase sample size and thus limit violations of the minimum expectation assumption, or perhaps to increase statistical power. ODA was used to compare group 1 and group 2, separately for each testing session, to ensure that combining groups did not produce paradoxical confounding.[7-9] Group was treated as the class variable in these analyses.

Two sets of analyses were conducted: the first compared groups 1 and 2 with respect to the 6-point scale (including Option 4) treated as being a categorical attribute (all p's>0.78); the second compared groups 1 and 2 with respect to a 5-point scale (excluding Option 4) treated as an ordered (i.e., categorical ordinal[7,8]) attribute (all p's>0.27). Groups 1 and 2 cannot be discriminated from each other at any of the testing periods, and when the data are combined the estimate of the percentage of sample using each response category for the combined sample fell between the corresponding percentages for groups 1 and 2. Thus, there is no evidence of paradoxical confounding for group 3 data which are thus used henceforth in analyses herein.

Solutions for the 6-point categorical scale were dominated: lower accuracy for larger cohorts was dismissed so as to correctly classify the one or two Option 4 responses. Thus, Option 4 was omitted from subsequent ordered analyses conducted using the 5-point scale. The final set of six ODA analyses tested all six possible paired comparisons between the four testings.

The first three analyses compared the 5-point, group 3 response scale score that was recorded at the first testing session versus the corresponding score recorded at each of the three follow-up testings. The same ODA model emerged in all three comparisons: if score$\leq$3 (indicating lower tobacco consumption) predict testing=follow-up, otherwise predict testing= interview. The models had greater sensitivity for predicting follow-up responses (range=66.0-69.5) versus interview responses (54.0 for all three models). All three comparisons of the response scale score recorded at the interview versus recorded at follow-up were statistically reliable, and all three returned a relatively weak effect that was stable in jackknife analysis: *vs*. two-weeks, ESS=20.0, p<0.011; *vs*. one-month, ESS=21.4, p<0.0069; *vs*. two-months, ESS= 23.5, p<0.0075.

These findings support the author's first statistical summary statement: "For the two-week, one-month and two-month follow-up periods, both groups (i.e., group 3) maintained or increased…withdrawal from cigarette smoking". However, the ODA finding that the response scale score at one-month post-discharge is *significantly lower* than was recorded at the interview refutes the author's second summary statement: "The discrepancy between the number (of cigarettes) smoked while in the hospital and one month after discharge is *not statistically significant* (emphasis in original)".

Finally, there were no statistically reliable differences between response scale scores between testings 2 and 3 (ESS=2.2, p<0.98), testings 2 and 4 (ESS=3.5, p<0.98), or testings 3 and 4 (ESS=2.1, p<0.99).

2155-0182/10/$3.00

References

[1]Reed K (1967). Smoking control with an in-patient general hospital population. In: Zagona SV (Ed.), *Studies and issues in smoking behavior*. Tucson, AZ: University of Arizona Press (pp. 107-109).

[2]Yarnold JK (1970). The minimum expectation in χ^2 goodness of fit tests and the accuracy of approximations for the null distribution. *Journal of the American Statistical Association*, *65*, 864-886.

[3]Yarnold PR (2013). Analyzing categorical attributes having many response categories. *Optimal Data Analysis*, *2*, 172-176.

[4]Yarnold PR (2013). Univariate and multivariate analysis of categorical attributes with many response categories. *Optimal Data Analysis*, 2, 177-190.

[5]Yarnold PR (2015). UniODA *vs*. chi-square: Deciphering *R* x *C* contingency tables. *Optimal Data Analysis*, *4*, 156-158.

[6]Yarnold PR (2016). UniODA *vs*. chi-square: Describing baseline data from the National Pressure Ulcer Long-Term Care Study (NPULS). *Optimal Data Analysis*, *5*, 24-28.

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

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

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

Author Notes

The study analyzed de-individualized data and was exempt from Institutional Review Board review. No conflict of interest was reported.

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