Computer Lab Session #3

Load the file MASSCOLL.MTW from the Minitab example data sets, Student 12 subdirectory. This contains several variables measured on 56 four-year colleges in Massachusetts for 1995. We will concentrated on y="%WhoGrad" (percentage of students graduating), x₁="CSAT" (median combined SAT math and verbal score) and x₂="SFRatio" (student faculty ratio).

As usual, produce numerical and graphical summaries for y, x₁ and x₂ to get a sense of the data in terms of the individual variables.

Fit two separate simple regression models for y="%WhoGrad" on x₁="CSAT", and y="%WhoGrad" on x₂="SFRatio". Use Stat > Regression > Regression. Remember to also produce the fitted line plots. Use Stat > Regression > Fitted Line Plot ("linear")

• What percent of the variability in y="%WhoGrad" is explained by x₁="CSAT"? Compute the correlation between the two using Stat > Basic Statistics > Correlation. Is the determination coefficient from the regression output equal to the square of the correlation coefficient?

• Address the same questions for y="%WhoGrad" and x₂="SFRatio".

• Which predictor does a better job at explaining the variability in percentage of graduating students across MA colleges?

• Do the estimated regression slopes suggest positive or negative relationships? Are they significant? (i.e. can one reject the hypothesis that the slope is 0 in each of the regressions). Identify the relevant test statistics and p-values in the regression output, comment on the p-values.

• Interpret the differences in coefficients of determination and slope signs. Do these make sense? What are they telling you about the determinants of percentage of graduating students?

• Construct the 95% confidence interval for the slopes in each regression. You can find the pivot (point estimate of the slope) and the standard error in the regression output. To determine the multiplier, use Calc > Probability distributions > t. You have to be careful with the degrees of freedom here. Look at the regression output; in both regressions, not all cases/units (i.e. colleges) were used, because the data contains missing values (these are indicated as asterisks in Minitab). Consequently, in both regressions the actual n is not 56.

• Practice the language required in interpreting the confidence intervals.

• As a pointer for the next lectures, for each of the two regressions go back to Stat > Regression > Fitted Line Plot ("linear"). In “Options”, check “display confidence intervals” and “display prediction intervals”, at 95% confidence level. The resulting fitted line plots will now have bands about the regression lines. What do you think these bands are indicating? Why do you think the prediction bands are larger than the confidence bands? Why do you think the bands “curve” outward?