Ben Goodwin

DS6371

HW12

Question 1)

Part A

SAS input for non-transformed data:

FILENAME REFFILE '/folders/myfolders/sasuser.v94/Crab17.csv';

PROC IMPORT DATAFILE=REFFILE

DBMS=CSV

OUT=crabby;

GETNAMES=YES;

RUN;

proc sgplot data=crabby;

styleattrs wallcolor=blanchedalmond;

scatter x=Height y=Force / group = Species

markerattrs = (symbol=CircleFilled size=7)

FILLEDOUTLINEDMARKERS;

run;

SAS output for non-transformed data:

Chart, scatter chart

Description automatically generated

SAS input for transformed data:

DATA logdata;

SET crabby;

LOGVAR=log10(Force);

RUN;

proc sgplot data=logdata;

styleattrs wallcolor=blanchedalmond;

scatter x=Height y=LOGVAR / group = Species

markerattrs = (symbol=CircleFilled size=7)

FILLEDOUTLINEDMARKERS;

run;

SAS output for transformed data:

Chart, scatter chart

Description automatically generated

Part B

Mu\_hat|height,species=beta\_0+beta\_1\*Height+beta\_2\*Species+beta\_3\*interaction(Height\*Species)

Part C

SAS Input:

proc glm data=crabby plots=all;

class Species(ref="Lophopanopeus bellus");

model Force = Height|Species /solution clparm;

run;

SAS Output:

Graphical user interface, text, application

Description automatically generated

Table

Description automatically generated

Graphical user interface, table

Description automatically generated

Part D

Residual Plot:

Chart, scatter chart

Description automatically generated

Appropriateness: The residuals seem to be spread out in a random manner, there is no obvious trend to the data.

Studentized Residual Plot:

Chart, scatter chart

Description automatically generated

Appropriateness: The studentized residuals indicate the data is very similar to one another, there is slight visual evidence for 1 outlier.

Histogram of Residuals:

Chart, histogram

Description automatically generated

Appropriateness: The residuals look very close to normally distributed

Quantile-Quantile Plot:

Chart, scatter chart

Description automatically generated

Appropriateness: The qq-plot looks about as ideal as one should look. The residuals match up very neatly with the normal quantiles and even match at the tails.

In general the model assumptions are met, including linearity, which is confirmed by the scatterplots. Normality looks good from the histograms. Equal-standard-deviations are met from the scatterplots.

Additionally, we will assume independence is met.

Part E

I have determined that the model is sufficient.

Interpretation of force coefficient:

Interpretation: With a height of 0 the Lophopanopeus bellus crab would exert -17.25 force units.

With a height of 0 the cancer productus would exert 8.24823 force units than the baseline Lophopanopeus bellus.

With a height of 0 the Hemigrapsus Nudus would exert 20.41259323 force units than the baseline Lophopanopeus bellus. We will speak in terms of medians because we have log transformed the data.

Interpretation of height coefficient:

For a one unit increase in median height the Lophopanopeus bellus crab we can expect a 3.8486 increase in force units. The Cancer Productus, for a one unit increase in median height in the Lophopanopeus Bellus will produce a -1.1687 increase in force units. And finally for a a one unit increase in median height Hemigrapsus will gain a -3.078 increase in force units. We will speak in terms of medians because we have log transformed the data.

Part F

Mu{force|Height for Hemigrapsus Nudus}: -17.25249079+3.84863387\*height

Mu{force|Height for Cancer Productus}: -9.00425074+2.67983032\*height

Mu{force|Height for Lophopanopeus Bellus }…3.1601025+0.14480222\*height

Question 2)

Part A Plots:

R input:

#read data

brainDat <- read.csv(“brainDat.csv”)

#look at data

head(brainDat)

#transform data

brainDat$logBrain <- log(brainDat$Brain)

brainDat$logBody <- log(brainDat$Body)

brainDat$logGestation <- log(brainDat$Gestation)

brainDat$logLitter <- log(brainDat$Litter)

brainDat$bodyGroup <- cut(brainDat$logBody,c(-Inf,-2,0,2,4,6,Inf))

levels(brainDat$BodyGrouping) <- c(“Group1”,”Group2”,”Group3”,”Group4”,”Group5”,”Group6”)

#Create linear model

brainDatlm <- lm(logBrain~logBody+logGestation+loglitter,data=brainDat)

#Look at summary stats

summary(brainDatlm)

#Non-transformed data (Gestation and Brain)

p <- ggplot(brainDat,aes(Gestation,Brain))

p+geom\_point(aes(color=factor(bodyGroup)),data = brainDat)+ xlab(“Gestation time”) + ylab(“Brain weight in Grams”)+theme\_minimal()+ theme(legend.title = element\_blank())+ggtitle(“Brain Weight in Grams and Gestation Time”)

#Non-transformed data (Litter and Brain )

q <- ggplot(brainDat,aes(Litter,Brain))

q+geom\_point(aes(color=factor(bodyGroup)),data = brainDat)+ xlab(“Litter Size”) + ylab(“Brain weight in Grams”)+ theme\_minimal()+ theme(legend.title = element\_blank())+ggtitle(“Brain Weight in Grams and Litter Size”)

#Non-transformed data (Body and Brain )

r <- ggplot(brainDat,aes(Body,Brain))

r+geom\_point(aes(color=factor(bodyGroup)),data = brainDat)+ xlab(“Body Size”) + ylab(“Brain weight in Grams”)+ theme\_minimal()+ theme(legend.title = element\_blank())+ggtitle(“Brain Weight in Grams and Body Size”)

#Looks at logged plots

#Transformed data (Gestation and Brain)

p <- ggplot(brainDat,aes(logGestation,logBrain))

p+geom\_point(aes(color=factor(bodyGroup)),data = brainDat)+ xlab(“Log Gestation time”) + ylab(“Brain weight in Grams”)+theme\_minimal()+ theme(legend.title = element\_blank())+ggtitle(“Log Brain Weight in Grams and Log Gestation Time”)

#Transformed data (Litter and Brain )

q <- ggplot(brainDat,aes(logLitter,logBrain))

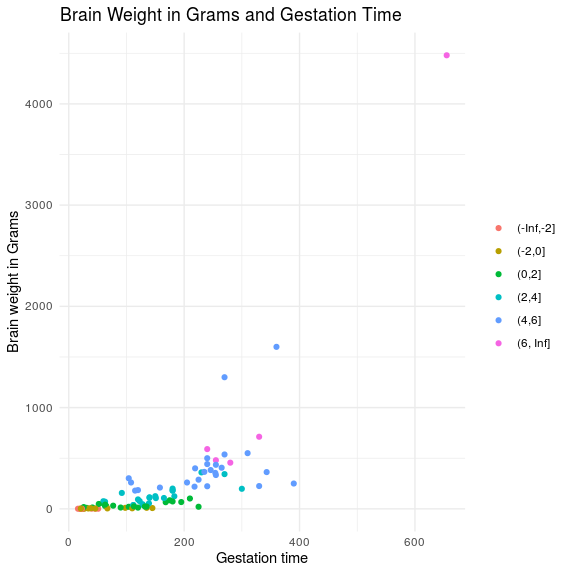
q+geom\_point(aes(color=factor(bodyGroup)),data = brainDat)+ xlab(“Log Litter Size”) + ylab(“Brain weight in Grams”)+ theme\_minimal()+ theme(legend.title = element\_blank())+ggtitle(“Log Brain Weight in Grams and Log Litter Size”)

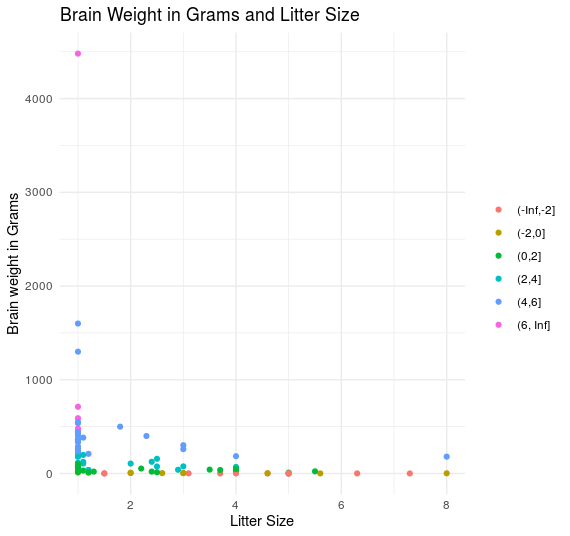
#Transformed data (Body and Brain )

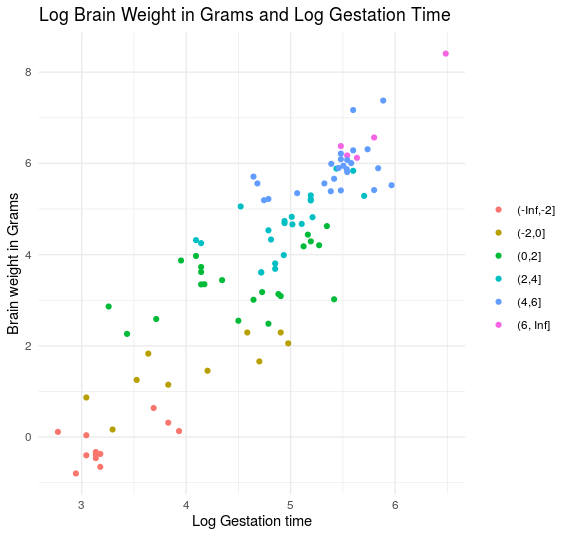
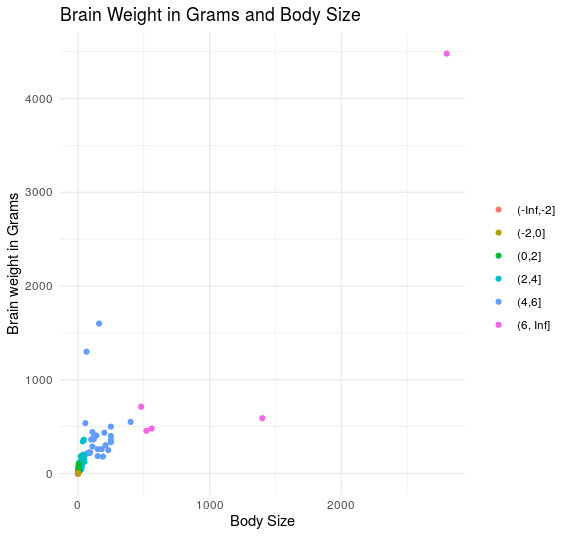
r <- ggplot(brainDat,aes(logBody,logBrain))

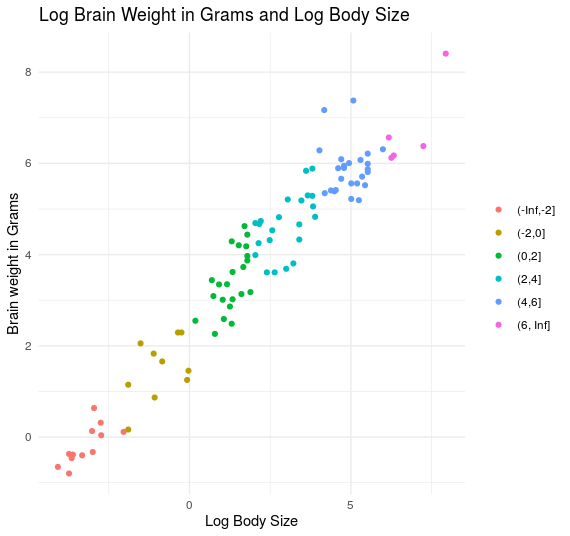
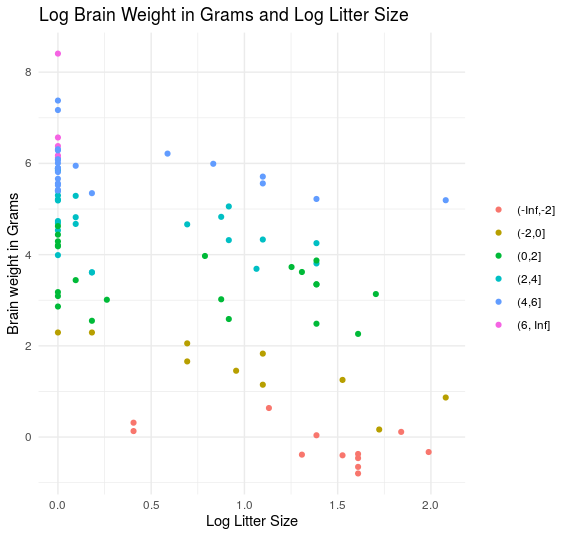
r+geom\_point(aes(color=factor(bodyGroup)),data = brainDat)+ xlab(“Log Body Size”) + ylab(“Brain weight in Grams”)+ theme\_minimal()+ theme(legend.title = element\_blank())+ggtitle(“Log Brain Weight in Grams and Log Body Size”)

R output:









Part B

Model mu\_hat{brain\_size|weight,gestitiation,body}=beta\_0+beta\_1\*gestitation+beta\_2\*Litter+beta\_3\*body

Part C

R input:

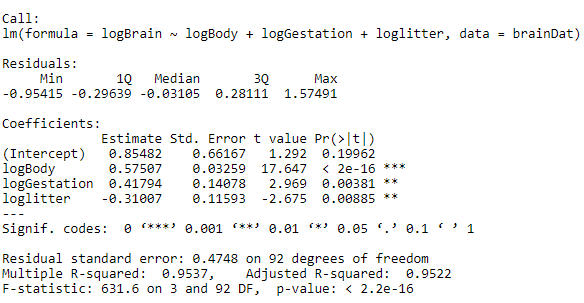
#Create linear model

brainDatlm <- lm(logBrain~logBody+logGestation+loglitter,data=brainDat)

#Look at summary stats

summary(brainDatlm)

R output:



Part D

R input:

plot(brainDatlm)

Brainres <- resid(brainDatlm)

par(mfrow=c(2,2))

plot(Brainres)

plot(brainDat$logGestation,Brainres, ylab="Residuals", xlab="log Gestation Period")

abline(0,0)

plot(brainDat$logBody,Brainres, ylab="Residuals", xlab="log Body")

abline(0,0)

plot(brainDat$loglitter,Brainres, ylab="Residuals", xlab="log Litter")

abline(0,0)

#Make histogram with nrml curve overlaid

#Save residuals

studentizedResidualsBrainDat <- rstudent(brainDatlm)

#Create histogram

hist(studresBrain, freq=FALSE, main="Distribution of Studentized Residuals",xlab="Studentized Residuals", ylab="Density", ylim=c(0,0.5))

##Make x and y values for normal range

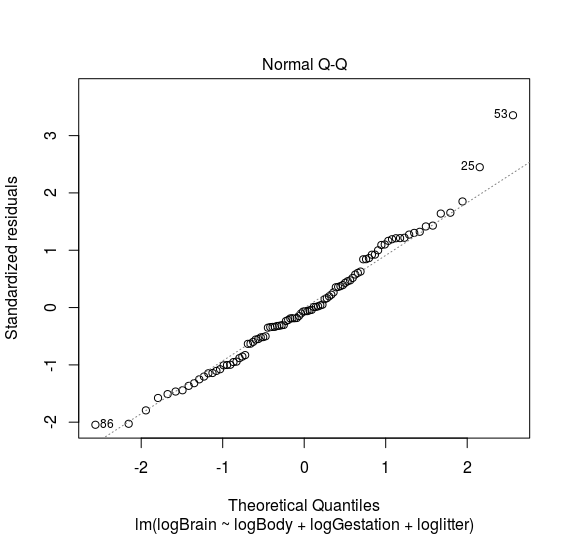
xfit <- seq(min(studentizedResidualsBrainDat)-1, max(studentizedResidualsBrainDat)+1, length=40)

yfit <- (dnorm(xfit))

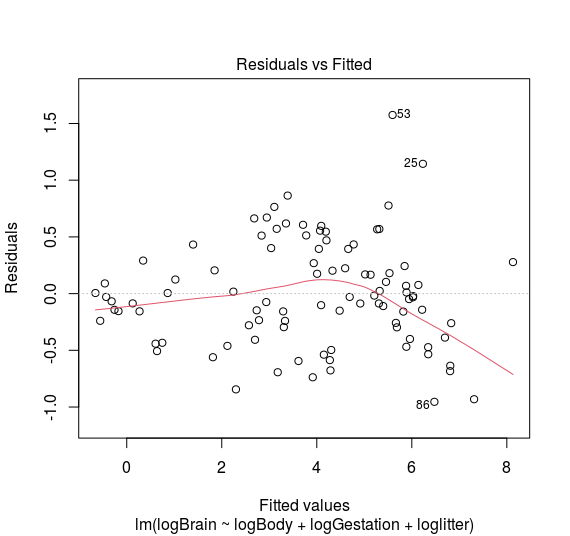
##Add normal curve

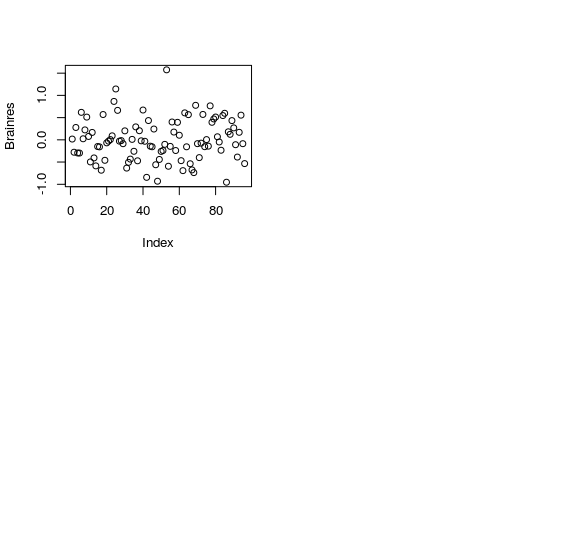
lines(xfit, yfit, ylim=c(0,0.5))

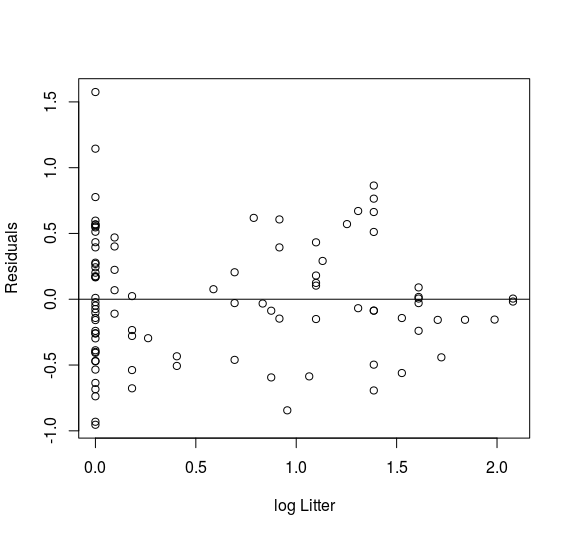
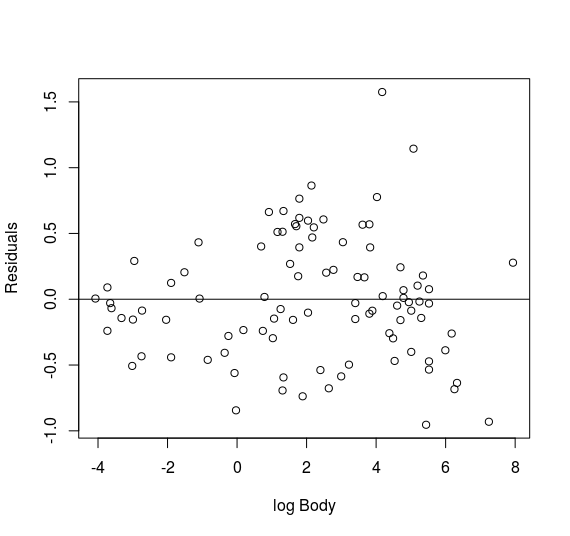
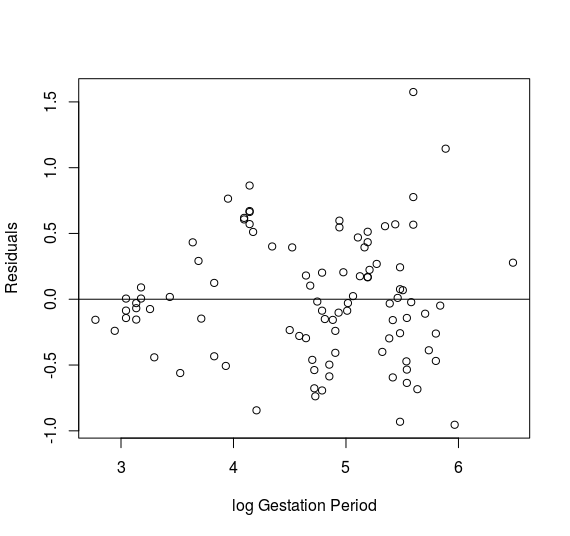
R Output:



Based on the quantile-quantile plot above, it seems that the data follows a normal distribution.



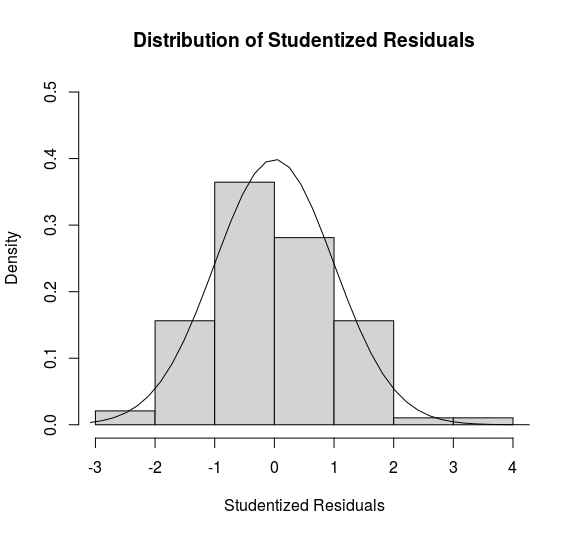




Based on each of the residual plots above, the log-transformed data does not have any obvious trends or irregularities.



The studentized residual plot looks good and shows that the data are generally near each other and there is little evidence for outliers.



The histogram of the studentized residuals follows closely to the normal distribution and seems like the log-transformed data makes a good candidate to model on.

In general the model assumptions are met, including linearity, which is confirmed by the scatterplots. Normality looks good from the histograms. Equal-standard-deviations are met from the scatterplots.

Additionally, we will assume independence is met.

Part E

Interpretation of model: In the general case we can say that for a one unit increase in gestation we can expect median brain size increase by .42. Again in the general case we can say that for a one unit increase in litter we can expect median brain size to increase by -.31. Finally we can assess body size, which for a one unit increase in median body size we can expect brain size to increase by .58. We will speak in terms of medians because we have log transformed the data.

Bonus:

A) 32 DOF

B) 19.761

E) 92 DOF

F 0.216