**§3.1.8 Initial Multiple Linear Regression Exercise**

A strategy for multiple linear regression analysis:

i. State the model.

ii. Perform a descriptive statistical analysis for Y and X1, X2,…,Xk etc.

iii. Plot Y vs Xi (i=1,…,k). Inspect plots for linear pattern1.

iv. Calculate and characterize rY|Xi and interpret2.

v. Fit the model.

vi. Perform a complete regression diagnostics workup3.

vii. Perform the overall F-test. If not significant, report this finding and stop working with the current model --it is not significant; do not interpret the slope estimates (the failure to reject the null means that you have not been able to show that any of the slopes are anything other than zero in value).

If significant, proceed with additional inferences/interpretations (below).

viii. Present estimate of the model. Perform partial tests and/or determine confidence intervals for regression slopes, and present and interpret as appropriate.

ix. Proceed with additional model inferences, if appropriate and wanted: test on Y-intercept, prediction intervals, confidence intervals on μY.

1 In multiple linear regression, plots known as ‘partial plots’ will be more appropriate than

simple scatterplots. We will revise the solution to our exercise once we have learned about these plots. (SAS will produce these plots for us)

2 In multiple linear regression, ‘partial’ correlations are more appropriate than simple correlations; we will revise our strategy. We will revise the solution to our exercise once we have learned about these plots. (SAS will produce these correlations for us)

3 A ‘regression diagnostics workup’ includes: (i) assumption checking, (ii) outlier detection, and (iii) collinearity detection. We will revise the solution to our exercise once we have learned about the relevant techniques.

Exercise: Data were collected on birthweight (g, variable bwt), mothers age (years, variable age), weight (lbs, variable wt), smoking status (0 = did not smoke during pregnancy, 1 = smoked during pregnancy), hypertension (0 = none, 1 = history of hypertension). Run a regression with birthweight as the dependent variable.

SAS code:

\*I will leave the details of importing the excel file, examining the contents of the imported data set, initially summarizing/describing the data for every variable to you! These are tasks you have performed many times in BIOS 500 and 591P;

\*Here's the correlation, scatterplot and multiple linear regression code

that you need...;

**PROC** **CORR** DATA=WORK.bwt PLOTS=MATRIX; \*I imported the data into this data set;

VAR age wt smoke ht;

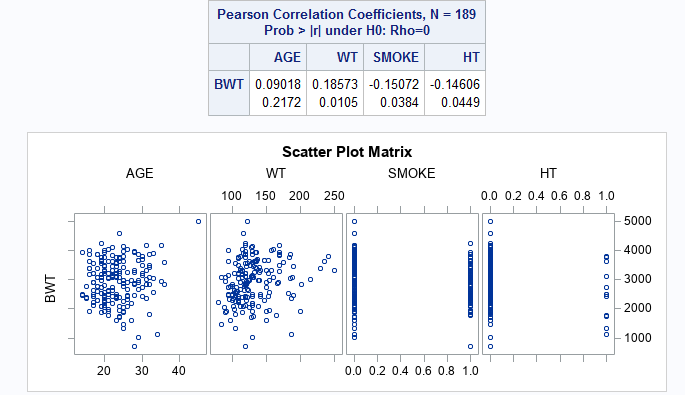
WITH bwt;

**RUN**;

**PROC** **REG** DATA=WORK.bwt;

MODEL bwt = age wt smoke ht;

**RUN**;

(Edited) SAS output: