University of Colorado Department of Civil, Environmental and Architectural Engineering CVEN 5454 Quantitative Methods

Homework #5

Due April 28th, 2005

Topics: Linear Regression, Logistic Regression, Nonparametric Regression and Design of Experiments – Chapters 11,12 and 13, Helsel and Hirsch – Chapter 15

1. Stream temperature data on the Truckee river at Reno is available at http://civil.colorado.edu/~balajir/CVEN5454/spr-05/streamtemp.dat A diagram of the study region is in my publication #28 (http://bechtel.colorado.edu/~balajir/mywebpage/publications.html)

As a water manager on this river concerned about the stream temperature and consequently, the water quality - you would like to come up with a best linear model that captures the variability in the stream temperature which can be used in hydrologic decision making. This would entail

- (i) Identifying the "best" subset of predictors to be included in the model (use Cp, AIC and adjusted R2 criteria and describe the differences)
- (ii) Performing model diagnostics
- (iii) Checking the model performance on new data. Which means you withhold some of the data and predict them using the fitted model.
- (iv) Plot the confidence interval of the fitted line and the predicted interval for the predicted points.
- 2.Problem 7. Page 458
- 3. Problem 15.1 (Helsel and Hirsch Book) Logistic Regression
- 4. Problem 11, Page 485 Single Factor design
- 5. Compute the mutual information between the response variable (stream temperature) and the predictors. Compare them with the correlation.

Bonus Question

- 6.You have become aware that local regression (nonparametric regression) schemes might be an attractive alternative in capturing any arbitrary dependence structure. To this end, use Local Regression Method (LOCFIT) on the best set of predictors you identified in 1 to get the fit and the confidence intervals. Comment on the differences you observe and which one you would recommend based on your analyses.
- 7. Problem 9 Page 501 (Complete Block Design)