Optimal Data Analysis
Vol. 5 (October 7, 2016), 112-115

2155-0182/10/$3.00

CTA *vs. Not* Chi-Square: Differentiating Statistical and Ecological Significance

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

Statistical and ecological significance are rarely distinguished in reports employing chi-square analysis to test statistical hypotheses. This note elucidates this distinction, comparing chi-square with CTA using an application relating smoking knowledge and smoking behavior.[1]

Data are drawn from a study comparing *success* (>90% reduction) *and failure* (<10% reduction) in quitting smoking (the class variable, dummy-coded for all observations using 1 and 0, respectively) rates *of aided and unaided* (the first attribute, treated as categorical, dummy-coded for observations as 1 and 0, respectively), *male and female* smokers (the second attribute, also categorical and dummy-coded as 1 and 0, respectively), recorded at three- and six- month post-intervention *follow-up testing periods* (the third and final attribute, treated as ordered with values of 3 and 6, respectively).

The design matrix consisted of 16 cells created by cross-classifying the class variable (2 levels) x aid (2 levels) x gender (2 levels) x test period (2 levels). Values used to populate matrix cells were computed from corresponding percentages that were reported as truncated integers in the original article: corresponding round-off error effectively subtracted a modest constant from half the design cells thus creating a sample that has 12 fewer observations than the unreported actual data. However, the author states: "In most cases these percentages will not add to 100 per cent because only those findings which are of major interest are given (pp. 96-97). The relative magnitude of this difference however is inconsequential when considered in light of the pattern of the results: that is, many small inter-cell differences and only a few large inter-cell differences emerged. Table 1 gives the number of observations falling into each matrix cell.

Table 1: Smoking Study Data for the Exposition

		Three Months		Six Months	
Gender	Aid	Succeed	Fail	Succeed	Fail
Male	Yes	26	23	22	21
Male	No	25	15	17	20
Female	Yes	32	19	26	23
Female	No	15	45	12	46

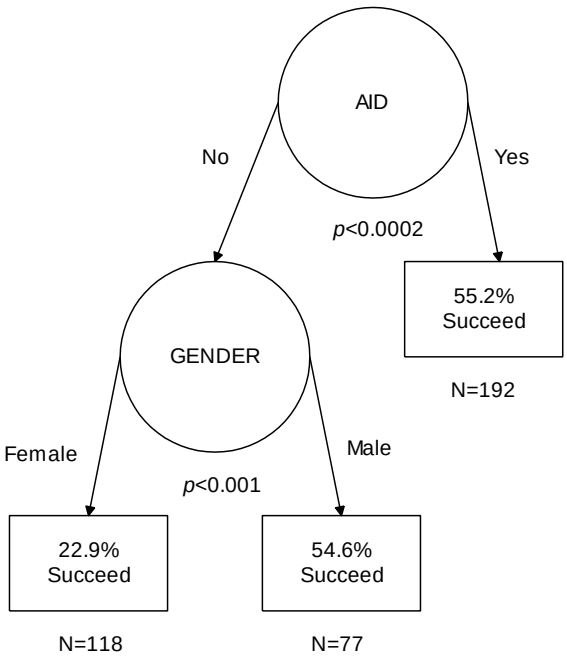
The "repeated measures" feature of the design is problematic for chi-square analysis that assumes every observation in the sample appears once in the design matrix. The original study therefore evaluated the data using eight chi-square analyses: four comparing the *success*

2155-0182/10/$3.00

rate between the aided versus unaided groups, separately by gender and testing period; and four comparing the *failure* rate between aided versus unaided groups, separately by gender and testing period. All analyses involving males had $p>0.05$ and all involving females had $p<0.05$. With respect to comparisons between aided versus unaided groups, it was concluded: "…men maintain the same success and failure rates… while the women have markedly lower success and markedly higher failure rates" (p. 95). Here the qualitative assessment "markedly" refers to the raw difference in success and failure rates of aided versus unaided groups of females, which was $\leq$20% at three, and $\leq$24% at six months.

CTA was used to analyze the data in this design as described earlier: success/failure is the class variable, aid and gender are categorical attributes, and test period is an ordered attribute. Figure 1 is the three-strata model that emerged.

Figure 1: CTA Model Predicting Success or Failure in Quitting Smoking



For CTA the maximum raw difference in success rate between model strata was >32% for both pairwise comparisons involving the left-most endpoint, versus <24% for all chi-square analyses. And, Type I error rates for the CTA model (p's<0.001) are an order of magnitude smaller than for chi-square (p's<0.01). The differences identified by CTA might thus be qualitatively assessed as "greater than markedly different" to be linearly consistent with the original report. However, statistical analysis is needed to evaluate the *reliability* of the raw differences, and the number of zeroes to the right of the decimal (for p) is not a measure of the *magnitude* of the difference, but rather of the *reliability* of the difference—the likelihood that a difference as large as was observed might have occurred by chance.

The strength of the difference, that is of the effect, is computed separately from p on the basis of the model's sensitivities (the ability of the model to accurately predict the actual class status of the observations in the sample).[4,5] For example, the confusion matrix for this three-strata model is presented in Table 2: the model sensitivity for predicting failure is 42.9%, and the model sensitivity for predicting success is 84.6%. A sensitivity of 50% is expected for each of the two class categories—failure and success—by chance.[4,5]

Table 2: Confusion Matrix for CTA Model Predicting Success/Failure in Quitting Smoking

		Predicted Outcome		
		Failure	Success	
Actual Outcome	Failure	91	121	42.9%
	Success	27	148	84.6%

The effect strength for sensitivity or ESS statistic is computed to summarize predictive accuracy achieved by the model.[4,5] For every analysis ESS=0 represents the accuracy that is expected by chance, and ESS=100 represents perfect, errorless accuracy. For this example the CTA model achieved ESS=27.5. Simulation research identified rules-of-thumb for making qualitative summaries of the strength of an

2155-0182/10/$3.00

effect normed against chance: ESS<25 is a relatively weak effect; ESS<50 is a moderate effect; and ESS values of 50 and greater reflect various levels of strong performance.[4,5] Thus the present effect is qualitatively described by this rule as (barely) being of moderate strength. In summary, CTA identified a statistically reliable, cross-generalizable (total sample and jackknife accuracy were identical) effect of moderate strength (ESS=27.5).

The substantive implications of the CTA model are tantalizing. The root variable is aid, which dominates the analytic solution and thus uncontestably indicates that—compared to the other attributes, and to chance—the intervention was efficacious. Quantitatively, half (55.2%) of the people who were given aid to stop smoking were in fact able to stop.

However, of the people who were not given aid to stop, half (54.55%) of the males, and 3 of 10 (22.9%) females were able to stop smoking. Recall that chi-square found that no males improved, and all females improved. The CTA model refutes the findings of chi-square, but the CTA model examined all attributes simultaneously in order to explicitly identify the most accurate model possible for the sample, whereas chi-square only considered aid in the absence of the other attributes. The CTA results suggest that males will succeed at the same high rate regardless of whether they receive aid. In contrast, females who don't receive aid fail 80% of the time. Therefore, without consideration of base rates, this pattern suggests that the majority (if not all) of the recruitment resources should be allocated for females, to ensure that more than 20% succeed.

Finally, a more modern, powerful legacy method for addressing the data presented herein is the log-linear model. Research comparing the log-linear model with structural decomposition analysis (SDA)—the maximum-accuracy analogue to principal components analysis[6] that maximizes predictive accuracy rather than variance[4,5]—consistently found the latter models are more accurate, usually the difference is between a very strong model versus a very weak model, respectively.[5,8-11]

Regardless, the use of chi-square in the manner presented herein is ubiquitous in the literature, and it is important that researchers understand the distinction between statistical reliability (*p*) and ecological significance (ESS) when interpreting and discussing their findings.

References

[1]Leventhal H, Singer R, Jones S (1965). Effects of fear and specificity of recommendation upon attitudes and behavior. *Journal of Personality and Social Psychology, 2*, 20-29.

[2]Leventhal H (1967). Effect of fear communications in the acceptance of preventive health practices. In: Zagona SV (Ed.), *Studies and issues in smoking behavior*. Tucson, AZ: University of Arizona Press (pp. 17-27).

[3]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.

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

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

[6]Grimm LG, Yarnold PR (1995). *Reading and Understanding Multivariate Statistics*. Washington, DC: APA Books.

[7]Grimm LG, Yarnold PR (2000). *Reading and Understanding More Multivariate Statistics*. Washington, DC: APA Books.

2155-0182/10/$3.00

[8]Yarnold PR (2015). UniODA-based structural decomposition *vs*. log-linear model: Statics and dynamics of intergenerational class mobility. *Optimal Data Analysis*, *4*, 179-181.

[9]Yarnold PR (2015). Modeling religious mobility by UniODA-based structural decomposition. *Optimal Data Analysis*, *4*, 192-193.

[10]Yarnold PR (2015). UniODA-based structural decomposition *vs*. legacy linear models: Statics and dynamics of intergenerational occupational mobility. *Optimal Data Analysis*, *4*, 194-196.

[11]Yarnold PR (2015). UniODA-based structural decomposition *vs*. legacy linear models: Statics and dynamics of intergenerational occupational mobility. *Optimal Data Analysis*, *4*, 194-196.

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