

# Mammals

*Yang*

*May 22, 2019*

## Load dataset

```
library(MASS)
data(mammals)
attach(mammals)
head(mammals)
```

```
##           body brain
## Arctic fox    3.385  44.5
## Owl monkey    0.480  15.5
## Mountain beaver 1.350   8.1
## Cow          465.000 423.0
## Grey wolf     36.330 119.5
## Goat         27.660 115.0
```

## Baseline Model

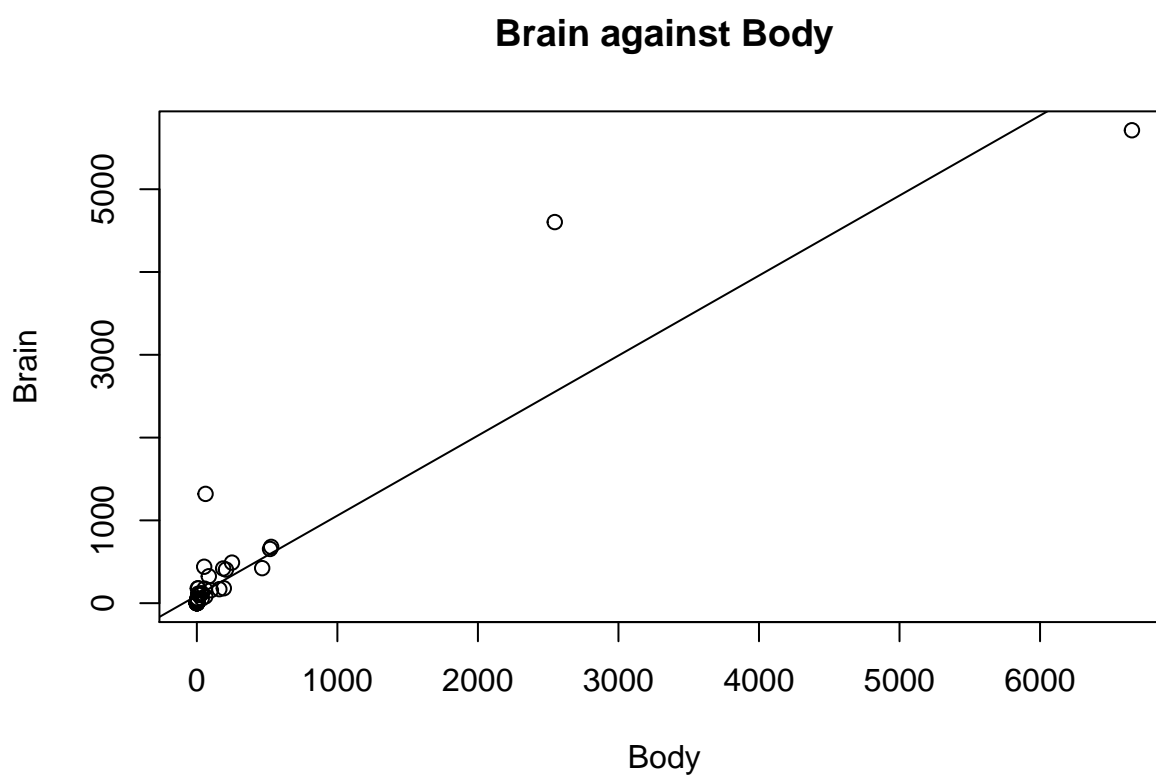
```
Body=mammals$body
Brain=mammals$brain
model_baseline<-lm(Brain~Body)
summary(model_baseline)
```

```
##
## Call:
## lm(formula = Brain ~ Body)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -810.07  -88.52  -79.64  -13.02  2050.33
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  91.00440   43.55258    2.09  0.0409 *
## Body         0.96650    0.04766   20.28 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 334.7 on 60 degrees of freedom
## Multiple R-squared:  0.8727, Adjusted R-squared:  0.8705
## F-statistic: 411.2 on 1 and 60 DF,  p-value: < 2.2e-16
```

## Take Log Transformations

```
#take log transformations
Log_Brain<-log(Brain)
Log_Body<-log(Body)
#try different model
model1=lm(Log_Brain~Body)
model2=lm(Brain~Log_Body)
model3=lm(Log_Brain~Log_Body)
bestmodel=model3#Best after log transformation
```

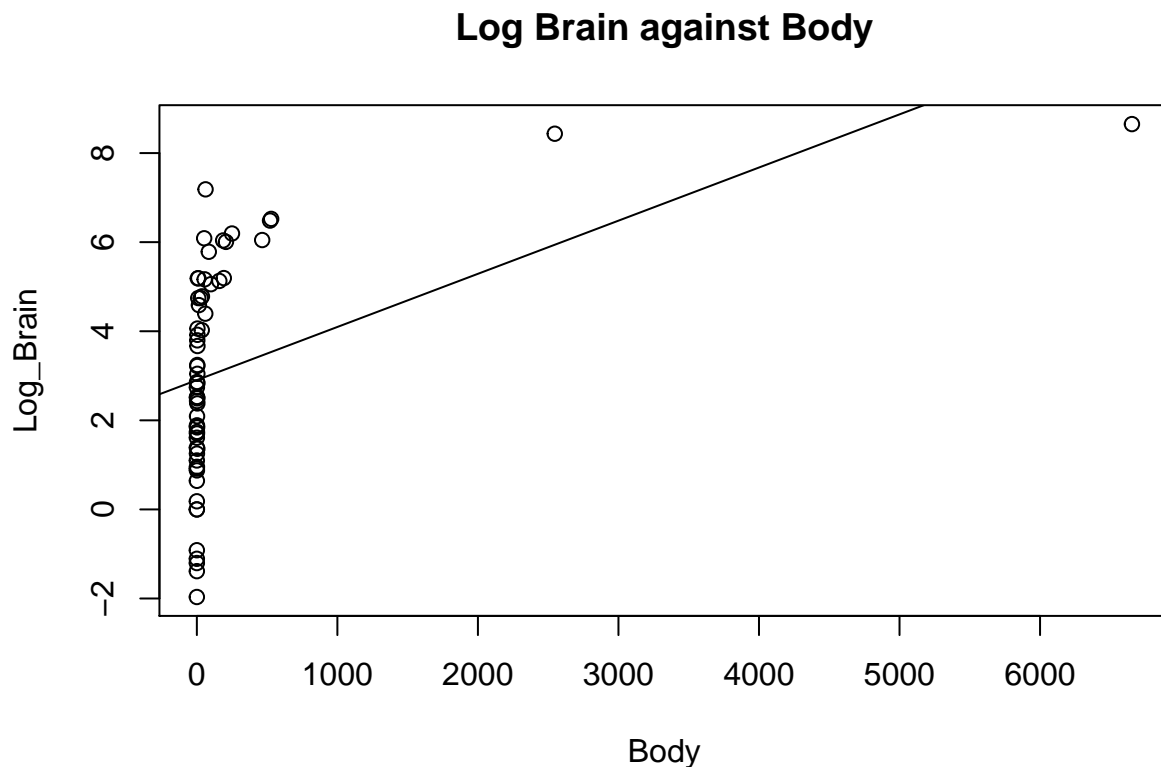
```
plot(Body, Brain,main="Brain against Body")
abline(model_baseline)
```



```
summary(model_baseline)
```

```
##
## Call:
## lm(formula = Brain ~ Body)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -810.07  -88.52  -79.64  -13.02  2050.33
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
##
```

```
## (Intercept) 91.00440 43.55258 2.09 0.0409 *
## Body 0.96650 0.04766 20.28 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 334.7 on 60 degrees of freedom
## Multiple R-squared: 0.8727, Adjusted R-squared: 0.8705
## F-statistic: 411.2 on 1 and 60 DF, p-value: < 2.2e-16
plot(Body, Log_Brain, main="Log Brain against Body")
abline(model1)
```

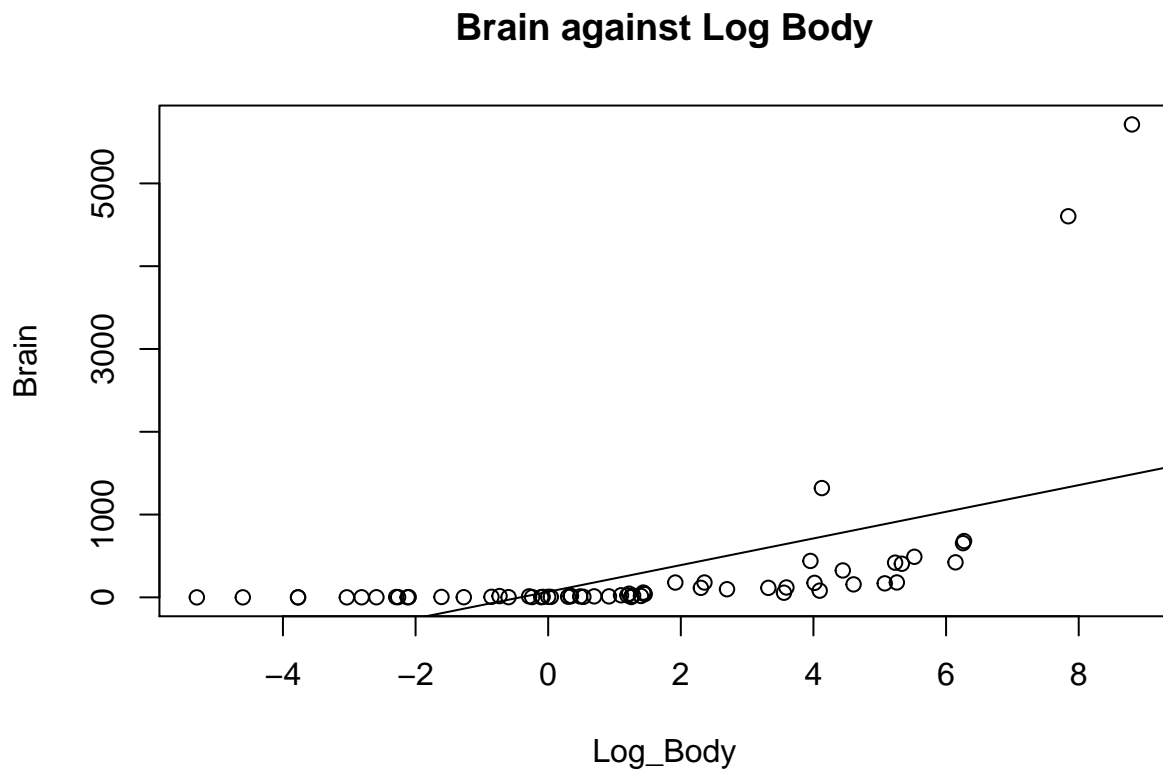


```
summary(model1)
```

```
##
## Call:
## lm(formula = Log_Brain ~ Body)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.8691 -1.5389 -0.1187  1.9847  4.2084
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.9030241   0.2884703   10.064 1.68e-14 ***
## Body         0.0011931   0.0003157    3.779 0.000364 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.217 on 60 degrees of freedom
## Multiple R-squared:  0.1923, Adjusted R-squared:  0.1788
## F-statistic: 14.28 on 1 and 60 DF,  p-value: 0.0003642
```

```
plot(Log_Body, Brain,main="Brain against Log Body")
abline(model2)
```

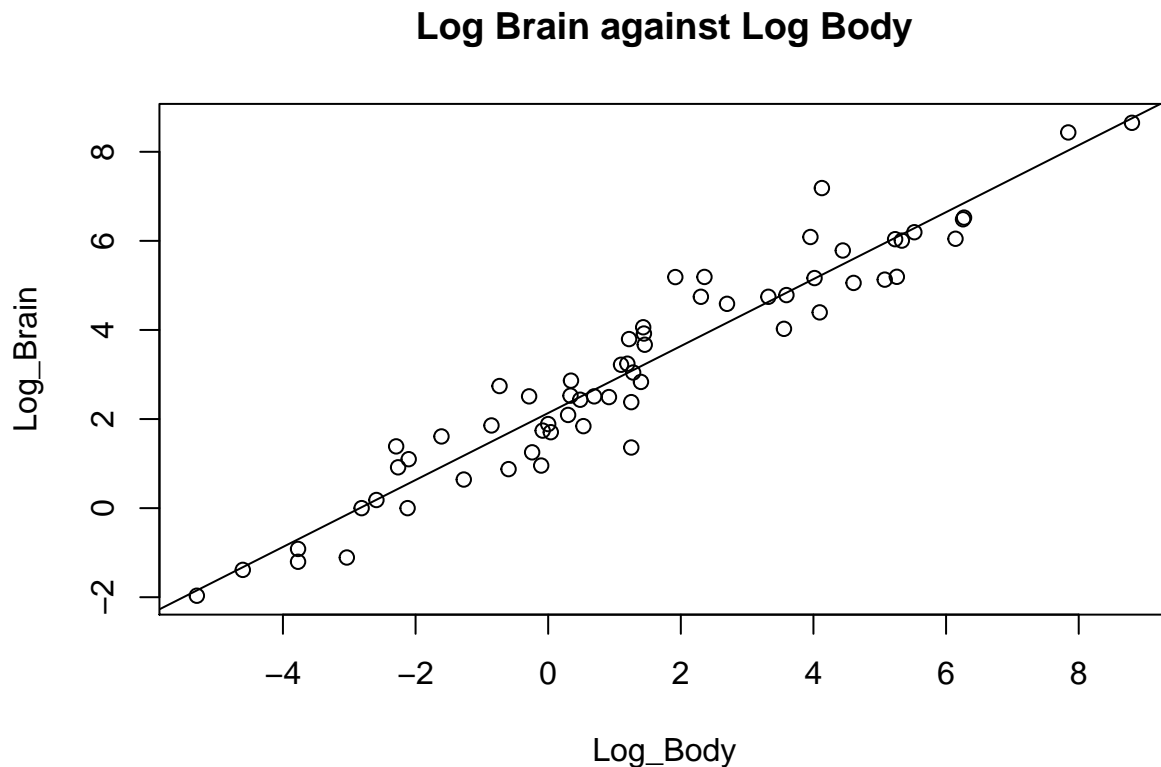


```
summary(model2)
```

```
##
## Call:
## lm(formula = Brain ~ Log_Body)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -734.2  -378.9  -182.3   123.4  4227.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    67.79    109.17   0.621   0.537
## Log_Body      161.00     32.35   4.976 5.77e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 789.2 on 60 degrees of freedom
```

```
## Multiple R-squared:  0.2921, Adjusted R-squared:  0.2803
## F-statistic: 24.76 on 1 and 60 DF,  p-value: 5.769e-06
```

```
plot(Log_Body, Log_Brain,main="Log Brain against Log Body")
abline(model3)
```



```
summary(model3)
```

```
##
## Call:
## lm(formula = Log_Brain ~ Log_Body)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71550 -0.49228 -0.06162  0.43597  1.94829
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.13479    0.09604   22.23  <2e-16 ***
## Log_Body       0.75169    0.02846   26.41  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6943 on 60 degrees of freedom
## Multiple R-squared:  0.9208, Adjusted R-squared:  0.9195
## F-statistic: 697.4 on 1 and 60 DF,  p-value: < 2.2e-16
```

Based on the Adjusted R-squared value, applying log transformation to **both** Brain and Body is the best

choice. The model is  $\log(\text{Brain}) = 0.75169\log(\text{Body}) + 2.13479 + \epsilon$  with  $\beta_0 = 2.13479$  and  $\beta_1 = 0.75169$ . This means if we increase log of Body weight by 1, we expect an increment in log of Brain weight by 0.75169. And the expected log of Brain weight with log of Body weight being 0 is 2.13479.

## Gamma GLM with log-link

### Fit

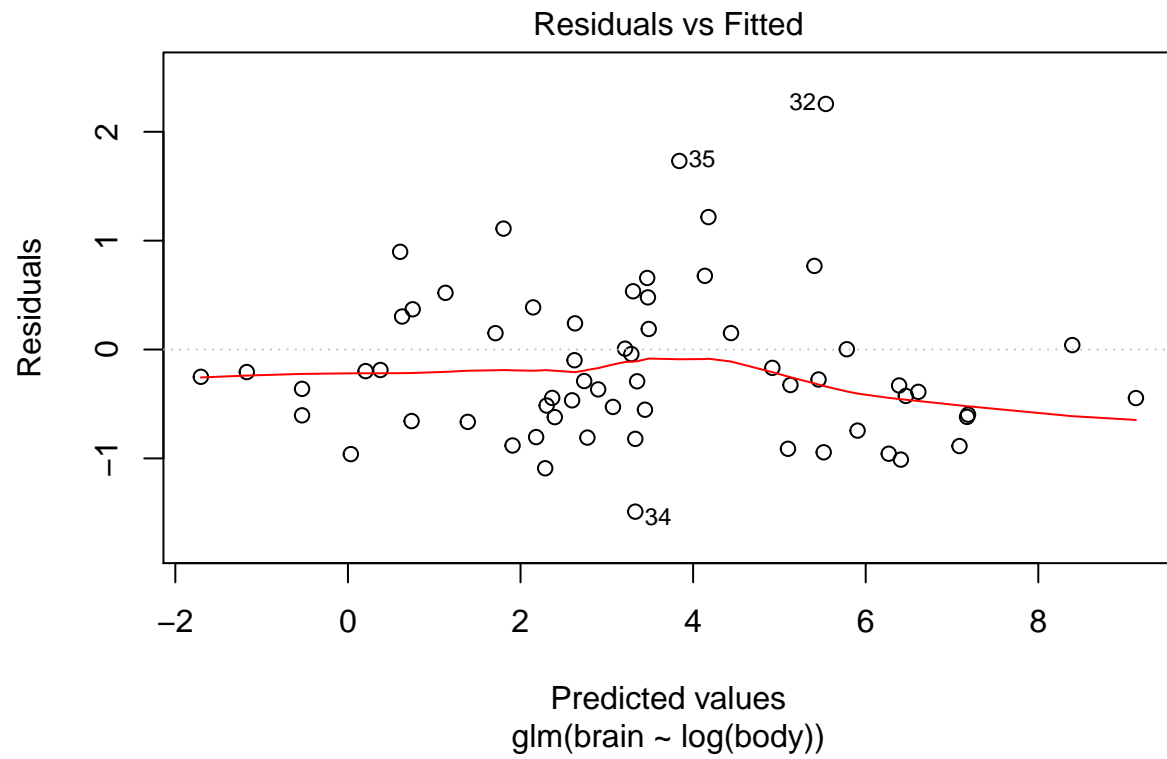
```
m3 <- glm(brain~log(body),family=Gamma(link="log"))
summary(m3)
```

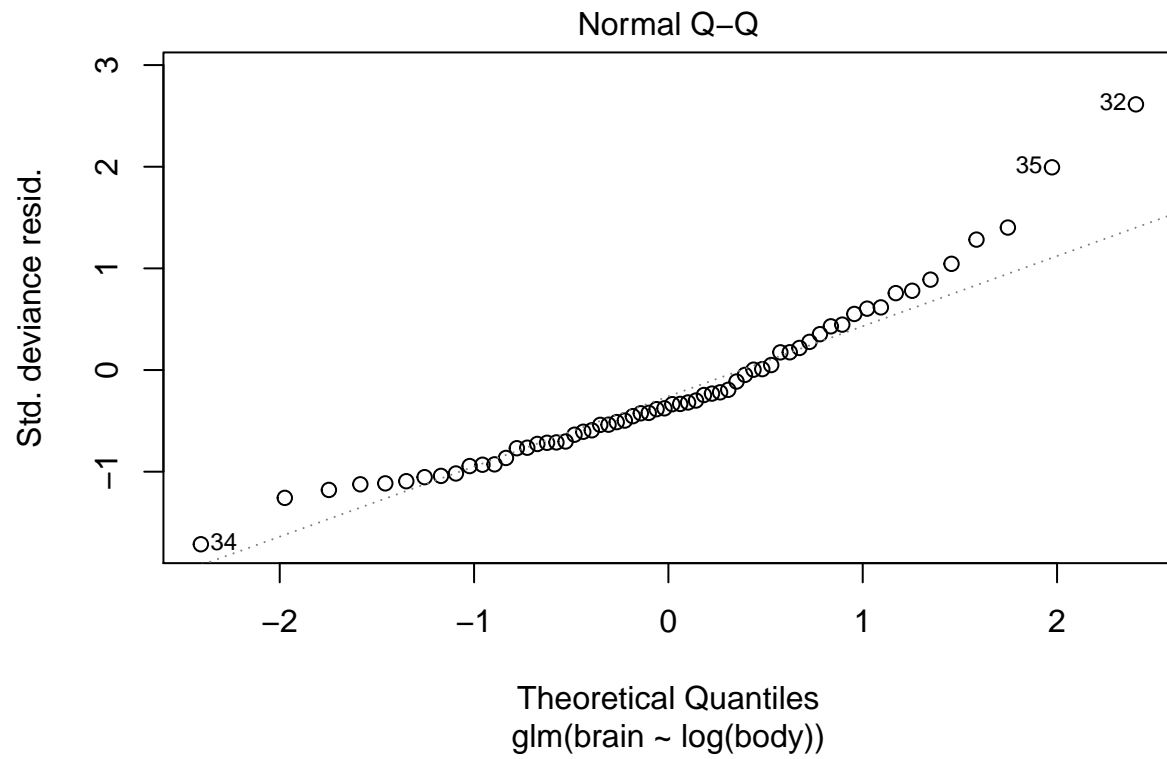
```
##
## Call:
## glm(formula = brain ~ log(body), family = Gamma(link = "log"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4889  -0.6204  -0.3089   0.1794   2.2555
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.36702    0.12116   19.54  <2e-16 ***
## log(body)    0.76846    0.03591   21.40  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.7671765)
##
##      Null deviance: 310.710  on 61  degrees of freedom
## Residual deviance:  31.578  on 60  degrees of freedom
## AIC: 534.72
##
## Number of Fisher Scoring iterations: 6
```

```
coefficients(m3)
```

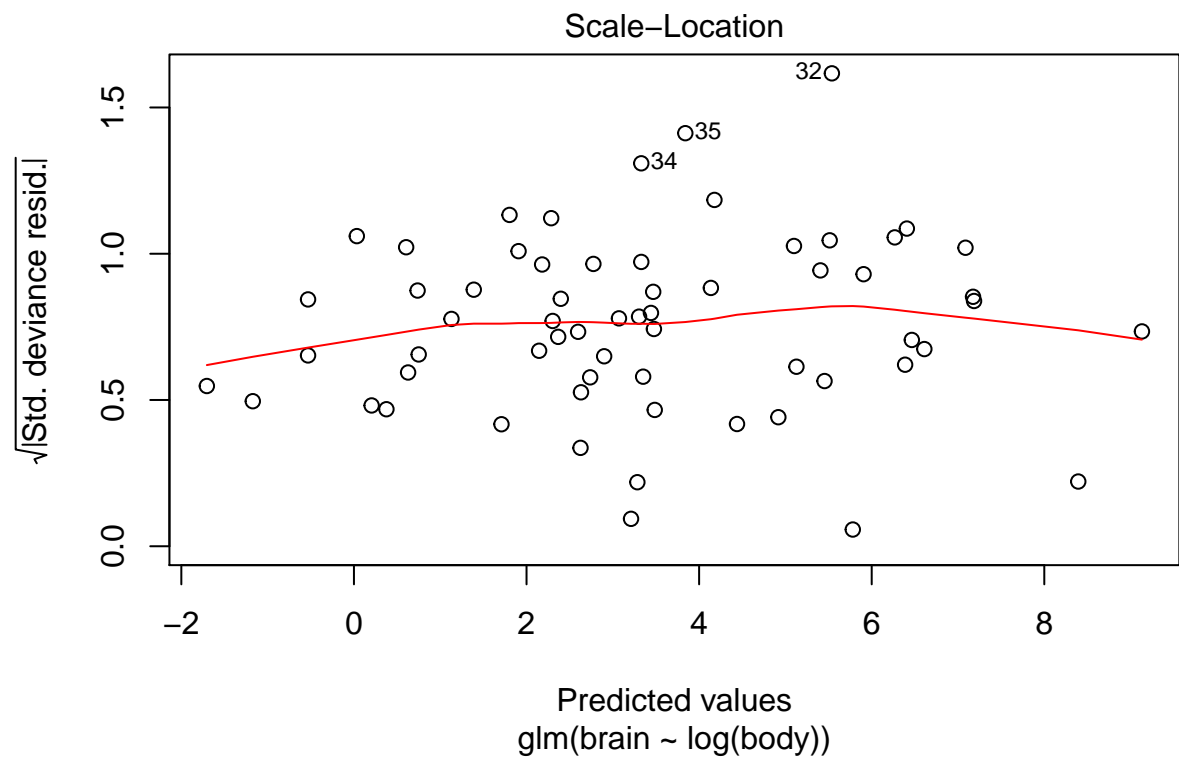
```
## (Intercept)    log(body)
##    2.3670159    0.7684591
```

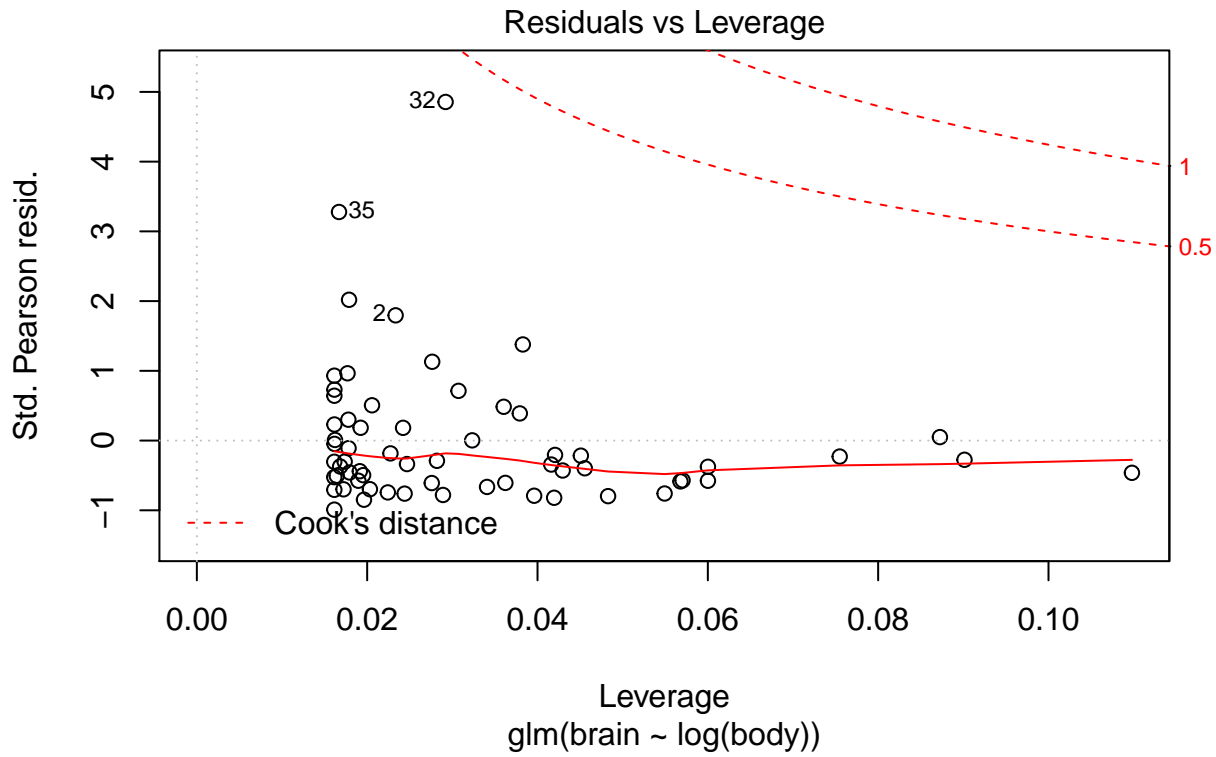
```
plot(m3)
```











## Interpretation

$\log(E(Y|X)) = \beta_0 + \beta_1 \cdot X \rightarrow E(Y|X) = e^{\beta_0} e^{\beta_1 X}$  suggests that the expected value of Y given X increases by  $e^{2.3670159}$ .

And  $E(Y|X = x) = e^{\beta_0} e^{\beta_1 x}$   $E(Y|X = x - 1) = e^{\beta_0} e^{\beta_1 (x-1)} = \frac{e^{\beta_0} e^{\beta_1 x}}{e}$  suggests that if we change X by 1, we expect a multiplicative increase of  $E(Y|X)$  by e.

## Gamma GLM with reciprocal link

### Fit

```
m4 <- glm(brain~log(body),family=Gamma(link="inverse"))
summary(m4)
```

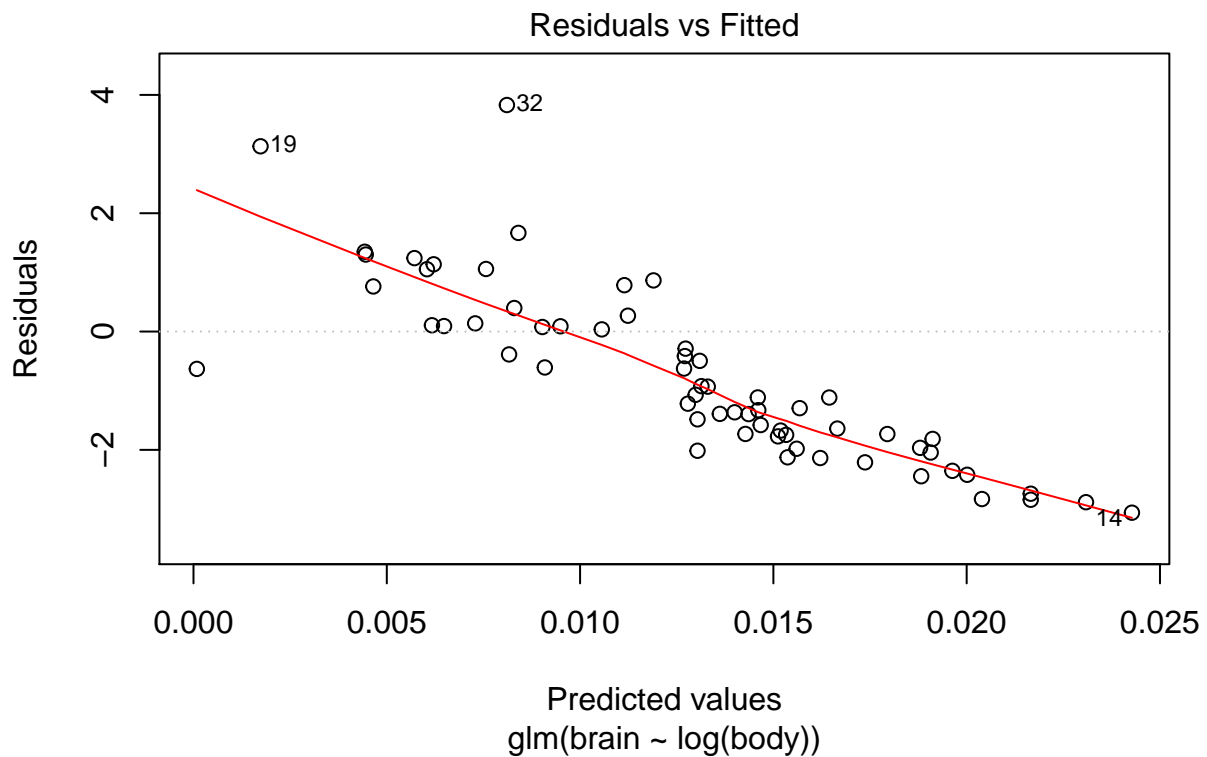
```
##
## Call:
## glm(formula = brain ~ log(body), family = Gamma(link = "inverse"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.062  -1.804  -1.114   0.103   3.829
##
## Coefficients:
```

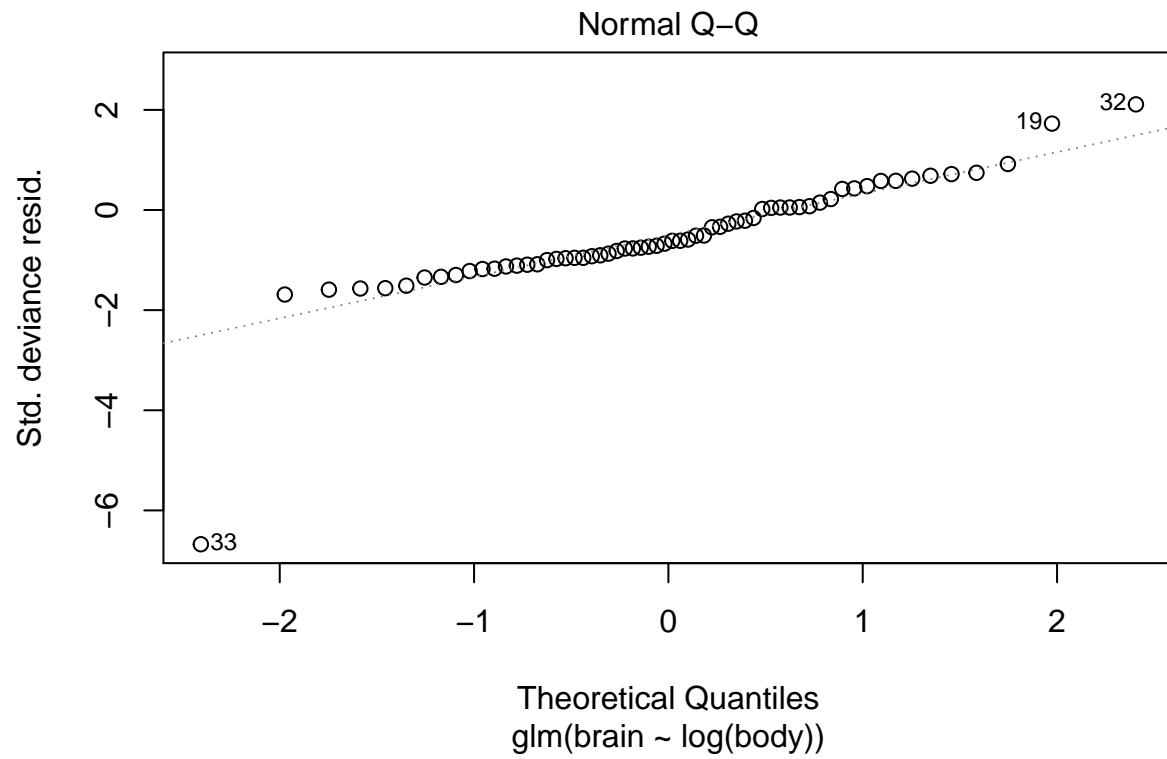
```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.015187   0.003567   4.257 7.39e-05 ***
## log(body)    -0.001715   0.000406  -4.225 8.26e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 3.345482)
##
## Null deviance: 310.71  on 61  degrees of freedom
## Residual deviance: 170.27  on 60  degrees of freedom
## AIC: 658.95
##
## Number of Fisher Scoring iterations: 8
```

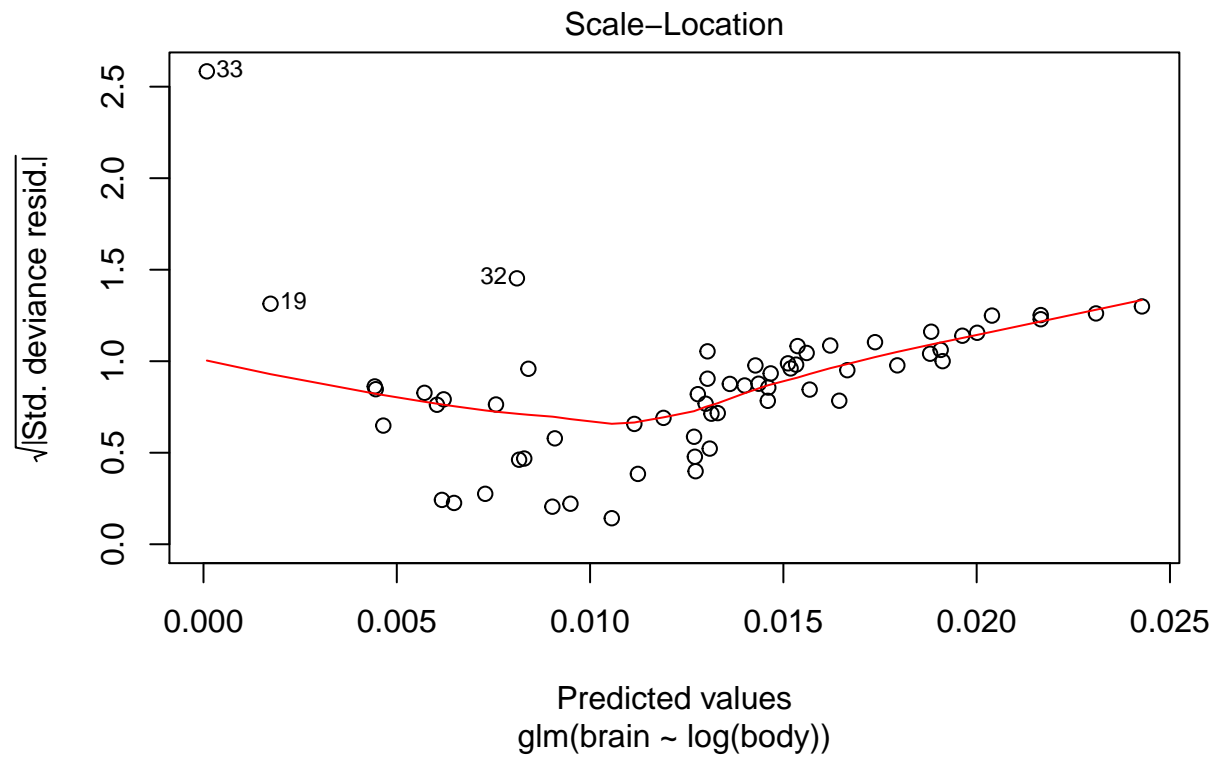
```
coefficients(m4)
```

```
## (Intercept)    log(body)
##  0.015186825 -0.001715374
```

```
plot(m4)
```

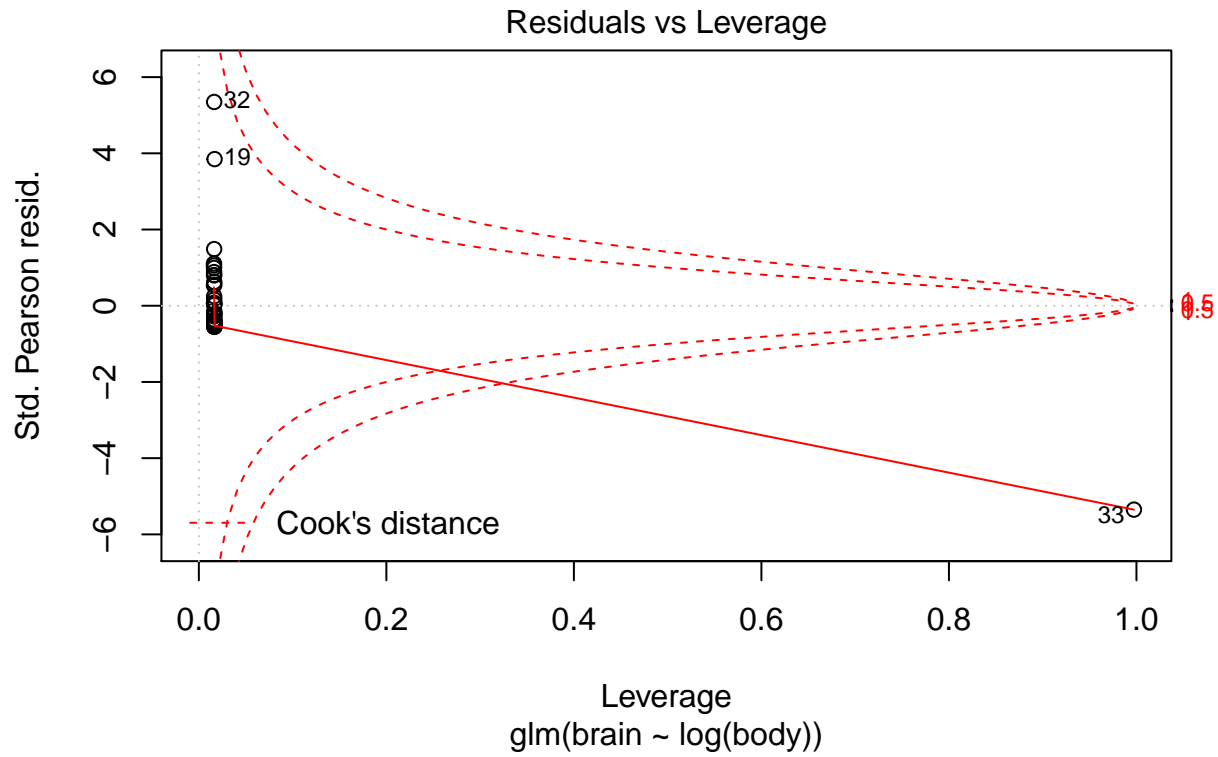






```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



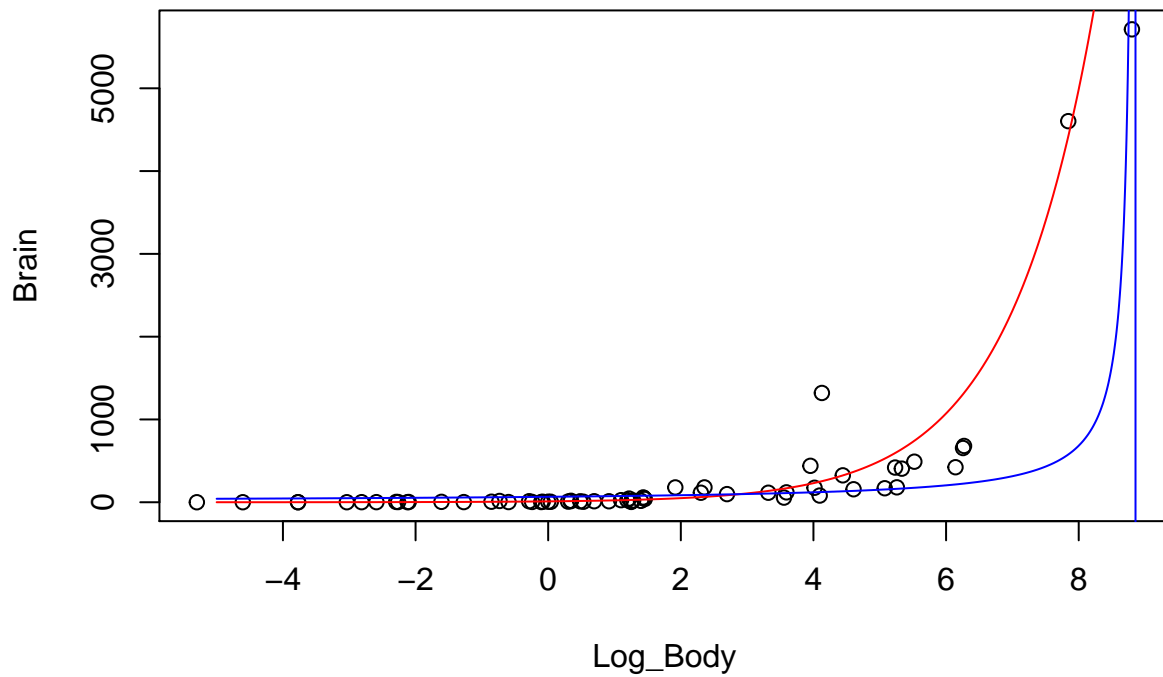
## Comparing Gamma GLM with log link(m3) and reciprocal link(m4)

I prefer the model m3 since:

1. According to the residual plot, there is an obvious pattern of linearity in m4 in part d while the residuals are distributed evenly around 0 in m3.
2. Lower AIC value of m3 than that of m4, suggesting m3 is a better fit.

## Fit plot

```
x<-seq(from=-5,to=9,by=0.01)
#model m3
beta3<-as.numeric(coef(m3))
y_predicted3<-exp(beta3[1]+x*beta3[2])
#model m4
beta4<-as.numeric(coef(m4))
y_predicted4<-(beta4[1]+x*beta4[2])^(-1)
plot(Log_Body,Brain)
lines(x,y_predicted3,col="red")
lines(x,y_predicted4,col="Blue")
```



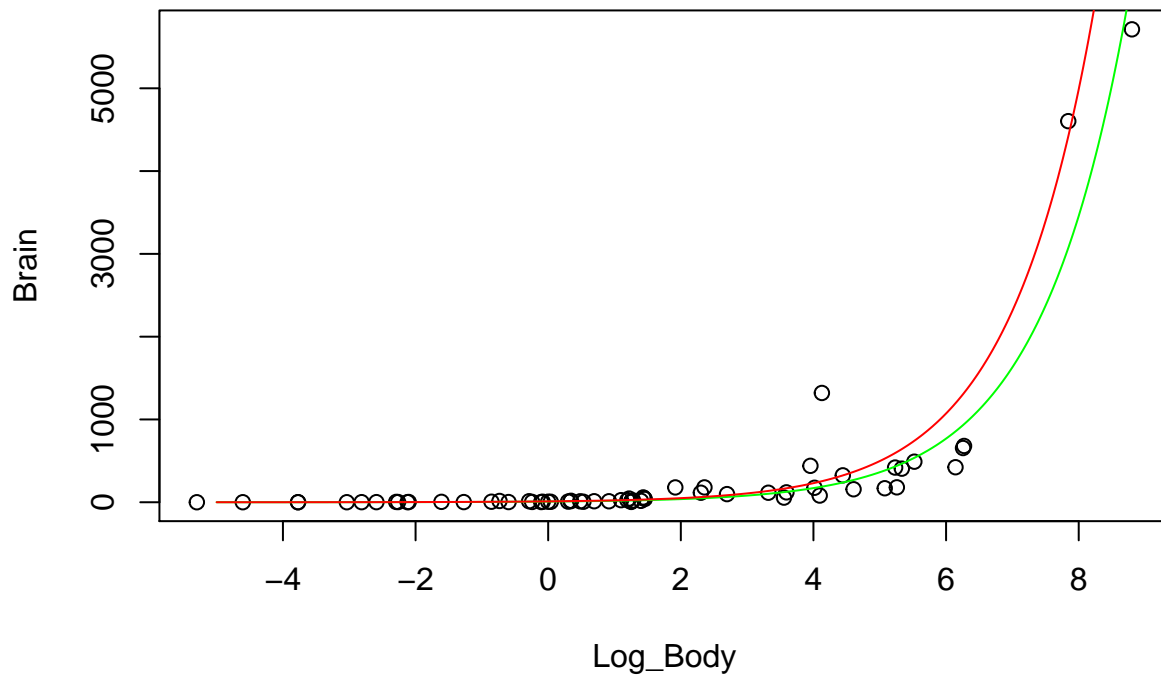
## Comparing m3 and best model with log transformation

```
m3:brain ~ log(body)
best_log: Log_Brain ~ Log_Body
```

```
bestmodel
```

```
##
## Call:
## lm(formula = Log_Brain ~ Log_Body)
##
## Coefficients:
## (Intercept)      Log_Body
##      2.1348      0.7517

beta_best=as.numeric(coef(bestmodel))
y_predictedbest=exp(beta_best[1]+x*beta_best[2])
plot(Log_Body,Brain)
lines(x,y_predictedbest,col='green')
lines(x,y_predicted3,col='red')
```



```
summary(bestmodel)
```

```
##
## Call:
## lm(formula = Log_Brain ~ Log_Body)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71550 -0.49228 -0.06162  0.43597  1.94829
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.13479    0.09604   22.23  <2e-16 ***
## Log_Body     0.75169    0.02846   26.41  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6943 on 60 degrees of freedom
## Multiple R-squared:  0.9208, Adjusted R-squared:  0.9195
## F-statistic: 697.4 on 1 and 60 DF,  p-value: < 2.2e-16
```

```
summary(m3)
```

```
##
## Call:
## glm(formula = brain ~ log(body), family = Gamma(link = "log"))
##
```



```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4889  -0.6204  -0.3089   0.1794   2.2555
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.36702    0.12116   19.54  <2e-16 ***
## log(body)    0.76846    0.03591   21.40  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.7671765)
##
##      Null deviance: 310.710  on 61  degrees of freedom
## Residual deviance:  31.578  on 60  degrees of freedom
## AIC: 534.72
##
## Number of Fisher Scoring iterations: 6
m3_Rsqr=1-m3$deviance/m3$null.deviance
m3_Rsqr

## [1] 0.8983679
```

The R-squared value for Log-Log model is 0.9195, which is greater than that of model3 0.8983679. Thus we prefer the log-log model we fitted from part(a).

## Make prediction

Now suppose the average body weight of a male polar bear is 450 kg, the predicted brain weight using three models above are: ##f

```
Body_new=log(450)
Brain_predicted_loglog=exp(beta_best[1]+Body_new*beta_best[2])
Brain_predicted_model3=exp(beta3[1]+Body_new*beta3[2])
Brain_predicted_model4=(beta4[1]+Body_new*beta4[2])^-1
```

The predicted values from best model with log transformation,model3 and model4 are 834.659757, 1166.452784, and 212.44153 respectively. And as we discussed above according to AIC and adjusted R-square value, we would trust the best model with log transformation with predicted brain weight 834.659757.