**STAT 462 – Applied Regression Analysis**

**Fall 2017, Lab 4 Solutions**

Prepare a short report with relevant output, your comments, and answers to the questions (this does not need to be exhaustive or polished, but should contain enough to show that you completed all tasks and analyses).

Submit the report at the end of the lab session.

Consider the dataset *bears.txt*.

This contains several variables measured on n=141 “bear capturing” occasions, with the following variables:

*ID:* Identification number

*Age:* Bear's age, in months

*Month:* Month when the measurement was made. Sex. 1 = male 2 = female

*Head.L:* Length of the head, in inches

*Head.W:* Width of the head, in inches

*Neck.G:* Girth (distance around) the neck, in inches

*Length:* Body length, in inches

*Chest.G:* Girth (distance around) the chest, in inches

*Weight:* Weight of the bear, in pounds

*Obs.No:* Observation number for this bear. For example, the bear with ID=41 (Bertha) was measured on four occasions. The value of Obs.No goes from 1 to 4 for these observations

*Name:* The names of the bears given to them by the researchers.

The observations are not independent, because the same bear may have been captured more than once (see variables “Name”, “ID” and “Obs.No”).

* For each bear, select only the first observation, so that the new dataset will contain only independent observations. Why is that important for linear regression? How many bears do we have in the dataset?

Consider the variables y=“Weight”, x1=“Chest.G” and x2=“Head.W”.

Fit two separate simple regression models for y=“Weight” on x1=“Chest.G”, and y=“Weight” on x2=“Head.W” (you can use the *lm* function or the equations).

* Do the estimated regression slopes suggest positive or negative relationships? Is there a meaningful interpretation for the regression intercepts?
* Using the equation, estimate the variance sigma2 of the error term for the two models (you can check the result with the *summary* function, but you need to compute it using the equation).
* Using the equation, compute the coefficient of determination R2 for both regressions (you can check the result with the *summary* function, but you need to compute it using the equation). What is their interpretation?
* Between x1=“Chest.G” and x2=“Head.W”, which appears to be the best predictor for y=“Weight”? (Address this comparing the coefficients of determination R2 of the two regressions).

|  |  |
| --- | --- |
| |  | | --- | |  | |
| |  | | --- | |  | |

The following part will NOT be considered in grading the lab report.

Fit a multiple linear regression model with predictors x1=“Chest.G” and x2=“Head.W”.

* Using the equation, estimate the variance sigma2 of the error term for the new model (you can check the result with the *summary* function, but you need to compute it using the equation).
* Using the equation, compute the coefficient of determination R2 for the new regression (you can check the result with the *summary* function, but you need to compute it using the equation). What is its interpretation?
* Do you think this model is better that the one with only x1? Why?