STA 290 STATISTICAL LABORATORY **HOMEWORK 4**
Fall Semester 2002

Solutions should be written up using Latex; include graphs only where required.

Exercise (1)

Find the expected values, variances and covariance of the ordinary least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ (conditional on σ^2) under the assumption that Y_i are independent normal random variables with mean $\beta_0 + X_i\beta_1$ and variance σ^2.

$$\begin{aligned} \hat{\beta}_1 &= S_{XY}/S_{XX} \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1\bar{X} \end{aligned}$$

Exercise (2)

EDA: The `fish` data under the Datasets link contains information on mercury concentrations in (ppm) for large mouth bass filets from two North Carolina rivers. The weight (grams) and length (cm) of each fish were recorded as well as an indicator variable for the river (0=Lumber, 1 = Wacamaw) and a number for the station on that river (0, 1, ... 15). Each fish caught corresponds to a single row of the file. The goal is to develop a predictive model of mercury concentrations based on the other covariates. Note: S-Plus users will need to delete the lines with comments at the top of the data file before reading it in to S-Plus.

(a) Based on your previous exploratory analysis of the data, find a "good" regression model to predict mercury concentrations using the weight and length variables. Check for normality (Q-Q plot) and if there are any outliers/influential points. Refit if necessary until assumptions for regression appear to be met. Provide summaries of your model.

(b) Are both weight and length necessary or can a model with just one of them do "equally" well? (answer via a statistical hypothesis test) Provide summaries for the best single predictor model.

(c) For the best single predictor model, construct a plot of the predictor versus the posterior mean PCB concentration and add 95% credible intervals for the mean PCB concentration and 95% prediction intervals (be sure to transform back to the **original units** if you used any transformations).

(d) For the single predictor model, find the fish size such that the posterior probability of mercury levels exceeding 1 ppm is less than 0.95.

(e) In light of your new analysis, what recommendations can you make for these stations/rivers? Write up a one page summary; include any figures and calculations in an appendix. Provide enough details so that someone familiar with statistical methods would be able to reproduce your work.