Selected References

Hosmer, David V. & Lemeshow, Stanley (2000) *Applied Logistic Regression*, 2nd ed., Wiley. [Chapters 4 and 5 served as the primary reference for this talk.]

Outliers and Diagnostics and Goodness of Fit

Andrews, DF & Pregibon, D (1978) Finding the outliers that matter. *J of the Royal Statistical Society, Series B 40*, 85-93. [details and discussion on masking]

Cook, RD & Weisberg, S (1999) *Applied Regression including Computing and Graphics*. Wiley.

Fowlkes, EB (1987) Some diagnostics for binary regression via smoothing. *Biometrika, 74*, 503-515.

Hosmer, DV, Taber, S & Lemeshow, S (1991) The importance of assessing the fit of logistic regression models: A case study. *Amer J Public Health 81*, 1630-1635. [good description of diagnostics fundamentals – with generic example]

Jennings, DE (1986a) Judging inference adequacy in logistic regression. *J of the Amer Statistical Association 81*, 471-476.

Jennings, DE (1986b) Outliers and residual distributions in logistic regression. *J of the Amer Statistical Association 81*, 987-990.

Landwehr, JM Pregibon, D & Shoemaker, AC (1984) Graphical methods for assessing logistic regression models. *J of the American Statistical Association 79*, 61-83. [presents methods based on grouping and smoothing – with comments by Rubin and others and rejoinder]

Menard, Scott (2000) Coefficients of determination for multiple logistic regression analysis. *The American Statistician 54*, 17-24.

Pregibon, D (1981) Logistic regression diagnostics. *Annals of Statistics 9*, 705-724. [seminal theoretical work in the field – lays out fundamentals.]

Model and Variable Selection Issues

Collett, D & Stepniewska, K (1999) Some practical issues in binary data analysis. *Statistics in Medicine 18*, 2209-2221. [deeper discussion of variable selection]

Harrel, FE, Lee, KL, & Mark, DB (1996) Tutorial in biostatistics: Multivariable prognostic models: Issues in developing models, evaluating assumptions and measuring and reducing errors. *Statistics in Medicine 15*, 361-387. [models look better in development than validation]

Hosmer, DV, Jovanovic, BD & Lemeshow, S (1989) Best subsets logistic regression. *Biometrics 45*, 1265-1270 [simplifying application of Lawless/Singhal ISMOD results for this case]

Jovanovic, BD and Hosmer, DV (1997) A simulation of the performance of C_p in model selection for logistic and Poisson regression. *Computational Statistics & Data Analysis 23*, 373-379. [C_p appears to be conservative relative to LR and Score tests]

Mallows, CL (1973) Some comments on C_p. *Technometrics 15*, 661-675 [seminal paper on C_p]

Sommer, S & Huggins, RM (1996) Variables selection using the Wald Test and a Robust C_p. *Applied Statistics 45*, 15-29. [new criterion proposed with a nice application]

A Few Additional Recent Medical & Health Care Application Articles

Green, SM et al. (2000) Predictors of adverse events with intramuscular ketamine sedation in children. *Annals of Emergency Medicine 35*, 35-42. [validation using bootstrap resampling]

Hin, LY et al. (1999) Dichotomization of continuous measurements using generalized additive modeling – Application in predicting intrapartum caesarean delivery. *Statistics in Medicine 18*, 1101-1110. [motivation in the prediction model context for graphical use of diagnostics]

Nattinger, AB et al. (1996) The effect of legislative requirements on the use of breast-conserving surgery. *N Engl J Med 335*, 1035-1040. [residuals used to assess deviation from time trend]

Saez, M et al. (1999) A GEE Moving Average Analysis of the Relationship between Air Pollution and Mortality for Asthma in Barcelona, Spain *Statistics in Medicine 18,* 2077-2086. [partial residual plots to assess non-linearity, deviance residuals modeled using complex MA]

Thomas E. Love, Ph. D.   TEL3@po.cwru.edu (216) 778-1265