Data 621 Reading/Video Notes

Boxplot

halfnorm

Regsubset- finds best models. Can specify best, forward or backward modelling.

Poly – used to get polynomials for linear modes

Step- find best models, compare AIC and BIC values

LM- Linear Model, used for data with normally distributed errors and constant variance.

GLM-Generlized linear model. Used for nonnormal error and non-constant variance. Need a exponential family model for response, systematic linear predictor and a link function that connects means of the response to the linear predictor. Transforming mean log of mean/1-mean for coefficient and covariance. Reverse with e to the values. Also see Quasi GLMs, more flexible models.

GLS – Generalized Least Square. Used for data with correlated errors , spatial/temporal patters/trends.

Pairs(data,col=binary response)

Train = year<2005

Glm—use subset = train; Predict –use newdata = data[!train]

Logit/Probit- used for binary data, can show marginal effects

Logit = exp(z)/1+exp(z)

Probit= integral of z = infinity density function(df)(u)du

Goodness of fit – percent of correctly predicted values.

OLS-Ordinary Least Square. All points weighted equally , including influential points., changing the regression equation.

WLS – Weighted Least Square. Influential points are NOT weighted equally. Huber Weighting system is a form of WLS. of WLS.