Final Exam B7932 Fall 2008

Type I error (the p_value) on the improvement in fit with a Generalized Linear Model (GzLM) is calculated from the chisquare distribution. According to Feller (1968) the chisquare distribution assumes homogeneous, normal, and independent deviations centered on zero.

In this course we have used residual plots (deviance residuals vs fits, normal probability plots) to evaluate assumptions of linearity (for a regression line) and assumptions for using the chisquare distribution. We have looked only cursorily at whether the residuals are overdispersed, as measured by the ratio of the deviance to degrees of freedom (but see list of articles below). If over dispersion occurs, ignoring it will result in underestimating the standard errors of the parameter estimates, which may lead to incorrect conclusions.

1. Using available texts, summarize recommended practice for graphical and other evaluation of assumptions for computing Type I error when using the GzLM. A partial list of texts appears below.

2. For each data set that you analyze, identify and justify the best first choice of error distribution and link function, then assess the assumptions for this first choice. Where a revised choice of error distribution and/or link function is warranted, justify the choice and then evaluate the assumptions for the revised choice (*e.g*. quasi poisson for overdispersed residuals). Where a second revision in choice is warranted, justify the choice and assess the assumptions again.

3. Check your conclusions by conducting a permutation test for each successive choice of error distribution/link function. This should include a comparison of the permutation p-value with the chisquare p-value.

4. Based on multiple analyses, summarize the efficacy of the methods in this course (residual versus fit plot and normality plot) and any additional practice that you choose based on recommendations in the published literature on GzLM.

Texts.

Agresti, A. An Introduction to Categorical Data Analysis. Wiley-Interscience.
Cameron, A.C. P.K. Trivedi. 1998. Regression Analysis of Count Data. Cambridge University Press.
Dupont, W.D. 2002. Statistical Modeling for Biomedical Researchers. Cambridge University Press
Feller, W. 1968. An Introduction to Probability Theory and its Applications.
Hilbe, J. M. 2007. Negative Binomial Regression. Cambridge University Press.
Hosmer, D.W., S.Lemeshow. 2000 Applied Logistic Regression. John Wiley.
Dobson, A.J., A.G. Barnett. 2008. An Introduction to Generalized Linear Models (3rd Ed) Chapman and Hall.
Hoffmann, J.P. 2004. Generalized Linear Models: An Applied Approach. Boston: Allyn and Bacon.
Lindsey J.K. 2000 Applying Generalized Linear Models. Springer.
McCullagh, P. & J.A. Nelder. 1989. Generalised linear models. Chapman and Hall.
McCulloch C.E., Searle, S.R. 2001. Generalized, Linear, and Mixed Models. John Wiley.
Montgomery, D.C. Peck, E. A. Vining, G.G. 2006. Introduction to Linear Regression Analysis Wiley-Interscience.
Myers, R.H., D.C. Montgomery, G.G. Vining. 2002. Generalized Linear Models. Wiley-Interscience.
Thyregod, P., H. Madsen. 2009. Introduction to General and Generalized Linear Models. Chapman and Hall.

Articles.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. Journal of the Royal Statistical Society C33, 38-44
Breslow N. E. (1996). Generalized linear models: Checking assumptions and strengthening conclusions. Statistica Applicata 8, 23_41.
Breslow, N. E., and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. Journal of the American Statistical Association, 88, 9_25
Cook, R. D., and Weisberg, S. (1994). ARES plots for generalized linear models. Computational Statistics and Data Analysis, 17, 303_315.
Cox, D. R. (1983). Some remarks on overdispersion. Biometrika 70, 269-274
Dean, C., and Lawless, J. F. (1989). Tests for detecting overdispersion in Poisson regression models. Journal of the American Statistical Association, 84, 467-472
Pierce, D. A., and Schafer, D. W. (1986). Residuals in generalized linear models. Journal of the American Statistical Association. 81
Smith, P. J., and Heitjan, D. F. (1993). Testing and adjusting for departures from nominal dispersion in generalized linear models. Applied Statistics 42: 31-41