**STAT 462 – Applied Regression Analysis**

**Fall 2017, Lab 6**

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 again the dataset *bears.txt* used in Lab4-5.

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.

As you did in Lab4-5, consider only the first observation for each bear (bears\_indep=bears[bears$Obs.No==1,]).

Consider a multiple linear regression model with response y=“Weight” and predictors x1=“Head.L”, x2=“Head.W”, x3=“Neck.G”, x4=“Length” and x5=“Chest.G”.

* After fitting the model, perform a t test for each single beta\_i, and a global F test for all predictors to answer to the following questions (you can use the *lm* function and the *summary* function).
* Is the model significant?
* Which of the predictors are NOT significant in the model at alpha=0.05, if you consider them individually?

> lm.bears=lm(Weight~Head.L+Head.W+Neck.G+Length+Chest.G,data=bears)

> summary(lm.bears)

Call:

lm(formula = Weight ~ Head.L + Head.W + Neck.G + Length + Chest.G,

data = bears)

Residuals:

Min 1Q Median 3Q Max

-59.457 -17.969 -2.059 14.432 99.239

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -258.3771 20.8837 -12.372 < 2e-16 \*\*\*

Head.L -7.5230 3.3596 -2.239 0.0275 \*

Head.W 0.3087 3.3965 0.091 0.9278

Neck.G 8.5812 1.7639 4.865 4.65e-06 \*\*\*

Length 1.3305 0.7425 1.792 0.0764 .

Chest.G 7.8844 1.0190 7.738 1.19e-11 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 27.24 on 93 degrees of freedom

Multiple R-squared: 0.9456, Adjusted R-squared: 0.9427

F-statistic: 323.5 on 5 and 93 DF, p-value: < 2.2e-16

**For the F distribution with F = 323.5 and its p-value = 2.2 × 10^(-16)**

**Since p-value < α =0.05, we reject the null hypothesis test (H0: β1 = β2 = β3 = β4 = β5 =0).**

**Thus, the model is significant.**

**For each beta\_i, we observe from the chart that β for Head Width and Length have test statistics of 0.091 and 1.792 with p-values of 0.9278 and 0.0764.**

**Since the p-value of those two variables are both greater than α =0.05, they are not significant**.

* Consider the subset of predictors that are not significant at alpha=0.05. Perform a test to see if they are significant or not, when you consider them simultaneously; i.e. test the hypotheses H0: betaq = … = betap-1 = 0 vs H1: at least one != 0. Provide:
* Distribution of the test statistic under H0 (type and parameters of the distribution)
* RSS of the full model and RSS of the reduced model, computed using the equation (you can check the result with the *anova* function, but you need to compute it using the equation)
* Value of the test statistics, computed using the equation (you can check the result with the *anova* function, but you need to compute it using the equation)
* P-value, computed using the equation (you can check the result with the *anova* function, but you need to compute it using the equation)
* Interpretation of the result.

> lm\_full=lm(Weight~Head.L+Head.W+Neck.G+Length+Chest.G,data=bears)

> lm\_red=lm(Weight~Head.L+Neck.G+Chest.G,data=bears)

> RSS\_full=sum(lm\_full$residuals^2)

> RSS\_full

[1] 69003.64

> RSS\_red=sum(lm\_red$residuals^2)

> RSS\_red

[1] 71386.92

> n=nrow(bears)

> n

[1] 99

> p=length(lm\_full$coefficients)

> p

[1] 6

> q=length(lm\_red$coefficients)

> q

[1] 4

> F=((RSS\_red-RSS\_full)/(p-q))/(RSS\_full/(n-p))

> F

[1] 1.606035

> pvalue=pf(F,p-q,n-p,lower.tail=FALSE)

> pvalue

[1] 0.2061971

**Under null hypothesis, test statistic is F distribution with p = 6, q = 4, df1=2 and df2=93.**

**RSS of full model is 69003.64**

**RSS of reduced model is 71386.92**

**Test Statistic is F = 1.606035**

**p-value is 0.2061971**

**Since the p-value > α = 0.05, we accept null hypothesis.**

**Thus, Head Width and Length are not significant.**

R code:

setwd("//udrive.win.psu.edu/Users/j/q/jql5883/Desktop/math462")

getwd()

bears=read.csv("bears.txt", header=T, sep="")

bears=bears[bears$Obs.No==1,]

head(bears)

attach(bears)

lm.bears=lm(Weight~Head.L+Head.W+Neck.G+Length+Chest.G,data=bears)

summary(lm.bears)

lm\_full=lm(Weight~Head.L+Head.W+Neck.G+Length+Chest.G,data=bears)

lm\_red=lm(Weight~Head.L+Neck.G+Chest.G,data=bears)

RSS\_full=sum(lm\_full$residuals^2)

RSS\_full

RSS\_red=sum(lm\_red$residuals^2)

RSS\_red

n=nrow(bears)

n

p=length(lm\_full$coefficients)

p

q=length(lm\_red$coefficients)

q

F=((RSS\_red-RSS\_full)/(p-q))/(RSS\_full/(n-p))

F

pvalue=pf(F,p-q,n-p,lower.tail=FALSE)

pvalue

If you still have time, work on the applied part of the homework!