Optimal Data Analysis
Vol. 3 (March 29, 2014), 17-18

2155-0182/10/$3.00

# UniODA *vs*. Logistic Regression: Serum Cholesterol and Coronary Heart Disease and Mortality Among Middle Aged Diabetic Men

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

The effect of serum cholesterol level on coronary heart disease and mortality was assessed for middle aged diabetic men in a prospective population study.[1] Logistic regression was used to test the linear trend over quintiles, yielding estimated $p<0.02$. However the validity of the estimated Type I error rate is called into question because the minimum expectation for chi-square was violated. UniODA was applied to these data and identified a model yielding moderate classification accuracy (ESS=29.9, exact $p<0.0006$), which was stable in jackknife validity analysis.

In Table 1 the expected value for the cell having two entries is 4.71, which is less than the minimum expectation which is recommended for this design.[2]

Table 1: Number of diabetic men developing coronary heart disease or dying during follow-up, by quintiles of serum cholesterol[1]

| Serum Cholesterol | Heart Disease | NO Heart Disease |
|---|---|---|
| ≤ 5.5 mmol/l | 3 | 53 |
| 5.6-6.1 mmol/l | 2 | 34 |
| 6.2-6.6 mmol/l | 6 | 33 |
| 6.7-7.3 mmol/l | 5 | 48 |
| >7.3 mmol/l | 15 | 38 |

Note: Tabled are frequency counts.

Because it was hypothesized that serum cholesterol level is positively predictive of the development of heart disease and mortality, a directional (confirmatory) UniODA analysis was conducted on these data.[3,4] The UniODA model was: if serum cholesterol≤7.3 mmol/l then predict NO heart disease; otherwise predict positive for heart disease. In this model only the highest quintile of serum cholesterol level was positively predictive of disease and mortality. The model correctly classified 168 of 206 (82%) men without heart disease, and 15 of 31 (48%) men who were positive for heart disease. The model was correct 91% of the time it predicted no disease, and 28% of the time that heart disease was predicted to occur. The model performance was stable in jackknife analysis,

2155-0182/10/$3.00

which is thus expected to cross-generalize if used to classify an independent random sample.[3]

It has been noted elsewhere[3] but is worth repeating here, that making arbitrary splines on ordinal measures can mask actual effects or can create artificial effects, and therefore using each subject's *actual data* is recommended.

## References

[1]Rosengren A, Welin L, Tsipogianni A, Wilhelmsen L (1989). Impact of cardiovascular risk factors on coronary heart disease and mortality among middle aged diabetic men: A general population study. *British Medical Journal, 299*, 1127-1131.

[2]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.

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

[4]UniODA analysis was accomplished using the following MegaODA[5-7] code (commands are indicated in red). Heart *disease* status was dummy-coded as 0=negative, 1=positive; *serum* cholesterol level was indicated as quintile (1=lowest, 5=highest).[8]

```
open data;
output coronary.out;
vars disease serum;
data;
1 1 (repeated 3 times)
1 2 (repeated 2 times)
1 3 (repeated 6 times)
1 4 (repeated 5 times)
1 5 (repeated 15 times)
0 1 (repeated 53 times)
0 2 (repeated 34 times)
0 3 (repeated 33 times)
0 4 (repeated 48 times)
0 5 (repeated 38 times)
end;
class disease;
attr serum;
dir < 0 1;
mcarlo iter 25000;
loo;
go;
```

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

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

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

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

## Author Notes

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

eMail: Journal@OptimalDataAnalysis.com