MATH 523 - Assignment 1

Emir Sevinc 260682995

January 21, 2019

A1)

a)

We have that $H^T = (X(X^TX)^{-1}X^T)^T = (X^T)^T((X^TX)^{-1})^TX^T = X((X^TX)^T)^{-1}X^T = X(X^T(X^T)^T)^{-1}X^T = X(X^TX)^{-1}X^T = H$

and $H^2 = (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T) = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H$

So using those, we can write: $(I_n - H)^T = I_n^T - H^T = I_n - H$

and $(I_n - H)^2 = (I_n - H)(I_n - H) = I_n^2 - I_nH - I_nH + H^2 = I_n - H - H + H^2 = I_n - H - H + H = I_n - H$.

b)

This would be equivalent to showing $[X^TX - \frac{X^Tzz^TX}{z^Tz}] * [(X^TX)^{-1} + \frac{(X^TX)^{-1}X^Tzz^TX(X^TX)^{-1}}{z^T(I_n-H)z}] = I$, since if $A^{-1} = B$, then $AA^{-1} = AB \implies AB = I$

Multiplying out gives $[X^TX - \frac{X^Tzz^TX}{z^Tz}] * [(X^TX)^{-1} + \frac{(X^TX)^{-1}X^Tzz^TX(X^TX)^{-1}}{z^T(I_n-H)z}]$

$= X^TX(X^TX)^{-1} + \frac{X^TX(X^TX)^{-1}X^Tzz^TX(X^TX)^{-1}}{z^T(I_n-H)z} - \frac{X^Tzz^TX(X^TX)^{-1}}{z^Tz} - \frac{X^Tzz^TX(X^TX)^{-1}X^Tzz^TX(X^TX)^{-1}}{z^Tzz^T(I_n-H)z}$

$= I + \frac{X^Tzz^TX(X^TX)^{-1}}{z^T(I_n-H)z} - \frac{X^Tzz^TX(X^TX)^{-1}}{z^Tz} - \frac{X^Tzz^TX(X^TX)^{-1}X^Tzz^TX(X^TX)^{-1}}{z^Tzz^T(I_n-H)z}$

Grouping under a common denominator:

$= I + \frac{zz^TX^Tzz^TX(X^TX)^{-1} - X^Tzz^TX(X^TX)^{-1}z^T(I_n-H)z - X^Tzz^TX(X^TX)^{-1}X^Tzz^TX(X^TX)^{-1}}{z^Tzz^T(I_n-H)z}$

Facotring out $X^Tzz^TX(X^TX)^{-1}$; $= I + \frac{X^Tzz^TX(X^TX)^{-1}[zz^T - z^T(I_n-H)z - X^Tzz^TX(X^TX)^{-1}]}{z^Tzz^T(I_n-H)z}$

$= I + \frac{X^Tzz^TX(X^TX)^{-1}[zz^T - z^TI_nz + z^THz - X^Tzz^TX(X^TX)^{-1}]}{z^Tzz^T(I_n-H)z}$

$= I + \frac{X^Tzz^TX(X^TX)^{-1}[z^THz - X^Tzz^TX(X^TX)^{-1}]}{z^Tzz^T(I_n-H)z}$

Redistributing $X^Tzz^TX(X^TX)^{-1}$; $= I + \frac{X^Tzz^TX(X^TX)^{-1}z^THz - X^Tzz^TX(X^TX)^{-1}X^Tzz^TX(X^TX)^{-1}}{z^Tzz^T(I_n-H)z}$

Now using that $H = (X(X^TX)^{-1}X^T)$, this becomes $I + \frac{X^Tzz^TX(X^TX)^{-1}z^THz - X^Tzz^THzz^TX(X^TX)^{-1}}{z^Tzz^T(I_n-H)z}$

Note that z^THz is a constant, so we get: $I + \frac{X^Tzz^TX(X^TX)^{-1}z^THz - X^Tzz^TX(X^TX)^{-1}z^THz}{z^Tzz^T(I_n-H)z} = I + 0 = I$

So we're done.

c)

Note that X^* is obtained by adding one additional column to X, and so can be written as $\begin{bmatrix} X & z \end{bmatrix}$. And since $H^* = X^*(X^{*T}X^*)^{-1}X^{*T}$ we can write H^* as

$\begin{bmatrix} X & z \end{bmatrix} [(X^{*T}X^*)^{-1}] \begin{bmatrix} X^T \\ z^T \end{bmatrix}$ Now we need an expression for $[(X^{*T}X^*)^{-1}]$. This would be:

$(X^{*T}X^*)^{-1} = (\begin{bmatrix} X & z \end{bmatrix}^T \begin{bmatrix} X & z \end{bmatrix})^{-1} = (\begin{bmatrix} X^T \\ z^T \end{bmatrix}^T \begin{bmatrix} X & z \end{bmatrix})^{-1} = \begin{bmatrix} X^TX & X^Tz \\ z^TX & z^TZ \end{bmatrix}^{-1}$. We will now use the formula to invert a block matrix:

$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$ Using this gives:

$$\begin{bmatrix} X^T X & X^T z \\ z^T X & z^T z \end{bmatrix}^{-1} =$$
$$\begin{bmatrix} (X^T X)^{-1} + (X^T X)^{-1} X^T z(z^T z - z^T X(X^T X)^{-1} X^T z)^{-1} z^T X(X^T X)^{-1} & -(X^T X)^{-1} X^T z(z^T z - z^T X(X^T X)^{-1} X^T z)^{-1} \\ -(z^T z - z^T X(X^T X)^{-1} X^T z)^{-1} z^T X(X^T X)^{-1} & (z^T z - z^T X(X^T X)^{-1} X^T z)^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} (X^T X)^{-1} + (X^T X)^{-1} X^T z(z^T z - z^T Hz)^{-1} z^T X(X^T X)^{-1} & -(X^T X)^{-1} X^T z(z^T z - z^T Hz)^{-1} \\ -(z^T z - z^T Hz)^{-1} z^T X(X^T X)^{-1} & (z^T z - z^T Hz)^{-1} \end{bmatrix}$$
$$= \begin{bmatrix} (X^T X)^{-1} + (X^T X)^{-1} X^T z(z^T (I_n - H)z)^{-1} z^T X(X^T X)^{-1} & -(X^T X)^{-1} X^T z(z^T (I_n - H)z)^{-1} \\ -(z^T (I_n - H)z)^{-1} z^T X(X^T X)^{-1} & (z^T (I_n - H)z)^{-1} \end{bmatrix}.$$ Using

that $z^T (I_n - H)z$ is a constant, we can move the inverses of those terms to the denominators, so we get:

$$= \begin{bmatrix} (X^T X)^{-1} + \frac{(X^T X)^{-1} X^T zz^T X(X^T X)^{-1}}{z^T (I_n - H)z} & \frac{-(X^T X)^{-1} X^T z}{z^T (I_n - H)z} \\ \frac{-z^T X(X^T X)^{-1}}{z^T (I_n - H)z} & \frac{1}{z^T (I_n - H)z} \end{bmatrix}$$ Now we can write out the final expression,

$$H^* = X^*(X^{*T} X^*)^{-1} X^{*T} = \begin{bmatrix} X & z \end{bmatrix} \begin{bmatrix} (X^T X)^{-1} + \frac{(X^T X)^{-1} X^T zz^T X(X^T X)^{-1}}{z^T (I_n - H)z} & \frac{-(X^T X)^{-1} X^T z}{z^T (I_n - H)z} \\ \frac{-z^T X(X^T X)^{-1}}{z^T (I_n - H)z} & \frac{1}{z^T (I_n - H)z} \end{bmatrix} \begin{bmatrix} X^T \\ z^T \end{bmatrix}$$

$$= \begin{bmatrix} X(X^T X)^{-1} + \frac{X(X^T X)^{-1} X^T zz^T X(X^T X)^{-1}}{z^T (I_n - H)z} + \frac{-zz^T X(X^T X)^{-1}}{z^T (I_n - H)z} & -\frac{X(X^T X)^{-1} X^T z}{z^T (I_n - H)z} + \frac{z}{z^T (I_n - H)z} \end{bmatrix} \begin{bmatrix} X^T \\ z^T \end{bmatrix}$$

$$= \begin{bmatrix} X(X^T X)^{-1} + \frac{Hzz^T X(X^T X)^{-1}}{z^T (I_n - H)z} + \frac{-zz^T X(X^T X)^{-1}}{z^T (I_n - H)z} & \frac{-Hz}{z^T (I_n - H)z} + \frac{z}{z^T (I_n - H)z} \end{bmatrix} \begin{bmatrix} X^T \\ z^T \end{bmatrix}$$

$$= X(X^T X)^{-1} X^T + \frac{Hzz^T X(X^T X)^{-1} X^T}{z^T (I_n - H)z} + \frac{-zz^T X(X^T X)^{-1} X^T}{z^T (I_n - H)z} + \frac{-Hzz^T}{z^T (I_n - H)z} + \frac{zz^T}{z^T (I_n - H)z}$$

$$= H + \frac{Hzz^T H}{z^T (I_n - H)z} + \frac{-zz^T H}{z^T (I_n - H)z} + \frac{-Hzz^T}{z^T (I_n - H)z} + \frac{zz^T}{z^T (I_n - H)z}$$

$$= H + \frac{Hzz^T H - zz^T H - Hzz^T + zz^T}{z^T (I_n - H)z}.$$ Note that the expression we're trying to get is $H + \frac{(I_n - H)zz^T (I_n - H)}{z^T (I_n - H)z} =$

$H + \frac{zz^T - Hzz^T (I_n - H)}{z^T (I_n - H)z} = H + \frac{zz^T - Hzz^T - zz^T H + Hzz^T H}{z^T (I_n - H)z}$. These two are equal, so we're done.

d)

Note that $\phi^2 = \frac{(z^{*T} R_0)^2}{(z^{*T} z^*)(R_0^T R_0)}$. We're meant to show that this is equal to $\frac{SSE_0 - SSE_1}{SSE_0}$.

In matrix form: $SSE_0 = Y^T (I_n - H)Y$, $SSE_1 = Y^T (I_n - H^*)Y$

First the numerator: $z^{*T} R_0 = [(I_n - H)z]^T (I_n - H)Y = z^T (I_n - H)^T (I_n - H)Y = z^T (I_n - H)(I_n - H)Y = z^T (I_n - H)^2 Y = z^T (I_n - H)Y$. But this is just a number so it is equal to its transpose: $Y^T (I_n - H)z$. Therefore $(z^{*T} R_0)^2 = (z^{*T} R_0)(z^{*T} R_0)^T = Y^T (I_n - H)zz^T (I_n - H)Y$. For the denominator, we have, $(z^{*T} z^*)(R_0^T R_0) = (z^T (I_n - H)^T (I_n - H)z)(Y^T (I_n - H)^T (I_n - H)Y) = (z^T (I_n - H)(I_n - H)z)(Y^T (I_n - H)(I_n - H)Y) = z^T (I_n - H)zY^T (I_n - H)Y$. Therefore, we have that $\phi^2 = \frac{(z^{*T} R_0)^2}{(z^{*T} z^*)(R_0^T R_0)} = \frac{Y^T (I_n - H)zz^T (I_n - H)Y}{z^T (I_n - H)zY^T (I_n - H)Y}$

Now: $SSE_1 = Y^T (I_n - H^*)Y = Y^T (I_n - H - \frac{(I_n - H)zz^T (I_n - H)}{z^T (I_n - H)z})Y$ by part c). Now let $k = \frac{(I_n - H)zz^T (I_n - H)}{z^T (I_n - H)z}$

Then, $Y^T (I_n - H - \frac{(I_n - H)zz^T (I_n - H)}{z^T (I_n - H)z})Y = Y^T (I_n - H - k)Y = (Y^T - Y^T H - Y^T k)Y = Y^T Y - Y^T HY - Y^T kY = Y^T (Y - HY) - Y^T kY = Y^T (I_n - H)Y - Y^T kY$. Thus

$$= Y^T (I_n - H)Y - \frac{Y^T (I_n - H)zz^T (I_n - H)Y}{z^T (I_n - H)z} = SSE_0 - \frac{Y^T (I_n - H)zz^T (I_n - H)Y}{z^T (I_n - H)z}$$

So $\frac{SSE_0 - SSE_1}{SSE_0} = \frac{SSE_0 - SSE_0 + \frac{Y^T (I_n - H)zz^T (I_n - H)Y}{z^T (I_n - H)z}}{SSE_0} = \frac{\frac{Y^T (I_n - H)zz^T (I_n - H)Y}{z^T (I_n - H)z}}{SSE_0} = \frac{Y^T (I_n - H)zz^T (I_n - H)Y}{z^T (I_n - H)zY^T (I_n - H)Y}$. Note that

this is precisely the expression we found above, so we have shown that $\phi^2 = \frac{SSE_0 - SSE_1}{SSE_0}$

A2)

a)

$$\frac{\Gamma(y+\theta_z)}{\Gamma(y+1)\Gamma(\theta_z)} * \left(\frac{\theta_z}{\mu+\theta_z}\right)^{\theta_z} * \left(\frac{\mu}{\mu+\theta_z}\right)^y$$
$$= exp[ylog(\frac{\mu}{\mu+\theta_z}) + \theta_z log(\frac{\theta_z}{\mu+\theta_z}) + log(\frac{\Gamma(y+\theta_z)}{\Gamma(y+1)\Gamma(\theta_z)})]$$

We have the form we wanted, so we have that $\theta = log(\frac{\mu}{\mu+\theta_z})$

$b(\theta) = -\theta_z log(\frac{\theta_z}{\mu+\theta_z})$, $a(\phi) = 1$ and $c(y,\theta) = +log(\frac{\Gamma(y+\theta_z)}{\Gamma(y+1)\Gamma(\theta_z)})$

We need to write b in terms of θ, and we can do this by writing the expression as

$$exp[ylog(\frac{\mu}{\mu+\theta_z}) + \theta_z log(\frac{\mu+\theta_z}{\mu+\theta_z} - \frac{\mu}{\mu+\theta_z}) + log(\frac{\Gamma(y+\theta_z)}{\Gamma(y+1)\Gamma(\theta_z)})]$$

$= exp[ylog(\frac{\mu}{\mu+\theta_z}) + \theta_z log(1 - \frac{\mu}{\mu+\theta_z}) + log(\frac{\Gamma(y+\theta_z)}{\Gamma(y+1)\Gamma(\theta_z)})]$. Since we have $\theta = log(\frac{\mu}{\mu+\theta_z})$, we will have $(\frac{\mu}{\mu+\theta_z}) = e^\theta$, thus $b(\theta) = -\theta_z log(\frac{\theta_z}{\mu+\theta_z})$ can be written as $b(\theta) = -\theta_z log(1 - e^\theta)$.

So we have identified all of our functions, and showed that it is an exponential family.

b)

$$E[Y] = b'(\theta) = -\theta_z * \frac{1}{1-e^\theta} * -e^\theta = \frac{\theta_z * e^\theta}{1-e^\theta}$$

We have that $\theta = log(\frac{\mu}{\mu+\theta_z})$, thus $E[Y] = \theta_z \frac{\frac{\mu}{\mu+\theta_z}}{1-\frac{\mu}{\mu+\theta_z}} = \frac{\frac{\theta_z * \mu}{\mu+\theta_z}}{\frac{\theta_z}{\mu+\theta_z}} = \mu$

For the variance we will need $b''(\theta) = \frac{\theta_z * e^\theta * (1-e^\theta) + \theta_z * e^\theta * e^\theta}{(1-e^\theta)^2} = \frac{\theta_z * e^\theta - \theta_z * e^\theta * e^\theta + \theta_z * e^\theta * e^\theta}{(1-e^\theta)^2} = \frac{\theta_z * e^\theta}{(1-e^\theta)^2}$

As $\theta = log(\frac{\mu}{\mu+\theta_z})$ this is equal to $\frac{\theta_z * \mu}{\mu+\theta_z} * \frac{(\mu+\theta_z)^2}{(\theta_z)^2} = \frac{\mu(\mu+\theta_z)}{\theta_z}$ and since $a(\phi) = 1$ we have $Var[Y] = \frac{\mu(\mu+\theta_z)}{\theta_z}$.

Since this is directly in terms of the mean μ the mean-variance relationship is simply $v(u) = \frac{\mu(\mu+\theta_z)}{\theta_z}$

c)

Since we found $b'(\theta)$ to be $\frac{\theta_z * e^\theta}{1-e^\theta}$ we need simply to invert this function. Let $y(x) = \frac{ke^x}{1-e^x}$ where k is placeholder for θ_z. To invert we need to swap the x and y and solve for y. Proceeding;
$x = \frac{k*e^y}{1-e^y} \implies x(1-e^y) = k*e^y \implies x - x*e^y = k*e^y \implies x = k*e^y + x*e^y \implies x = e^y(k+x) \implies$
$e^y = \frac{x}{k+x} \implies log(e^y) = log(\frac{x}{x+k}) \implies y = log(\frac{x}{x+k})$
So the canonical link function is $g(x) = log(\frac{x}{x+\theta_z})$. What this means is we will have $g(E[Y/X]) = X\beta \implies$
$E[Y/X] = g^{-1}X\beta \implies E[Y/X] = \frac{\theta_z * e^{X\beta}}{1-e^{X\beta}}$

The way this could go wrong is the following: since the distribution is negative binomial and corresponds to an actual quantity, we require that $E[Y/X]$ is positive. But if $X\beta$ is positive we can see from the expression that $E[Y/X]$ will end up being negative, or even undefinted if $X\beta$ is somehow 0, so these could prove problematic and it may not be appropriate to model with this function if $X\beta$'s are positive for the data set.

A3

a)

```
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
```

3

```
## Cow                 465.000 423.0
## Grey wolf            36.330 119.5
## Goat                 27.660 115.0
```

Trying a simple linear regression model:

```r
body<-mammals$body
brain<-mammals$brain
fit_data1<-lm(brain~body,data=mammals)
plot(body,brain,pch=19)
title(main = 'Simple Linear Regression Model')
abline(coef(fit_data1),col='red')
```

Simple Linear Regression Model
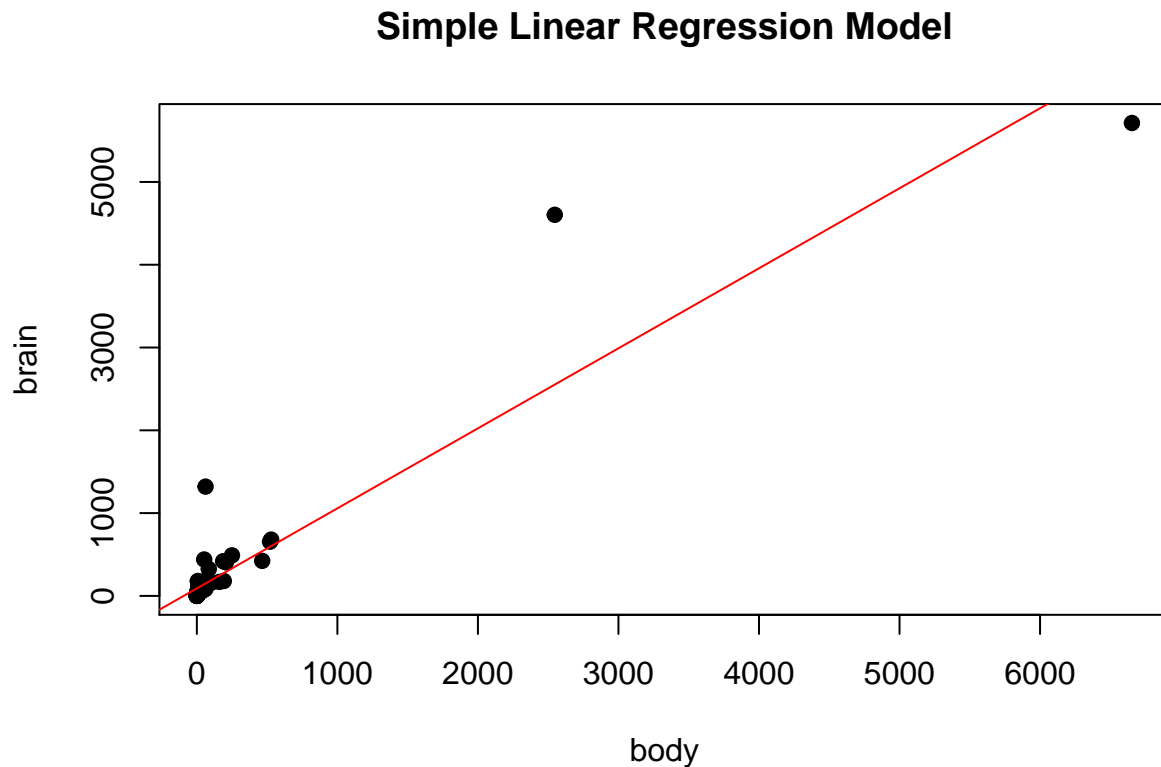

```r
summary(fit_data1)
```

```
##
## Call:
## lm(formula = brain ~ body, data = mammals)
##
## 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
```

There seems to be high variance of both the response and the regression, so let us try logging both.

```r
fit.Mammals_refit <- lm(log(brain) ~ log(body))
plot(log(body),log(brain),pch=19)
title(main = 'Simple Linear Regression Model, Logged Variables')
abline(coef(fit.Mammals_refit),col='red')
```

Simple Linear Regression Model, Logged Variables
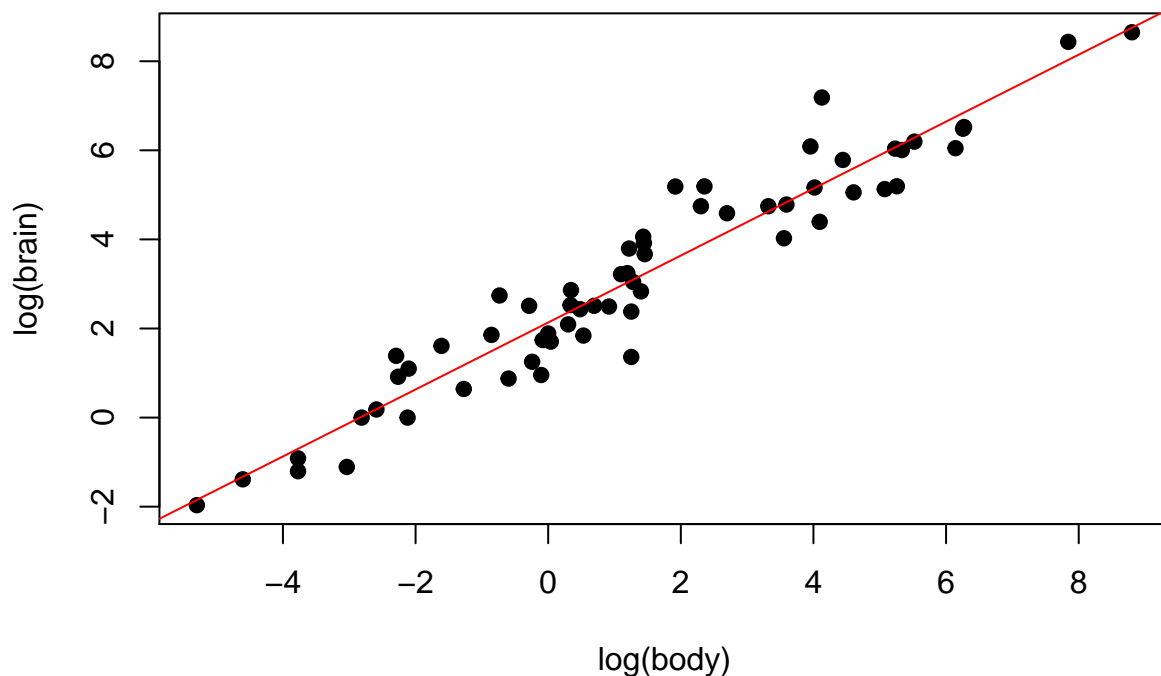

```r
summary(fit.Mammals_refit)
```

```
##
## 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
```

The fit appears vastly improved, with very low p values across the board that suggests that the parameters are significant, and a high R^2_{adj} value of 0.9195 suggesting that the model captures the variation well. Let's run diagnostics to confirm:
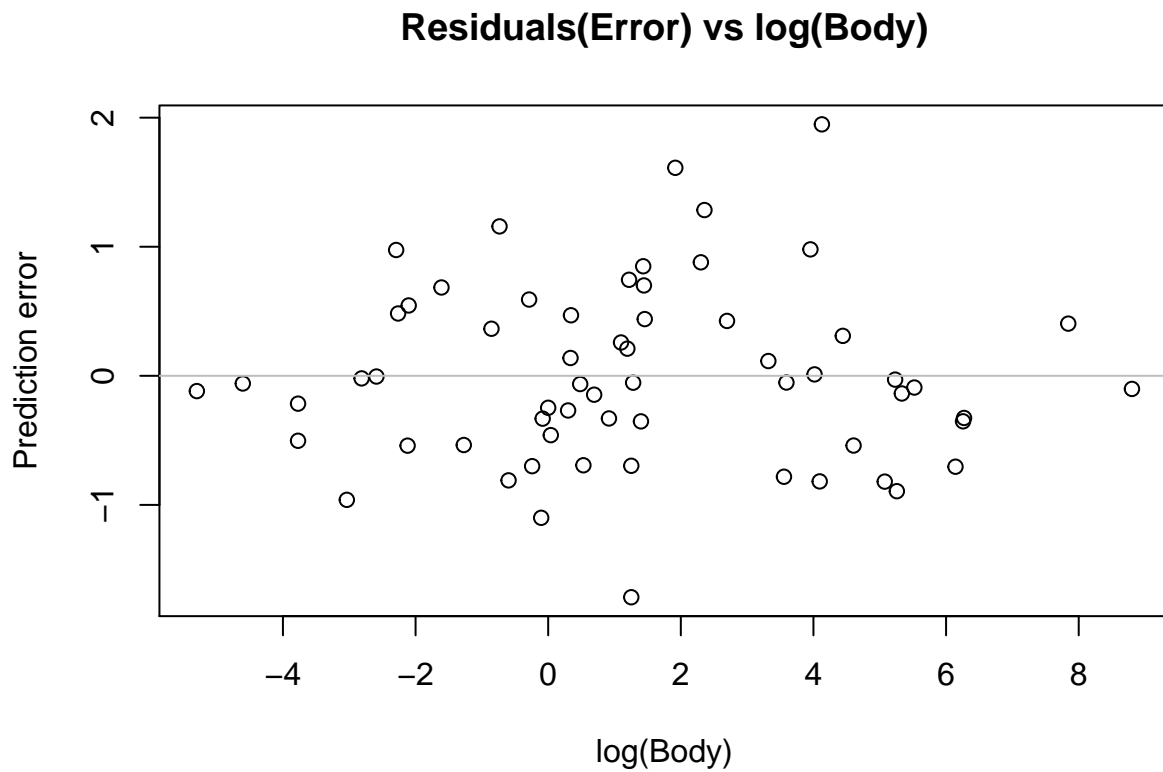
Residual plot:

```
data1_res = residuals(fit.Mammals_refit)
plot(data1_res,pch=19)
title(main = 'Residuals')
abline(h=0,col='red',lty=3)
```
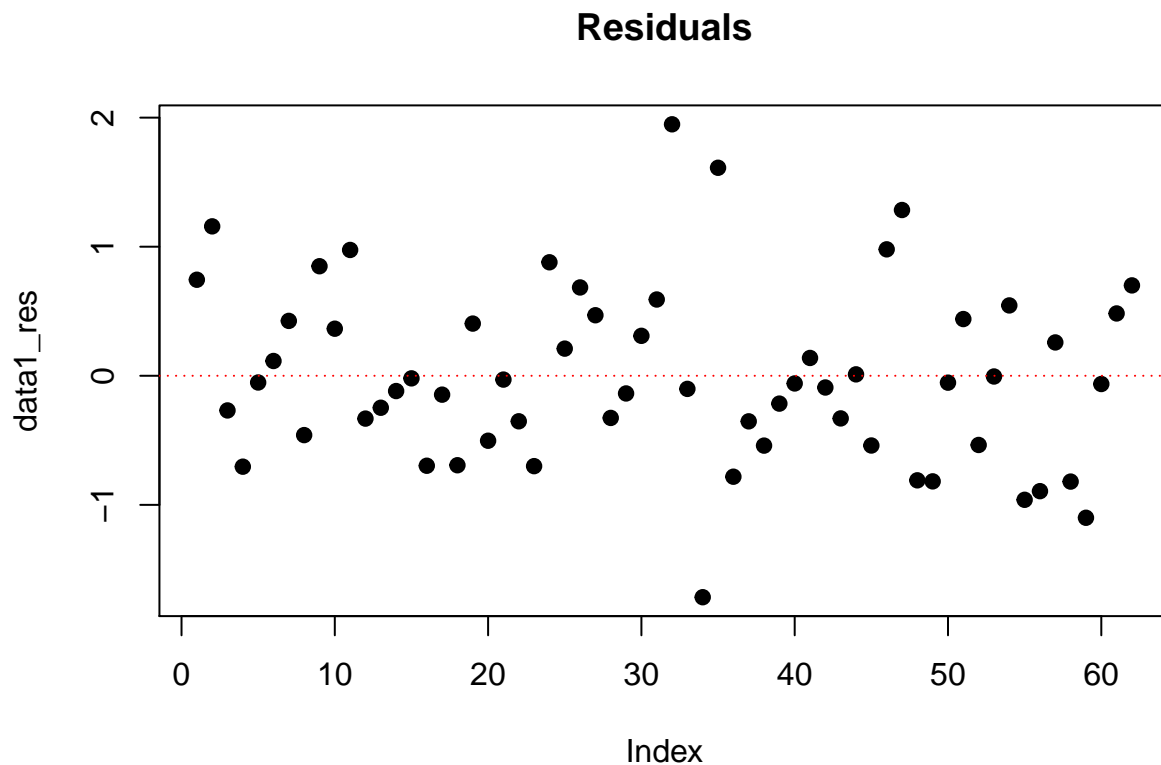

The residuals appear broadly centred about 0, but we can observe some "alternating"" pattern, so this is probably not a great sign.

Now we plot the residuals vs the regressor

```
plot(log(body), residuals(fit.Mammals_refit), main = "Residuals(Error) vs log(Body)",
xlab = "log(Body)", ylab = "Prediction error")
abline(h = 0, col = "grey")
```
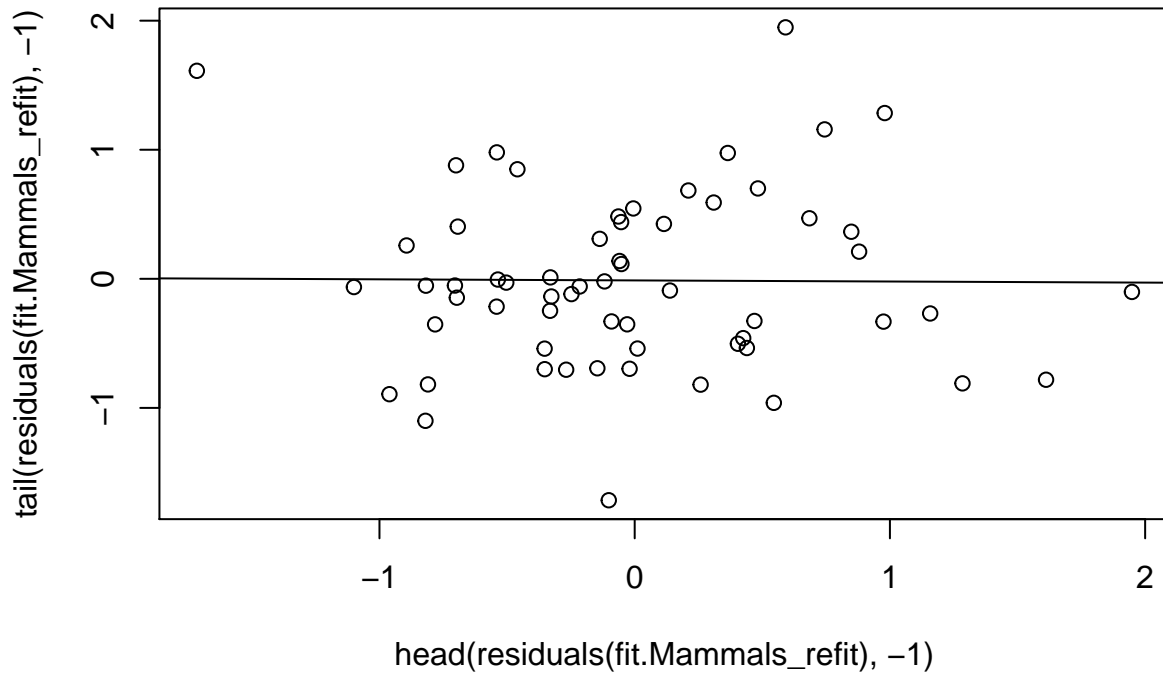
Residuals(Error) vs log(Body)

Not a "desired" picture; while this is centred around 0, it's not a blur and we have significant scattering.

Residuals vs Residuals:

```r
plot(head(residuals(fit.Mammals_refit), -1),
tail(residuals(fit.Mammals_refit), -1),main = "Residuals vs Residuals")
abline(lm(tail(residuals(fit.Mammals_refit),
-1) ~ head(residuals(fit.Mammals_refit),
-1)))
```
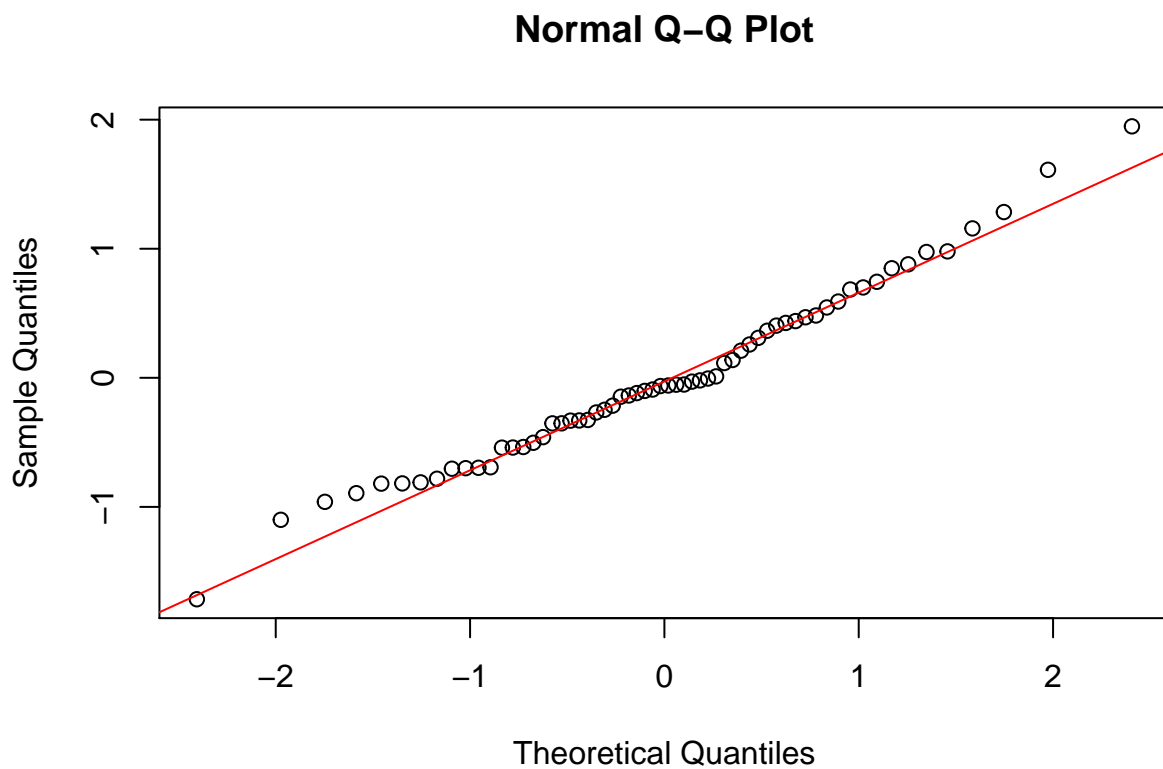
Residuals vs Residuals

This is no good, we would desire a blob with no patterning but we have once again a scattered plot here. Finally let's take a look at the residual distributions and see if we have normality:

```
qqnorm(residuals(fit.Mammals_refit))
qqline(residuals(fit.Mammals_refit), col='red')
```

Normal Q–Q Plot

Broadly normal, with some divergence at the beginning and the end, now we plot the residual vs a gaussian curve:

```r
hist(residuals(fit.Mammals_refit), breaks = 100,
freq = FALSE, xlab = "Residuals",
main = "Residual Distribution")
curve(dnorm(x, mean = 0, sd = sd(residuals(fit.Mammals_refit))),
add = TRUE, col = "blue")
```

Residual Distribution
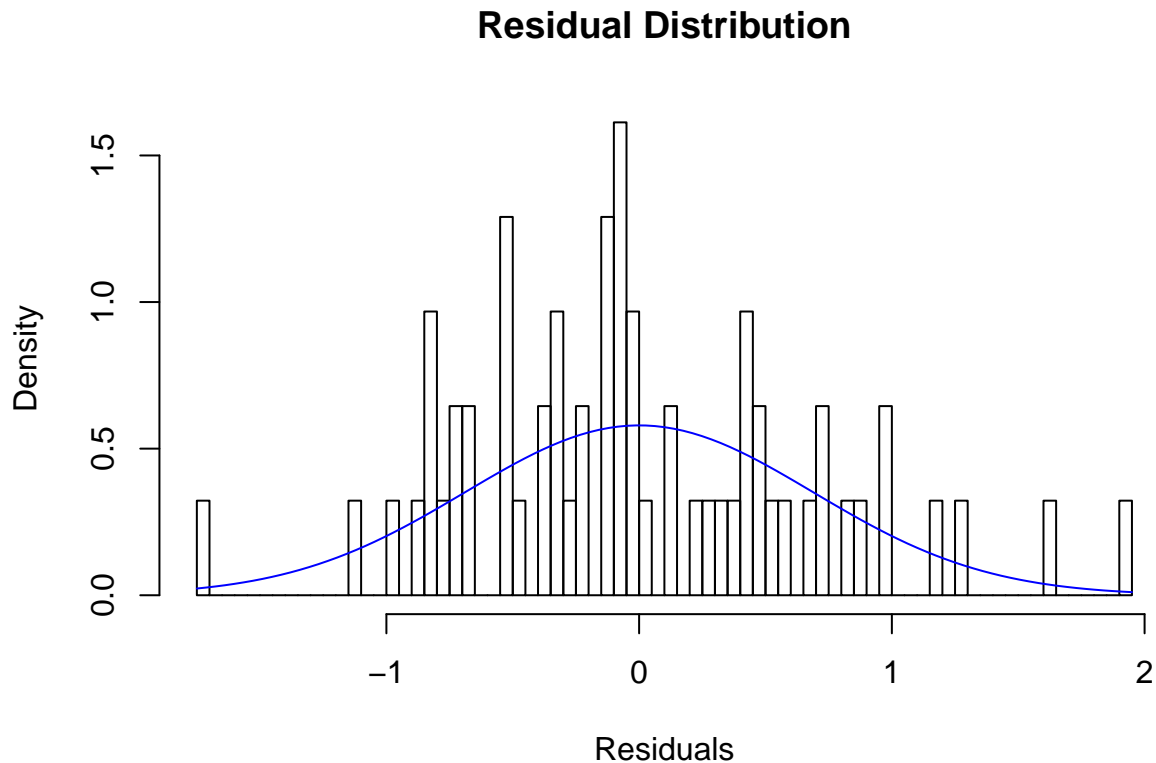
Conclusion: while visually the fit does not appear too bad, the residual diagnostics suggest that the model may not be optimal. Errors may not be fitting perfectly but we don't have many reasons to think that they aren't normal. Interpretation:

The fiited Model is: $log(Brain) = \beta_0 + \beta_1 * log(Body)$

That is, there is a statistically significant relationship between the logged values of brain and body. As the logged value of the body goes up by 1, the logged value of the brain can be expected to increase by 0.75169, the $\hat{\beta}_1$ parameter, or so the model says. The p values suggest that it is not likely that any of the parameters are 0 (so they're significant), and the R^2_{adj} suggest that the variation in the data is well captured.

b)

```
m1 <- lm(brain~body)
m2 <- glm(brain~body,family=gaussian(link="identity"))
summary(m1)
```

```
##
## 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
```

`summary(m2)`

```
##
## Call:
## glm(formula = brain ~ body, family = gaussian(link = "identity"))
##
## Deviance 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
##
## (Dispersion parameter for gaussian family taken to be 112037.3)
##
##     Null deviance: 52790554  on 61  degrees of freedom
## Residual deviance:  6722239  on 60  degrees of freedom
## AIC: 900.76
##
## Number of Fisher Scoring iterations: 2
```

i)

`coef(m1)`

```
## (Intercept)        body
##   91.0043962   0.9664964
```

`coef(m2)`

```
## (Intercept)        body
##   91.0043962   0.9664964
```

The coefficients appear to be identical, as are the standard errors.

ii)

we have that $334.7 = \frac{\hat{\sigma}\sqrt{n}}{\sqrt{n-2}}$ from the residual standard error: 334.7. Also $n - 2 = 60 \implies n = 62 \implies$ $334.7 = \frac{\hat{\sigma}\sqrt{62}}{\sqrt{60}} \implies \hat{\sigma} = 329.257$

The estimated dispersion parameter is 112037.3, note that $\sqrt{112037.3} = 334.7$, precisely equal to the residual standard error: 334.7 from the m1 output. So we will have $\sqrt{112037.3} = \frac{\hat{\sigma}*\sqrt{62}}{\sqrt{60}}$

11

iii)

```
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: brain
##           Df    Sum Sq  Mean Sq F value    Pr(>F)
## body       1 46068314 46068314  411.19 < 2.2e-16 ***
## Residuals 60  6722239   112037
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From this we have $SS_{Reg} = 46068314, SS_{Residual} = 6722239$ and $SS_{Total} = 6722239 + 46068314 = 52790553$. We can see that the "null deviance"" output from m2 is equal to the total sum of squares from m1, and the "residual deviance" outmput from m2 is equal to the residual sum of squares from the m1 output, and due to the usual parition ($SS_{Total} = SS_{Reg} + SS_{Residual}$) we will have that the sum of squares due to regression from m1 will be equal to the null deviance - residual deviance from the m2 output.

iv)

We have that the F statistic is equal to $\frac{SS_R/(p-1)}{SS_{RES}/n-p}$ where p is the number of parameters. But since we have found what SS_R and SS_{RES} are solely from the output of m2, we only need p and n. Since residual sum of squares will be on n-2 degrees of freedom and residual deviance from m2 is equal to this (and thus shares df), it is easily seen that n = 60+2 = 62. p can be seen to be equal to 2 simply due to the presence of 2 coefficient lines.

Thus, we have F=[(null deviance - residual deviance)/(2-1)]/[residual deviance/(62-2)] in the form of solely m2 output. Let's confirm numerically:

$F = \frac{(52790554-6722239)/(2-1)}{6722239/(62-2)} = 411.1$, which is precisely the F-statistic output that can be seen from m1.

c)

```
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
```

The estimated coefficients are 2.36702 for β_0 and 0.76846 β_1. The p values for the corresponding tests appear very small so the coefficients are are statistically significant. The inverse of log is the exponential function, so we will have that $E[Y/X] = g^{-1}[X\beta] = exp(X\beta)$. Here is what this means: Since the model is $E[Y/X] = exp(2.36702 + 0.76846 * X_{body})$, the XB term will increase by 0.76846 each time X, the (logged) body weight increases by 1, and this corresponds to, according to the model, an increase by a factor $e^{0.76846} = 2.15$. So every (logged) increase in body weight causes the brain size to increase 2.15 times.

d)

```
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
```

Plotting (blue corresponds to m3 and red corresponds to m4):

```
x <- seq(from=0,to=10,by=1)
plot(brain~log(body),xlim=c(0,10),ylim=c(0,5713),xlab="Body",ylab="Brain",pch=19)
lines(x,exp(coef(m3)[1]+coef(m3)[2]*x),col="blue")
lines(x,1/(coef(m4)[1]+coef(m4)[2]*x),col="red")
```
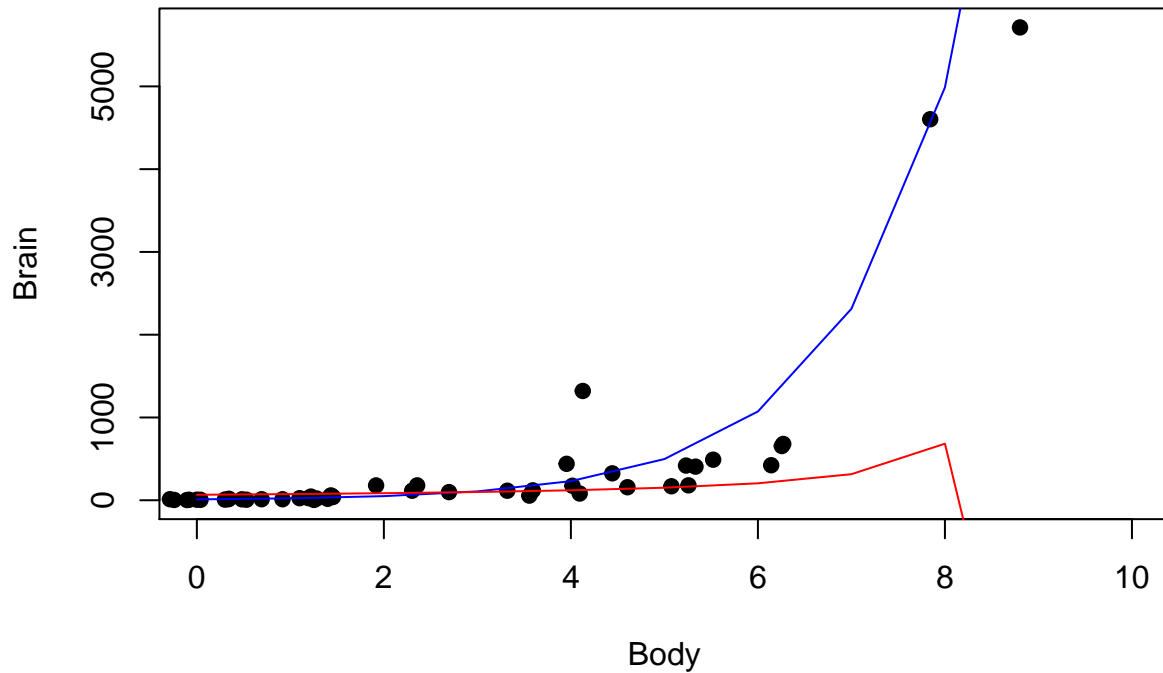
AIC Values for m3 and m4:

```
m3$aic
```

```
## [1] 534.7237
```

```
m4$aic
```

```
## [1] 658.9463
```

m4 is not preferable: the plot (red) seems to fail after a certain point (we can see it starts to go down and to negatives), and reciprocals are tricky due to potential singularities. In this case, we have that according to m4, $g(E[Y/X]) = X\beta \implies E[Y/X] = g^{-1}X\beta \implies E[Y/X] = \frac{1}{X\beta}$, thus we have that $E[Y/X] = \frac{1}{0.015187-0.001715*X}$. It is easy to see that at a certain point, the denominator could be equal to 0 (when $0.015187 = 0.001715 * X$), causing the expectation to be undefined.

In addition, if $0.001715 * X > 0.015187$, we would have a negative $E[Y/X]$ (we do not want this since Gamma is supposed to be modeling positive values) causing us to be unable to interpret it well.

Thus this model is not preferable if these conditons hold.

In addition, the plot of m3 (blue) seems to fit a lot better to the data points, and m3 has a smaller AIC value. All of these point to m3 being the more appropriate model.

e)

To get a ture comparison we need to apply m3 to the logged response (along with the regressor) but we can not do this since certain log(brain) values will become negative and the gamma distribution will not accept this.

The models do different things in the sense that the part a) model relates the logged response to the logged regressor, while m3) relates the response to the logged regressor, the latter which is better for interpretation; the part a) model only allows for an interpretation of how the logged value of the brain will change.

Having had to adjust the model less is a good sign too; m3 is able to capture the exponentially growing nature of the data without having to smooth it out first. Thus m3 would seem to be preferable.

f)

The predictions of m4, m3 and the part a) model are given respectively:

```r
(1/(coef(m4)[1]+coef(m4)[2]*log(450)))
```

```
## (Intercept)
##    212.4415
```

```r
(exp(coef(m3)[1]+coef(m3)[2]*log(450)))
```

```
## (Intercept)
##    1166.453
```

```r
(coef(fit.Mammals_refit)[1]+coef(fit.Mammals_refit)[2]*log(450))
```

```
## (Intercept)
##    6.727024
```

The m3's prediction takes the win here. We had discussed earlier how m3 was evidentally superior to m4, and due to it being a simple linear model we cant just take the exp of the prediction given by part a) and interpret it that way, so it only allows for an interpretation of how the logged value of the brain will change. Thus the m3 model's prediction is preferable.