Optimal Data Analysis
Vol. 3 (March 27, 2014), 9-11

2155-0182/10/$3.00

UniODA *vs*. Chi-Square: Discriminating Inhibited and Uninhibited Infant Profiles

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

Kagan and Snidman[1] investigated processes mediating early reactivity to stimulation in a longitudinal study of 94 four-month-old infants who displayed a combination of either high motor activity and frequent crying, or low motor activity and infrequent crying. Fearful behavior assessed at 9 and 14 months of age was examined in relation to these two infant typologies. Eyeball analysis, which was confirmed statistically using chi-square analysis, revealed that 40% of low motor activity infants displayed "low fear" (which was arbitrarily defined as one or fewer fears) at both 9 and 14 months, versus 0% of high motor activity infants. When UniODA was applied to these data it identified statistically reliable effects at 9- and 14-months: the strongest effect occurred at 14 months. Applying CTA to these data revealed that a multiattribute model wasn't feasible.

Data in this study feature extensive dispersion across number of fears at both 9- and 14-months within both infant typologies. Such complexity makes eyeball analysis a complex and difficult task. Indeed, as seen in Table 1, the chi-square-confirmed eyeball finding[1] (indicated in red) fails to address most data in the sample.

Such classification problems are readily solved using UniODA[2,3] and CTA.[4,5] Presently, infant typology is treated as the dummy-coded class variable: low-motor/infrequent crying=0; high-motor/frequent crying=1 (actual dummy-code values used are arbitrary).[6,7] The number of fears at 9 months, and at 14 months, are both treated as ordered attributes.[8]

For *9-month data* the UniODA model was: if number of attacks$\leq$1 then predict that class=low-motor infants; otherwise predict class=high-motor infants.[12] The model achieved a moderate ESS of 40.1, and the result was statistically significant (p<0.009). The model correctly classified 22 (63%) of 35 low-motor infants, and 17 (77%) of 22 high-motor infants. The model was correct 81% of the time that it predicted an infant was low-motor, and 57% of the time it predicted an infant was high-motor.

Classification performance fell in leave-one-out (jackknife) validity analysis: ESS=26.5, p<0.05. The model correctly classified 63% of low-motor infants, and 64% of high-motor infants. The model was correct 73% of the time that it predicted an infant was low-motor, and 52% of the time it predicted an infant was high-motor. This level of classification performance is expected if the present cut-point is used to classify independent random infant samples.[2]

2155-0182/10/$3.00

Table 1: Number of fears at 9 and 14 months of age for infants classified as either *high motor*-high cry or *low motor*-low cry at four months of age: red indicates the eyeball analysis finding

Number of Fears		Number of Infants	
9 Months	14 Months	Low-Motor	High-Motor
0	0	3	0
0	1	7	0
0	2	2	
0	3	3	2
1	0	1	0
1	1	3	0
1	3	2	1
1	5	1	
1	6		1
1	≥8		1
2	1	1	
2	2	2	1
2	3		1
2	4	1	1
3	0		1
3	1	1	
3	2	1	
3	3		1
3	4		2
3	5		1
3	6		2
3	7		1
3	≥8		1
4	0	2	
4	1	1	
4	3		1
4	4		1
4	6	1	
4	7	1	
5	2	1	
5	≥8		1
≥6	0	1	
≥6	6		2

For *14-month data* the UniODA model was: if number of attacks≤2 then predict that class=low-motor infants; otherwise predict class=high-motor infants. The model achieved a relatively strong ESS of 65.2, and the result was statistically significant (p<0.0001). The model correctly classified 74% of low-motor infants, and 91% of high-motor infants. The model was correct 93% of the time that it predicted an infant was low-motor, and 69% of the time it predicted an infant was high-motor. The model performance was stable in jackknife analysis, suggesting the finding is likely to cross-generalize with comparable strength if applied to an independent random sample of infants.

No multivariable model was possible: CTA[4,5] indicated that only the 14-month data entered the model.

These findings add to a growing body of literature which suggests that rather than strain eyeballs and rattle ancient inappropriate analytic methods in the hopes of making sense of data, it is easier and more productive to use state-of-the-art, exact maximum-accuracy methods to identify underlying relationships in exploratory research, and precisely evaluate confirmatory hypotheses—in all of empirical research.

References

[1]Kagen J, Snidman N (1991). Infant predictors of inhibited and uninhibited profiles. *Psychological Science*, *2*, 40-44.

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

[3]Yarnold PR, Soltysik RC (2010). Optimal data analysis: A general statistical analysis paradigm. *Optimal Data Analysis*, *1*, 10-22.

[4]Soltysik RC, Yarnold PR (2010). Automated CTA software: Fundamental concepts and control commands. *Optimal Data Analysis*, *1*, 144-160.

[5]Yarnold PR (2013). Initial use of hierarchically optimal classification tree analysis in medical research. *Optimal Data Analysis*, 2, 7-18.

2155-0182/10/$3.00

[6]Bryant FB, Harrison PR (2013). How to create an ASCII input data file for UniODA and CTA software. *Optimal Data Analysis*, *2*, 2-6.

[7]The ASCII data set, called kagan.txt, was constructed as follows: the variables are 9- and 14-month fears, and typology, respectively.

0 0 0 (repeated 3 times)
0 1 0 (repeated 7 times)
0 2 0 (repeated 2 times)
0 3 0 (repeated 3 times)
0 3 1 (repeated 2 times)
1 0 0
1 1 0 (repeated 3 times)
1 3 0 (repeated 2 times)
1 3 1
1 5 0
1 6 1
1 8 1
2 1 0
2 2 0 (repeated 2 times)
2 2 1
2 3 1
2 4 0
2 4 1
3 0 1
3 1 0
3 2 0
3 3 1
3 4 1 (repeated 2 times)
3 5 1
3 6 1 (repeated 2 times)
3 7 1
3 8 1
4 0 0 (repeated 2 times)
4 1 0
4 3 1
4 4 1
4 6 0
4 7 0
5 2 0
5 8 1
6 0 0
6 6 1 (repeated 2 times)

[8]UniODA analysis was accomplished using the following MegaODA[9-11] code: commands are indicated in red; a non-directional exploratory analysis was conducted because there was no *a priori* hypothesis.

```
open kagan.txt;
output kagan.out;
vars month9 month14 kid_type;
class kid_type;
attr month9 month14;
mcarlo iter 25000;
loo;
go;
```

[9]Soltysik RC, Yarnold PR. (2013). MegaODA large sample and BIG DATA time trials: Separating the chaff. *Optimal Data Analysis*, 2, 194-197.

[10]Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Harvesting the wheat. *Optimal Data Analysis*, 2, 202-205.

[11]Yarnold PR, Soltysik RC (2013). MegaODA large sample and BIG DATA time trials: Maximum velocity analysis. *Optimal Data Analysis*, 2, 220-221.

[12]UniODA and eyeball cutpoints are the same. UniODA applied this to 9-month data, Kagan to 9- and 14-month data. A different cut-point is appropriate for the 14-month data (see ahead).

Author Notes

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

eMail: Journal@OptimalDataAnalysis.com