Optimal Data Analysis
Vol. 2, Release 2 (November 24, 2013), 217-219

2155-0182/10/$3.00

Determining When Annual Crude Mortality Rate Most Recently Began Increasing in North Dakota Counties, I: Backward-Stepping Little Jiffy

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

Recent research tested the hypothesis that the annual crude mortality rate (ACMR) was higher after versus before 1998 in counties of North Dakota, due to increased exposure of the population to environmental toxins and hazards beginning approximately at that time.[1] This hypothesis was confirmed with experimentwise $p<0.05$ for 16 counties. This article investigates the ACMR time series for each of these counties using a backward-stepping little jiffy UniODA analysis to ascertain precisely when ACMR began to increase. As hypothesized the ACMR began increasing in Bowman and Kidder counties precisely in 1998. Consistent with the *a priori* hypothesis the initial (and presently sustained) increase in ACMR occurred in McLean county in 1997, and in Foster county in 1996. Significant sustained increases in ACMR initially began in Stark county in 1993, and in Burleigh county in 1988. The UniODA models identified hypothesized, recent, powerful, sustained, statistically significant increases in ACMR.

In initial research in this area ACMR in 1998 through 2005 (the latest year for which data are available) was compared with ACMR from the rest of the annual series, which began in 1937. In contrast, in the backward-stepping little jiffy (LJ) analysis, the number of observations (data points) *after* the class threshold and the number *before* the class threshold are held equal.[2] In this application eight data points were available after environmental challenge began (1998-2005), so a total of 16 data points were used in each LJ analysis. With comparatively few data points available for each LJ analysis presently, this is a conservative analytic methodology.[3,4]

Of the total of 16 counties which had a statistically significant increase in ACMR using all data in the series, six also had hypothesized statistically significant increases in ACMR as assessed using the LJ procedure. For Bowman and Kidder counties the *a priori* hypothesis was perfectly supported. Results of the backward-stepping LJ are summarized in Table 1.

2155-0182/10/$3.00

Table 1: Summary of Confirmatory Backward-Stepping Little Jiffy UniODA Analyses

Years Compared	Bowman	Kidder	McLean	Foster	Stark	Burleigh
1990-1997 *vs.* 1998-2005	75.0, 80.0	62.5, 63.5	62.5, 72.7	87.5, 88.9	75.0, 75.0	75.0, 80.0
1989-1996 *vs.* 1997-2004			75.0, 75.0	75.0, 75.0	75.0, 80.0	75.0, 80.0
1988-1995 *vs.* 1996-2003				75.0, 75.0	87.5, 88.9	75.0, 80.0
1987-1994 *vs.* 1995-2002					75.0, 75.0	87.5, 88.9
1986-1993 *vs.* 1994-2001					75.0, 75.0	100, 100
1985-1992 *vs.* 1993-2000					75.0, 80.0	100, 100
1984-1991 *vs.* 1992-1999						100, 100
1983-1990 *vs.* 1991-1998						87.5, 88.9
1982-1989 *vs.* 1990-1997						87.5, 88.9
1981-1988 *vs.* 1989-1996						100, 100
1980-1987 *vs.* 1988-1995						87.5, 88.9

The presence of an entry in Table 1 indicates a statistically significant UniODA model ($p<0.05$, experimentwise) was identified for the indicated county for the indicated comparison, and the absence of an entry indicates no statistically significant model was obtained. The left entry in each cell of Table 1 is ESS (an index of accuracy) for the UniODA model, and the right entry is ESP (an index of predictive value) for the model: on both indices, 0 is the performance expected by chance, and 100 is perfect, errorless classification.[5]

The hypothesized effect occurred exactly in 1998 for Bowman county and the effect was very strong.[6] The hypothesized effect likewise occurred exactly in 1998 for Kidder county, and while the effect was mildly weaker, it still was a relatively strong effect.[6]

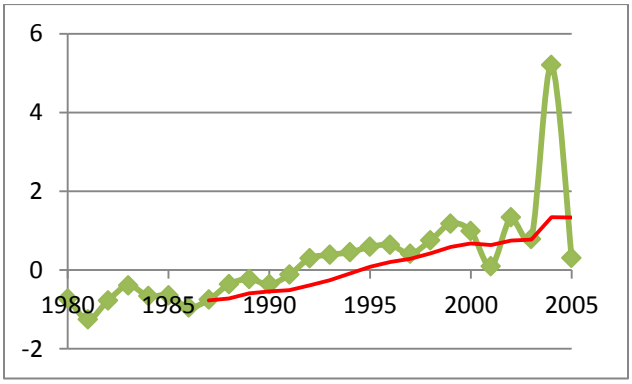
For McLean county the hypothesized effect initially occurred using 1997 as the class threshold, and the model was very strong. The effect was sustained in 1998, and the effect was relatively strong.

For Foster county the hypothesized effect initially occurred using 1996 as the class threshold, and the model was very strong. The effect remained very strong in 1997, and it achieved its strongest ESS in 1998.

For Stark county the hypothesized effect initially occurred in 1993—five years earlier than hypothesized, and the UniODA model was very strong. The effect has remained very strong since that time until today.

The ACMR series for Burleigh county had the most consistent sustained significant growth starting in 1988—a decade earlier than hypothesized, and in every year the increase in ACMR was a very strong effect. The UniODA model achieved perfect classification for three consecutive comparisons. Figure 1 shows the ipsative ACMR series for Burleigh county over all of the years compared in Table 1 (the 8-year moving average is indicated in red).

Figure 1: Ipsatively Standardized ACMR for Burleigh County, 1980-2005



2155-0182/10/$3.00

The enormous z-score reported in 2004 is clearly a statistical outlier. It is widely known that such extreme scores are problematic for parametric analyses because of their influence on mean and variance, and this data point would thus be omitted or transformed in an attempt to mitigate the associated violation of fundamental assumptions underlying parametric methods.[7,8] Traditional problems such as non-normality, exaggerated variance, skewness or kurtosis, and of course statistical outliers, are not a problem for exact univariate[6] or multivarate[9,10] optimal (maximum-accuracy)methods.

References

[1]Yarnold PR (2013). Statistically significant increases in crude mortality rate of North Dakota counties occurring after massive environmental usage of toxic chemicals and biocides began there in1998: An optimal *static* statistical map. *Optimal Data Analysis*, 2, 98-105.

[2]Yarnold PR (2013). The most recent, earliest, and *K*th significant changes in an ordered series: Traveling backwards in time to assess when annual crude mortality rate most recently began increasing in McLean County, North Dakota. *Optimal Data Analysis*, 2, 143-147.

[3]Soltysik RC, Yarnold PR (2013). Statistical power of optimal discrimination with one attribute and two classes: One-tailed hypotheses. *Optimal Data Analysis, 2*, 26-30.

[4]Yarnold PR (2013). UniODA and small samples. *Optimal Data Analysis, 2*, 71.

[5]Yarnold PR (2013). Standards for reporting UniODA findings expanded to include ESP and all possible aggregated confusion tables. *Optimal Data Analysis*, 2, 106-119.

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

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

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

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

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

Author Notes

ODA Blog: odajournal.com