Lab 2: Multiple Regression

Sepetember 25, 2017

### Names: Daniel Meza, Adolfo Huerta

##### What each person contributed to the lab:Daniel Meza did problem one and Adolfo

##### Huerta problem two. We then checked each others worked and discussed about it.

**Please make sure to show all R code and output after each question so that I can see your work.** Write a sentence for each numerical value produced describing its meaning in context with the proper units.

## Problem 1

A medical study was conducted to study the relationship between systolic blood pressure and the explanatory variables, weight (kg) and age (days) for infants. The data for 25 infants is provided in infants.csv. The researchers decide to consider the general linear model:

### 1.a.

State the statistical analysis research question for your Regression Analysis.

The researchers are interested in seeing if the variables age and weight can predict the systolic blood pressure well.

### 1.b.

Name the response variable of interest below and state the type (quantitative, qualitative).

The response variable is systolic blood pressure and it is quantitative.

### 1.c.

Name the explanatory variables of interest below and state the types (quantitative, qualitative).

Weight and age are both explanatory quantitative variables.

### 1.d.

Describe the data collection technique: is it experimental or observational?

This is an observational study since they cannot control the childrens weight.Also, they did not specify whether they tracked the children or not.

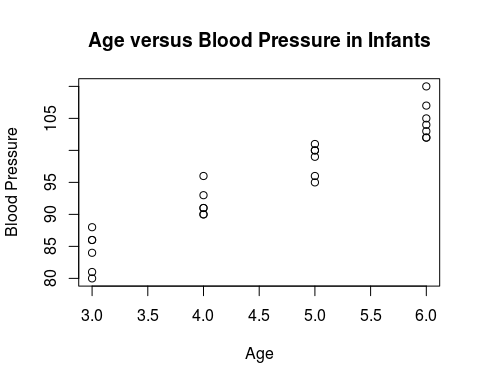
### 1.e.

Exploratory data analysis: provide the summary statistics (mean and standard deviation) for each explanatory variable and the response. In addition, provide a scatterplot of each explanatory variable versus the response. Graphs should have proper x and y axis labels. Discuss the observed relationships and possible implications on the structure of the regression model (first order model or higher order models).

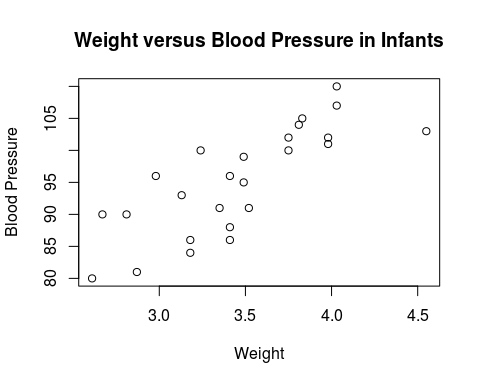
#Provide R code and output/graphs   
  
library(readr)  
infants <- read\_csv("~/Documents/infants.csv")

## Parsed with column specification:  
## cols(  
## Infant = col\_integer(),  
## Age = col\_integer(),  
## Weight = col\_double(),  
## SystolicBP = col\_integer()  
## )

View(infants)  
plot(infants$Age,infants$SystolicBP,main = "Age versus Blood Pressure in Infants",xlab="Age",ylab="Blood Pressure")



plot(infants$Weight,infants$SystolicBP,main = "Weight versus Blood Pressure in Infants",xlab="Weight",ylab="Blood Pressure")



x2=infants$Weight  
x1=infants$Age  
x3=infants$SystolicBP  
bl.model<-lm(x3~x1+x2)  
mean(x1)

## [1] 4.56

mean(x2)

## [1] 3.4584

sd(x1)

## [1] 1.157584

sd(x2)

## [1] 0.478406

mean(x3)

## [1] 95.2

sd(x3)

## [1] 8.301606

The mean age in the infants data set is 4.56 and the standard deviation is 1.157 for age.The mean weight of the infants was 3.4558 kg with a standard deviation of .04784.The mean blood pressure for the infants was 95.2 and the standard deviation was 8.301. We can observe a good relationship between wieght and blood pressure. There is also a sign of a possitve correlation with age and blood pressure. If age and weight are not too correlated then there could be a possibility for a high order linear model.

### 1.f.

Determine the coefficients for each variable in the model and provide an interpretation of each parameter estimate in the context of the problem.

#Provide R code and output/graphs  
  
bl.model<-lm(x3~x1+x2)  
coef(bl.model)

## (Intercept) x1 x2   
## 57.264442 5.804097 3.316238

The blood pressure when the age is zero and the wieght is zero is 57.264. Holding the weight constant, a one year increase in age will increase the blood pressure by 5.804.Holding the age constant, a one kg increase in weight will increase the blood pressure by 3.316.

### 1.g.

Test for possible correlations between the explanatory variables and assess the possible impact of collinearity. Provide measures of correlation, scatterplots and VIF assessment.

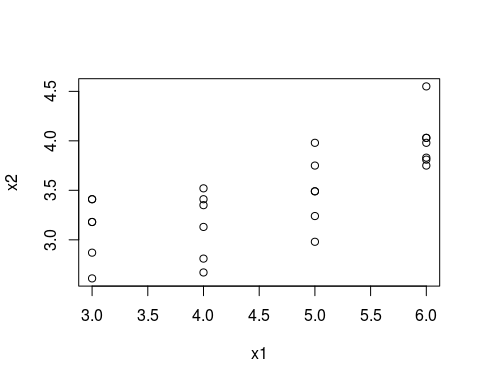
#Provide R code and output/graphs   
  
  
  
cor(infants)

## Infant Age Weight SystolicBP  
## Infant 1.0000000 0.2347532 0.4757208 0.2666474  
## Age 0.2347532 1.0000000 0.7382695 0.9504183  
## Weight 0.4757208 0.7382695 1.0000000 0.7886113  
## SystolicBP 0.2666474 0.9504183 0.7886113 1.0000000

library(car)  
vif(bl.model)

## x1 x2   
## 2.198005 2.198005

plot(x1,x2)



When we look at just the correlation for the variables in the model we can see that weight and age are highly correlated. The VIF does not prove colinearity, since the VIF for both variables is less than 5, so we cannot claim that there exits a worrisome collinearity between the variables.

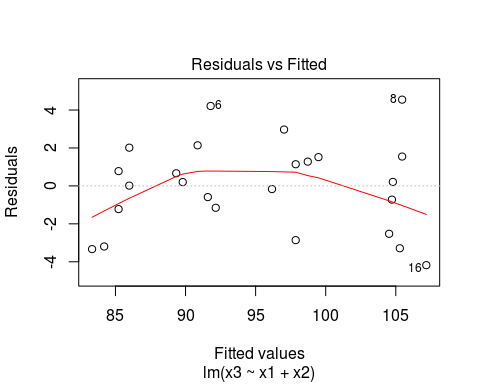
### 1.h.

Conduct appropriate testing for the assumptions of Regression Analysis.

#### 1.h.i.

Did the observations pass the equal variance assumption?

#Provide R code and output/graphs   
plot(bl.model,1)

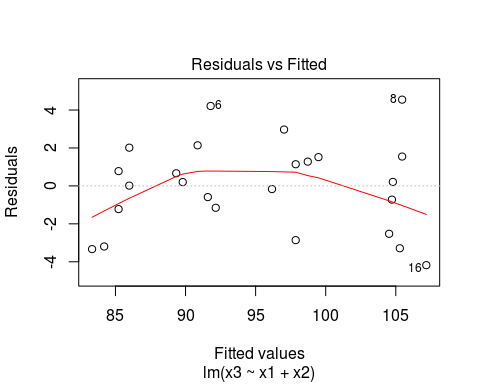


There exists constant variance since the residuals are spread out evenly througout the plot.

#### 1.h.ii.

Did the observations pass the independence assumption?

#Provide R code and output/graphs   
plot(bl.model,1)

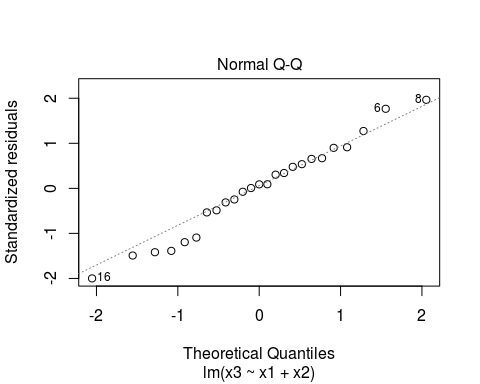


Yes, the assumption is satisfied since the residuals are randomly dispersed in the plot. We cannot observe a specific pattern.

#### 1.h.iii.

Did the observations pass the normality assumption?

plot(bl.model,2)



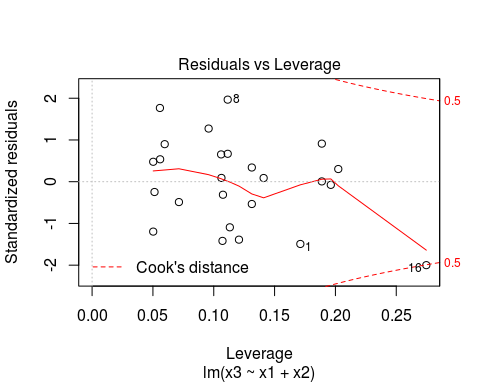
#Provide R code and output/graphs

Yes, the normaility assumption is satisfied since there are only a couple data points that begin to trail of the normal line.

#### 1.h.iv.

Did the observations pass the no outlier assumption?

#Provide R code and output/graphs   
plot(bl.model,5)



No, this assumption can’t be satisified since there is a outlier of concern.

### 1.i.

Now we will conduct the test to determine if the linear relationship is statistically significant by testing the *general* hypothesis for the full model.

#### 1.i.i.

State the *general* hypothesis for the full model (null and alternative) for Multiple Regression Analysis mathematically and in a complete sentence.

The model has no explanantory power of my response.

At least one

Atleast one explanantory variable has a statistically significant relationship with the response.

#### 1.i.ii.

Use R to find the ANOVA test statistics to test the general hypothesis test.

#Provide R code and output/graphs   
summary(bl.model)

##   
## Call:  
## lm(formula = x3 ~ x1 + x2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.1779 -1.2224 0.2005 1.5164 4.5465   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 57.2644 3.7986 15.075 4.44e-13 \*\*\*  
## x1 5.8041 0.6415 9.048 7.22e-09 \*\*\*  
## x2 3.3162 1.5522 2.136 0.044 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.454 on 22 degrees of freedom  
## Multiple R-squared: 0.9199, Adjusted R-squared: 0.9126   
## F-statistic: 126.3 on 2 and 22 DF, p-value: 8.696e-13

anova(bl.model)

## Analysis of Variance Table  
##   
## Response: x3  
## Df Sum Sq Mean Sq F value Pr(>F)   
## x1 1 1494.05 1494.05 248.1303 1.828e-13 \*\*\*  
## x2 1 27.48 27.48 4.5644 0.04402 \*   
## Residuals 22 132.47 6.02   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### 1.i.iii.

Interpret the meaning of the p-value in the context of the situation.

The probability that we observe our test statistics or more extreme is less than .0001.

#### 1.i.iv.

State the conclusion of the hypothesis test in the context of the problem.

Let F-statistic=126.3, p-valeu<.00001. Based on our sample, using a significance level of .05 we will reject the null hypothesis that the model has no explanatory power for the response. We have evidence to support the alternative that atleast one explanatory variable has some statistically significant relationship with the response.

### 1.j.

Now we will conduct the test to determine if each variables is statistically significant by testing the *specific* hypotheses for each variable.

#### 1.j.i.

State the *specific* hypotheses for each variable in the model (null and alternative) for Multiple Regression Analysis mathematically and in a complete sentence.

The variable age has no relationship with power with the response given that all other variables are in the model.

The variable age has a relationship with power with the response given that all other variables are in the model.

The variable weight has no relationship with power with the response given that all other variables are in the model.

The variable weight has a relationship with power with the response given that all other variables are in the model.

#### 1.j.ii.

Use R to find the t-test statistics to test each hypothesis test.

#Provide R code and output/graphs  
summary(bl.model)

##   
## Call:  
## lm(formula = x3 ~ x1 + x2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.1779 -1.2224 0.2005 1.5164 4.5465   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 57.2644 3.7986 15.075 4.44e-13 \*\*\*  
## x1 5.8041 0.6415 9.048 7.22e-09 \*\*\*  
## x2 3.3162 1.5522 2.136 0.044 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.454 on 22 degrees of freedom  
## Multiple R-squared: 0.9199, Adjusted R-squared: 0.9126   
## F-statistic: 126.3 on 2 and 22 DF, p-value: 8.696e-13

#### 1.j.iii.

Interpret the meaning of the p-values in the context of the situation.

The probability that we observe our test statistic of age or more extreme is less than .0001. The probability that we observe our test statistic of weight or more extreme is .0442.

#### 1.j.iv.

State the conclusion of the hypothesis tests in the context of the problem.

let t-statistic= 9.048 and p-value=.0001. Based on our sample, we will reject the null and support the alternative that the variable age has a relationship with the response given that all other variables are in the model.

let t-statistic= 2.136 and p-value=.0442. Based on our sample, we will reject the null and support the alternative that the variable weight has a relationship with the response given that all other variables are in the model.

### 1.k.

Interpret the correct value in the context of the model variables. Assess the strength of the predictive power of the model.

About 91 percent of the variability in blood pressure is explained by the linear relationship with weight and age.

### 1.l.

Do you have any concerns or issues with the formulation of the model and the included variables? Explain.

Most of the assumptions passed except the assumption on outliers. I have a small concern about why the variables weight and age are strongly correlated but, when we checked the VIF the value was small.

## Problem 2

A poultry scientist was studying various dietary additives to increase the rate at which chickens gain weight. One of the potential additives was studied by creating a new diet which consisted of a standard basal diet supplemented with varying amounts of the additive (0, 20, 50, 60, 80, 100 mg). In addition, they looked at the effects of high copper levels in the chick feed by adding either 0 or 400 ppm of cooper to the feed. There were sixty chicks available for the study. Each of the six diets for additives was randomly assigned to 10 chicks. Then 5 of the 10 chicks were randomly assigned to the high level of copper (400ppm) and the remaining 5 received 0ppm copper. At the end of four weeks, the feed efficiency ratio, measured as feed consumed (g) to weight gain (g), was obtained for the sixty chicks and is provided in the data file poultry.csv. The researchers decide to consider the general linear model:

### 2.a.

State the statistical analysis research question for your Regression Analysis.

The statistical analysis research question is wether the various dietary additives in combination with either high level of copper or low level of copper have predictive power for the rate at which chickens gain weight, which is measured as the feed efficiency ratio.

### 2.b.

Name the response variable of interest below and state the type (quantitative, qualitative).

The response variable of interest is the feed efficiency ratio, which is measured as feed consumed (g) to weight gain (g). This is a quantiative type of variable since it is grams over grams.

### 2.c.

Name the explanatory variables of interest below and state the types (quantitative, qualitative).

The explanatory variables of interest are the amount of additive (mg), which is quantitative, and wether it is high copper level (400 ppm) or low copper level (0 ppm), which would be categorical.

### 2.d.

Describe the data collection technique: is it experimental or observational?

The data collection technique is experimental since there are two types of groups given different treatments, in this case in the shape of different diets.

### 2.e.

Exploratory data analysis: provide the summary statistics (mean and standard deviation) for each explanatory variable and the response. In addition, provide a scatterplot of each explanatory variable versus the response. Graphs should have proper x and y axis labels. Discuss the observed relationships and possible implications on the structure of the regression model (first order model or higher order models).

#Provide R code and output/graphs   
  
  
library(readr)  
poultry <- read\_csv("~/Documents/poultry.csv")

## Parsed with column specification:  
## cols(  
## CuLevel = col\_integer(),  
## Additive = col\_integer(),  
## FeedEffRatio = col\_double()  
## )

View(poultry)  
attach(poultry)  
names(poultry)

## [1] "CuLevel" "Additive" "FeedEffRatio"

summary(poultry)

## CuLevel Additive FeedEffRatio   
## Min. : 0 Min. : 0 Min. :1.140   
## 1st Qu.: 0 1st Qu.: 20 1st Qu.:2.192   
## Median :200 Median : 50 Median :2.455   
## Mean :200 Mean : 50 Mean :2.778   
## 3rd Qu.:400 3rd Qu.: 80 3rd Qu.:3.320   
## Max. :400 Max. :100 Max. :5.090

sd(CuLevel)

## [1] 201.6878

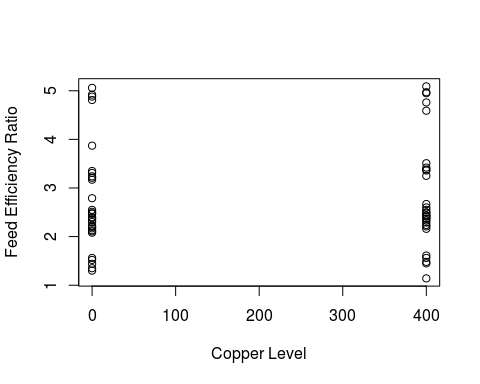
sd(Additive)

## [1] 34.44475

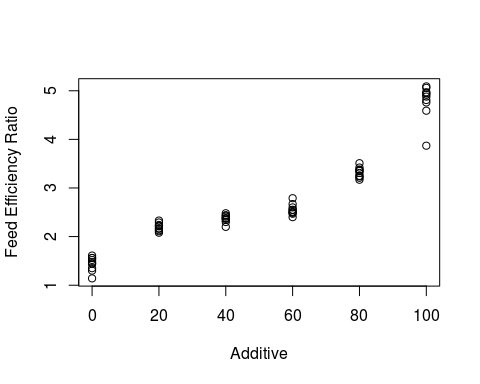
sd(FeedEffRatio)

## [1] 1.077504

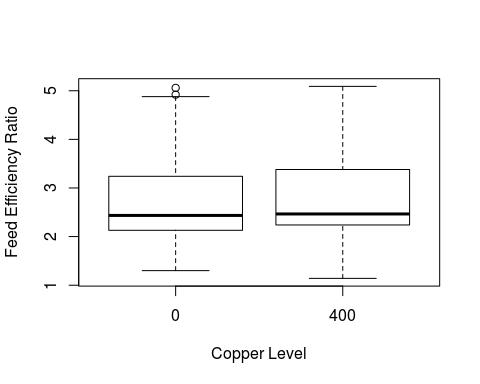
plot(poultry$CuLevel,poultry$FeedEffRatio, xlab="Copper Level", ylab = "Feed Efficiency Ratio")



plot(poultry$Additive,poultry$FeedEffRatio, xlab="Additive", ylab = "Feed Efficiency Ratio")



plot(factor(poultry$CuLevel), poultry$FeedEffRatio, xlab = "Copper Level", ylab = "Feed Efficiency Ratio")



The dataset consists the respone variable being FeedEffRatio and the two explanatory variables being CuLevel and Additive. The CuLevel is wether the chickens diet consisted of low Copper(0ppm) or high Copper(400ppm), which could be viewed as a category so the mean and median of 200 would not be to useful, it has a standard deviation of 201.6878ppm. The diet consisted of a certain amout of additive ranging from 0 to 100 mg, it has a mean of 50mg and a standard deviation of 34.44475mg. The responese variable, the Feed Efficiency Ratio has a mean of 2.778g and a standard deviation of 1.077504g. From the observed scatterplots, it appears that the amount of additive and the Feed Efficiecy Ration have a positive linear relationship. We can also observe that between the two differenct Copper levles and the Feed Efficiency Ratio, there appears to be no relationship, as the feed efficienciy ratio is similar for either group. ###2.f0.  
Determine the coefficients for each variable in the model and provide an interpretation of each parameter in the context of the problem. (Use the model proposed by the researchers, regardless of your conclusions in 2.d.)

#Provide R code and output/graphs   
model.fer <- lm( poultry$FeedEffRatio ~ poultry$Additive + poultry$CuLevel)  
coef(model.fer)

## (Intercept) poultry$Additive poultry$CuLevel   
## 1.28704762 0.02900571 0.00020500

summary(model.fer)

##   
## Call:  
## lm(formula = poultry$FeedEffRatio ~ poultry$Additive + poultry$CuLevel)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.62739 -0.31153 -0.07828 0.26537 0.87238   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.2870476 0.1073076 11.994 <2e-16 \*\*\*  
## poultry$Additive 0.0290057 0.0015435 18.792 <2e-16 \*\*\*  
## poultry$CuLevel 0.0002050 0.0002636 0.778 0.44   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4084 on 57 degrees of freedom  
## Multiple R-squared: 0.8612, Adjusted R-squared: 0.8564   
## F-statistic: 176.9 on 2 and 57 DF, p-value: < 2.2e-16

From the summary of the proposed model we can see that the feed efficiency ratio is 1.287 when the additive is zero and the Copper level is zero as well, this is the intercept coefficient. Also, holding the Copper level constant, a one unit increase in additive (mg) will increase the feed efficiency ratio by 0.0290g. Finally, when we hold the additivie constant, a high level of copper (400ppm) will increase the feed efficiency ratio by 0.0002g.

### 2.g.

Interpret the correct value in the context of the model variables. Assess the strength of the predictive power of the model.

The correct value in the context of the model is the Adjusted R-squared of 0.8564 since we are doing multiple regression. This tells us that 85.64% of variability in the feed efficiency ratio is explained by a linear model with additive and Copper level.

### 2.h.

Test for possible correlations between the explanatory variables and assess the possible impact of collinearity. Provide measures of correlation, scatterplots and VIF assessment.

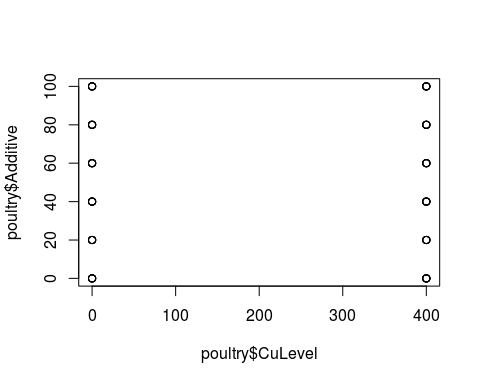
#Provide R code and output/graphs   
cor(poultry)

## CuLevel Additive FeedEffRatio  
## CuLevel 1.00000000 0.0000000 0.03837203  
## Additive 0.00000000 1.0000000 0.92723075  
## FeedEffRatio 0.03837203 0.9272308 1.00000000

library(car)  
vif(model.fer)

## poultry$Additive poultry$CuLevel   
## 1 1

plot(poultry$CuLevel, poultry$Additive)



There is zero correlation between the explanatory variables, but we can see a strong positive correlation between additive and feed efficiency ratio. Also, after running the VIF test, we have VIF scores of one for both explanatory variables, which implies that collinearity is not a problem here.

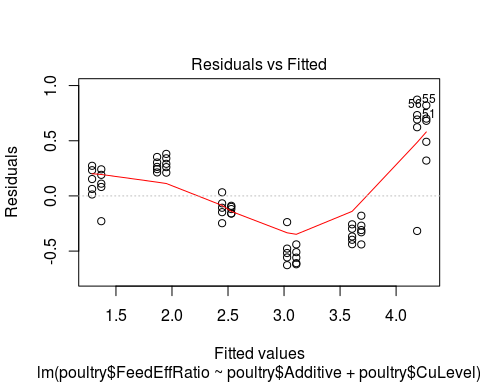
### 2.i.

Conduct appropriate testing for the assumptions of Regression Analysis.

#### 2.i.i.

Did the observations pass the equal variance assumption?

#Provide R code and output/graphs   
plot(model.fer, 1)

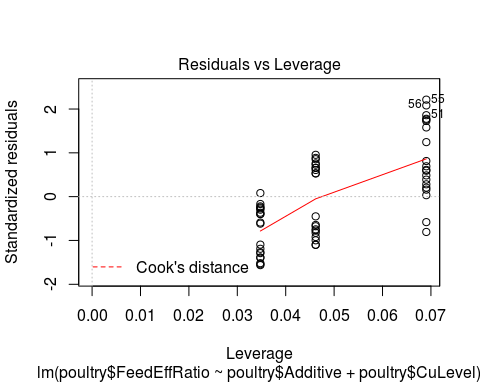
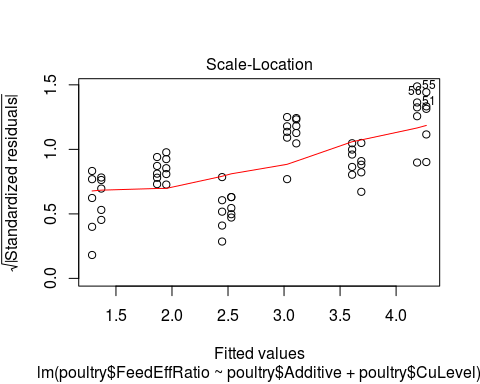
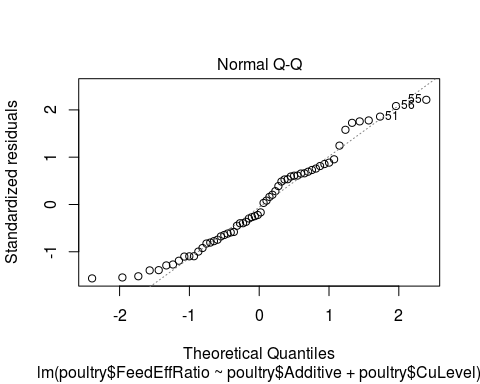
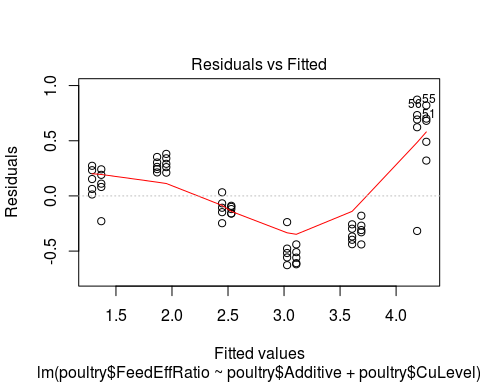


The observations appear to meet the equal variance assumption because the data points are evenly spread out from zero. About half the data points have a negative residuals and half of them have a positive residual and have a parabola shape.

#### 2.i.ii.

Did the observations pass the independence assumption?

#Provide R code and output/graphs   
plot(model.fer, 1)  
plot(model.fer)

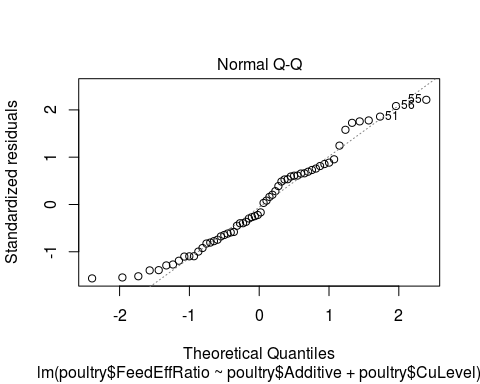


In this case the observations appear to have a pattern resembling curvature in the residuals versus the fitted values. They are not randombly distributed about zero, therefore the assumption of independence is not satisfied.

#### 2.i.iii.

Did the observations pass the normality assumption?

#Provide R code and output/graphs   
plot(model.fer, 2)

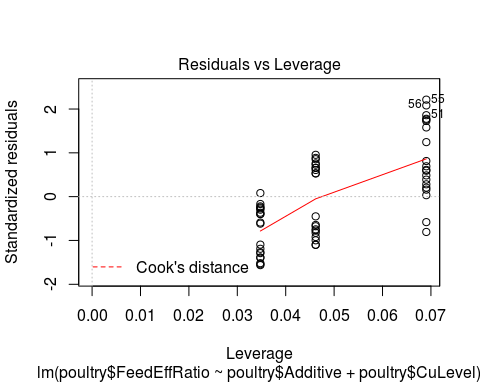


Judging from the QQ plot, we can say that the data does appear to meet the normality assumption. There is approximatley a straight line relationship between the quantiles of our data and the quantiles from the normal distribution.

#### 2.i.iv.

Did the observations pass the no outlier assumption?

#Provide R code and output/graphs   
plot(model.fer, 5)



There appears to be no outliers in our data, since all observations are under a 0.5 Cook’s Distance from the central cluster of observations. Thefore, our data does meet the assumption of no outliers points that might unduly alter our analysis and estimate of slope.

### 2.j.

The researchers decide to switch to a model that also includes a 2nd order and 3rd order term for Additive Level.

#### 2.j.i.

Write out the new model using , values, and values. (You might want to borrow the model from the Problem 2 description and modify it.)

#### 2.j.ii.

Determine the coefficients for each variable in the model. If possible, interpret the parameter estimates as usual. If not possible, explain why not.

#Provide R code and output/graphs   
add2 <- poultry$Additive \* poultry$Additive   
add3 <- poultry$Additive \* poultry$Additive \* poultry$Additive  
model.fer2 <- lm(FeedEffRatio~Additive + CuLevel + add2 + add3)  
summary(model.fer2)

##   
## Call:  
## lm(formula = FeedEffRatio ~ Additive + CuLevel + add2 + add3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.8973 -0.0526 0.0139 0.1018 0.2927   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.415e+00 5.768e-02 24.524 < 2e-16 \*\*\*  
## Additive 5.411e-02 5.171e-03 10.463 1.08e-14 \*\*\*  
## CuLevel 2.050e-04 1.109e-04 1.848 0.0699 .   
## add2 -1.148e-03 1.285e-04 -8.933 2.71e-12 \*\*\*  
## add3 9.420e-06 8.437e-07 11.165 9.30e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1718 on 55 degrees of freedom  
## Multiple R-squared: 0.9763, Adjusted R-squared: 0.9746   
## F-statistic: 566.3 on 4 and 55 DF, p-value: < 2.2e-16

The coefficient of the intercept is equal to 1.415e+00, which is the Feed Efficiency Ratio when all other variables are zero. However, it is not possible to interpret the parameter estimates as usual since we can not change the variable Additive and leave and fixed.

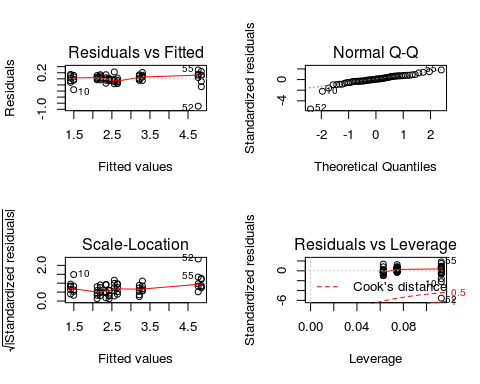
#### 2.j.iii.

Calcuate appropriate VIF values and assess the possible impact of collinearity.

#Provide R code and output/graphs   
library(car)  
vif(model.fer2)

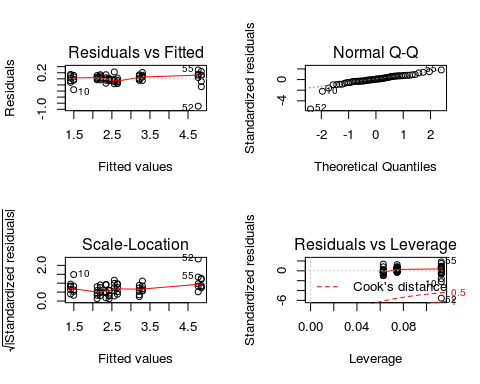
## Additive CuLevel add2 add3   
## 63.40664 1.00000 424.90046 186.38117

par(mfrow=c(2,2))  
plot(model.fer2)



From this VIF test we can see that we have a problem with collinearity. Three of our variables dependent on Additive have significantly large VIF scores of 63.4 and greater.  
####2.j.iv. Conduct appropriate testing for the assumptions of Regression Analysis. (Test the 4 assumptions other than linearity and collinearity; that is, equal variance, independence, and normality of residuals; and outliers)

#Provide R code and output/graphs  
par(mfrow=c(2,2))  
plot(model.fer2)

 >From the Residuals vs Fitted plot we can see that the data points are randomly distributed about zero (no trumpeting) and have no obvious pattern or shape (i.e. curvature), therefore this tells us the assumptions for constant variance and independence of residuals is satisfied. From the QQ plot, we can observe that the data does appear to be normally distributed since the quantiles from our data approximately align with the quantiles from a normal distribution. Finally, from the Cook’s Distance plot, we can see that there is only one outlier of concern, specifically data point 52, which is between 0.5 to 1 Cook’s Distance. This could be considered an outlier and would have to be furter investigated to ensure assumption of no outliers in data that might unduly alter our analysis and estimates of slope.

#### 2.j.v.

Now we will conduct the test to determine if the linear relationship is statistically significant by testing the *general* hypothesis for the full model.

##### 2.j.v.1

State the *general* hypothesis for the full model (null and alternative) for Multiple Regression Analysis mathematically and in a complete sentence.

All partial slopes are equal to zero. The model has no explanatory power of my response.

At least one

At least one partial slope is not equal to zero. The model has some predictive power.

##### 2.j.v.2

Use R to test the general hypothesis.

#Provide R code and output/graphs   
summary(model.fer2)

##   
## Call:  
## lm(formula = FeedEffRatio ~ Additive + CuLevel + add2 + add3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.8973 -0.0526 0.0139 0.1018 0.2927   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.415e+00 5.768e-02 24.524 < 2e-16 \*\*\*  
## Additive 5.411e-02 5.171e-03 10.463 1.08e-14 \*\*\*  
## CuLevel 2.050e-04 1.109e-04 1.848 0.0699 .   
## add2 -1.148e-03 1.285e-04 -8.933 2.71e-12 \*\*\*  
## add3 9.420e-06 8.437e-07 11.165 9.30e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1718 on 55 degrees of freedom  
## Multiple R-squared: 0.9763, Adjusted R-squared: 0.9746   
## F-statistic: 566.3 on 4 and 55 DF, p-value: < 2.2e-16

##### 2.j.v.3

Interpret the meaning of the p-value in the context of the situation.

The probability that we observe our test statistic or more extreme is less than 2.2e-16 < .0001.

##### 2.j.v.4

State the conclusion of the hypothesis test in the context of the problem.

Let F-statistic=566.3, p-value<.00001. Based on our sample, using a significance level of .05 we will reject the null hypothesis that the model has no explanatory power for the response. We have evidence to support the alternative that at least one explanatory variable has some statistically significant relationship with the response.

#### 2.j.vi.

Interpret the correct value in the context of the model variables. Assess the strength of the predictive power of the model.

We see from this model that 97.46% of the variability in Feed Efficiency Ratio is explained by the linear regression with Additive, Copper Level, Additive squared and Additive cubed variables.

### 2.k

Compare the two models that you considered for predicting Feed Effect Ratio. Which model would be the best (relative to each other) in predicting Feed Effect Ratio? Justify your answer using statistical results.

Comparing the two models above for predicting Feed Effeciency Ratio, we can say that the second model(model.fer2) is the best relative to the first. We can see that comparing them on the adjusted R-squared scores, the variables from model two explain 97.4% of the variability in Feed Efficiency Ratio, while model one only explains about 85.6% of the variability in Feed Efficiency Ratio. However, there is a concern in the second model because the VIF test for collinearity between explanatory variables fails with significantly high scores.

*The end.*