

final_analysis.R

Feipeng Huang

2022-12-07

```
library(here)

## here() starts at /Users/stonehuang/Documents/environmental_data

delomys = read.csv(here("data", "delomys.csv"))

summary(delomys$body_mass)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      5.00   35.00   43.00   44.06   54.00   105.00

summary(delomys$body_length)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      49.0   106.0   116.0   114.7   124.0   265.0

shapiro.test(delomys$body_mass)

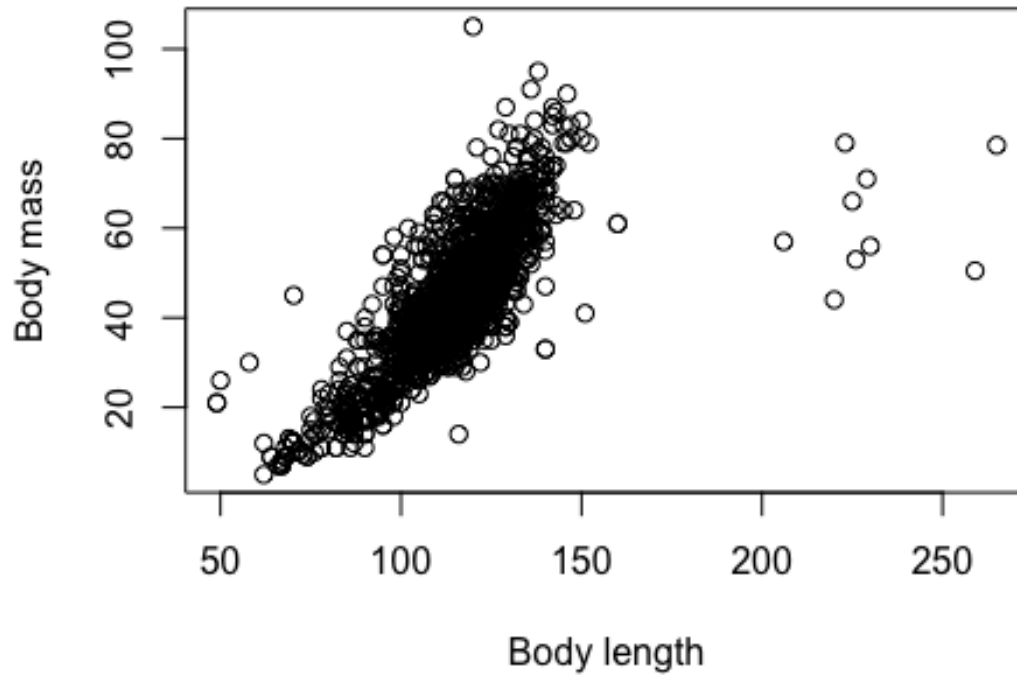
##
##  Shapiro-Wilk normality test
##
## data:  delomys$body_mass
## W = 0.99506, p-value = 4.33e-05

shapiro.test(delomys$body_length)

##
##  Shapiro-Wilk normality test
##
## data:  delomys$body_length
## W = 0.87609, p-value < 2.2e-16

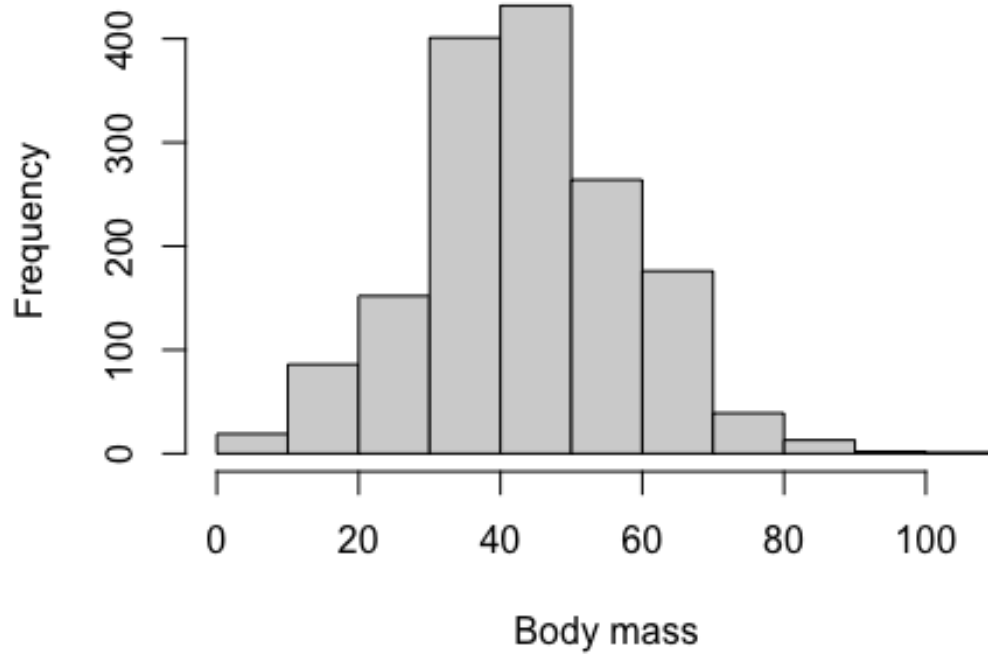
plot(body_mass ~ body_length, data = delomys, xlab = "Body length", ylab =
"Body mass", main = "Scatterplot of body mass and body length")
```

Scatterplot of body mass and body length



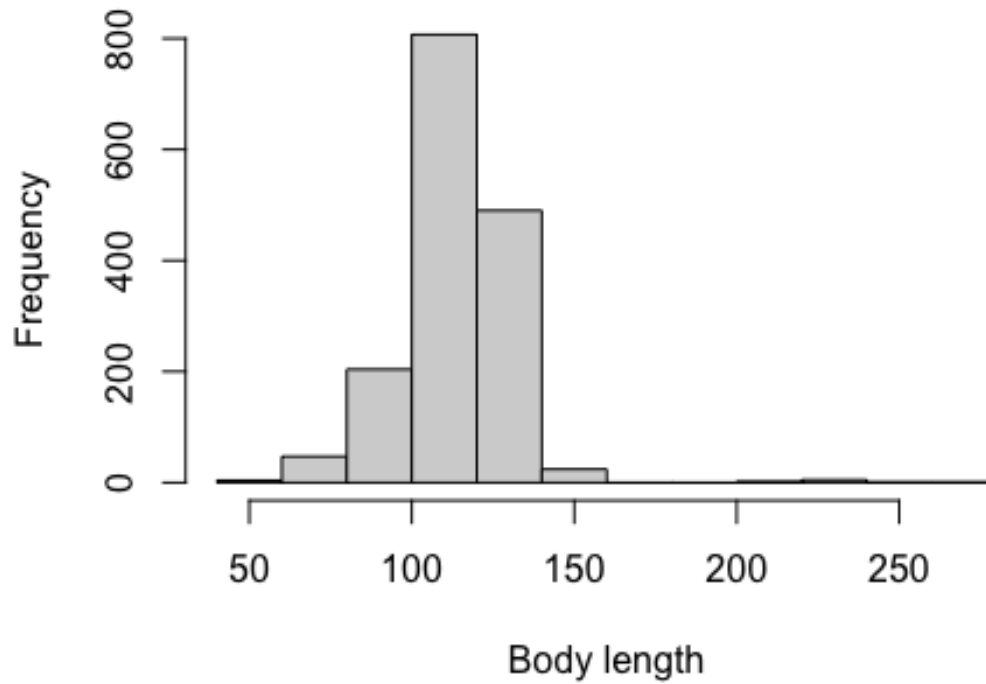
```
hist(delomys$body_mass, xlab = "Body mass", main = "Histogram of body mass")
```

Histogram of body mass



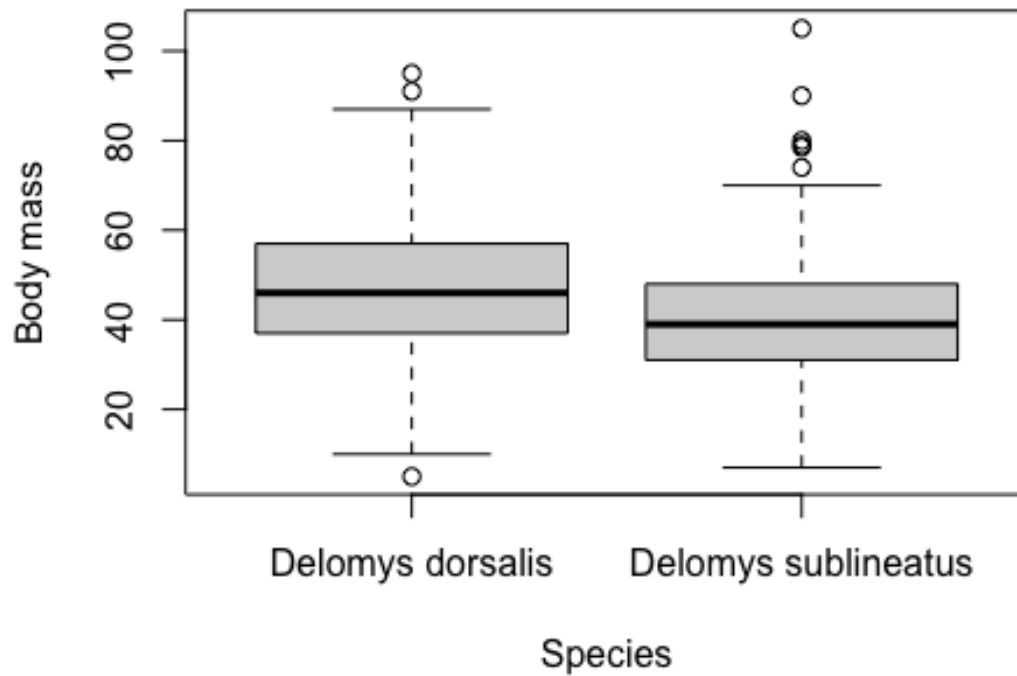
```
hist(delomys$body_length, xlab = "Body length", main = "Histogram of body length")
```

Histogram of body length



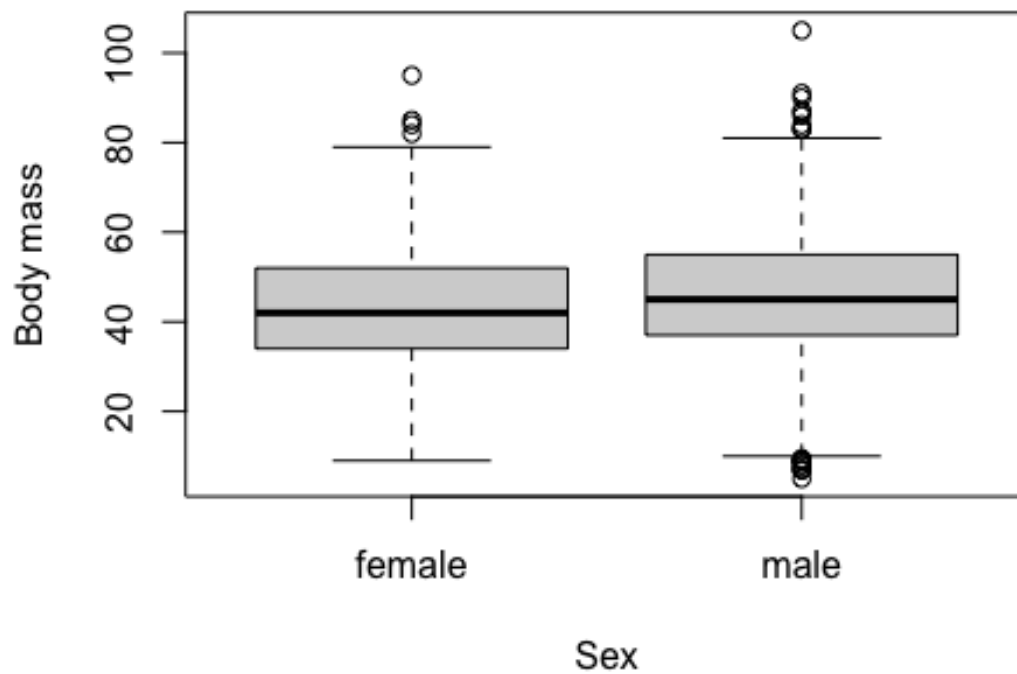
```
boxplot(body_mass ~ binomial, data = delomys, xlab = "Species", ylab = "Body mass", main = "Boxplot of body mass conditioned on species")
```

Boxplot of body mass conditioned on species



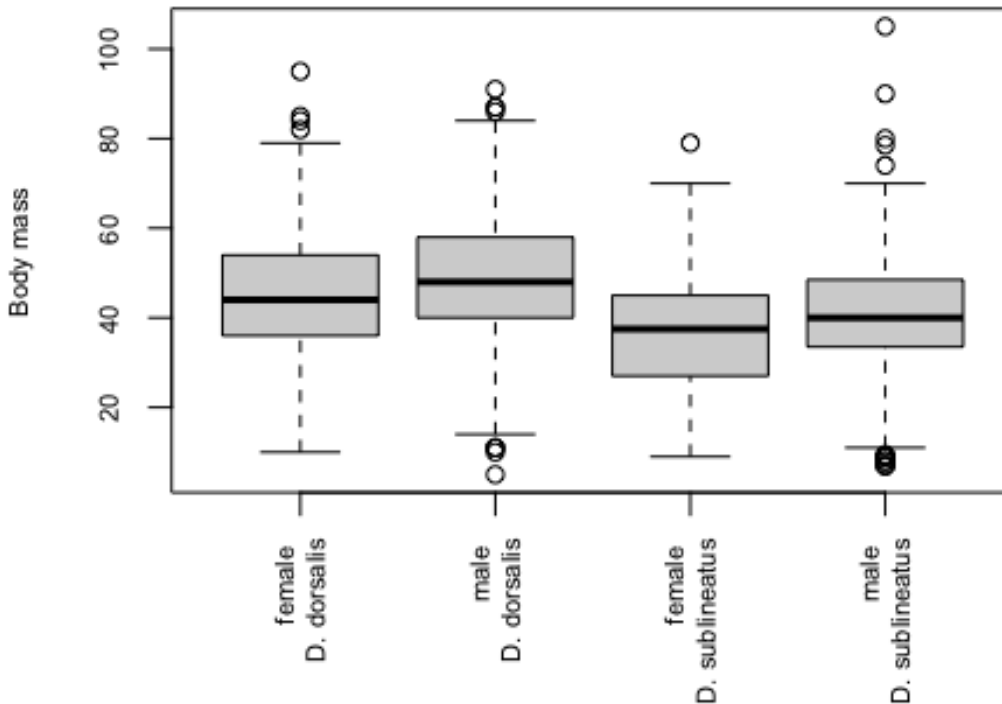
```
boxplot(body_mass ~ sex, data = delomys, xlab = "Sex", ylab = "Body mass",  
main = "Boxplot of body mass conditioned on sex")
```

Boxplot of body mass conditioned on sex



```
boxplot(body_mass ~ sex*binomial, data = delomys, las = 3, names = c("female \n D. dorsalis", "male \n D. dorsalis", "female \n D. sublineatus", "male \n D. sublineatus"), ylab = "Body mass", xlab = "", main = "Boxplot of mass conditioned on sex and species", cex.axis = 0.7, cex.lab = 0.7)
```

Boxplot of mass conditioned on sex and species



#Q1 The relationship between body mass and length seems to be linear. Mass is positive correlated with length.

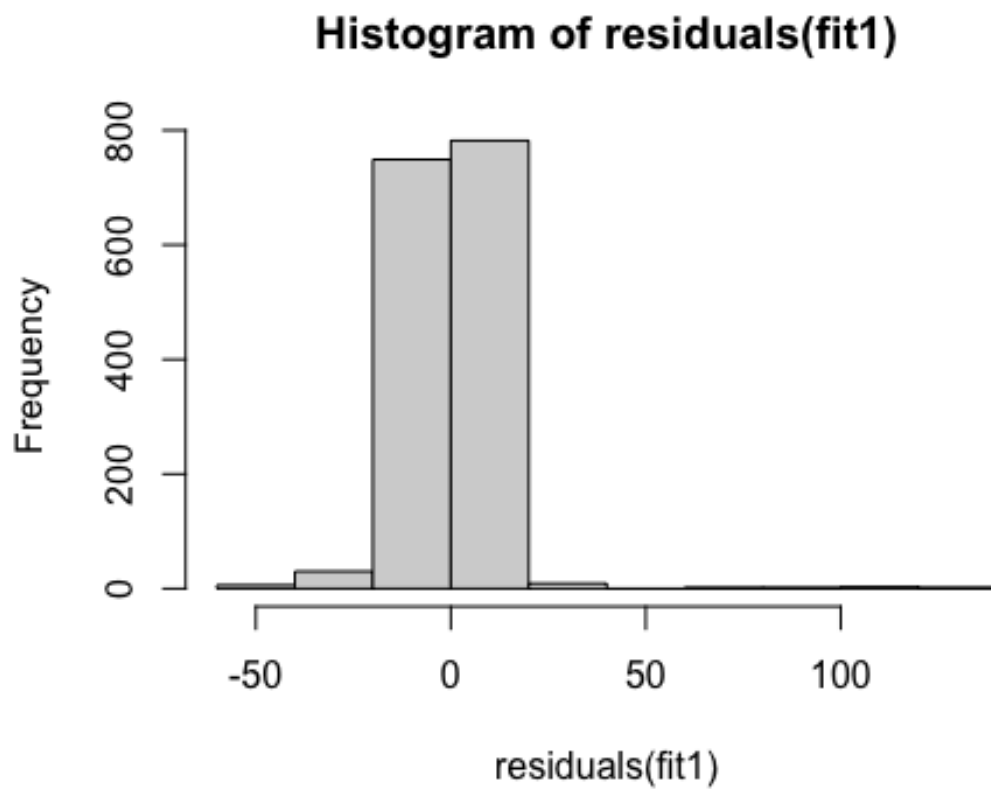
#Q2 The data, especially body length, does not appear normally-distributed because the histograms are not symmetrical.

#Q3 I think the (unconditioned) body masses and body length are not normally-distributed (histograms are skewed, p -values < 0.05). My visual assessment of normality matched the results of the numerical normality tests.

#Q4 *D. dorsalis* seems to weigh more than *D. sublineatus* and male seems to weigh more than female, but not by much. There is no significant difference.

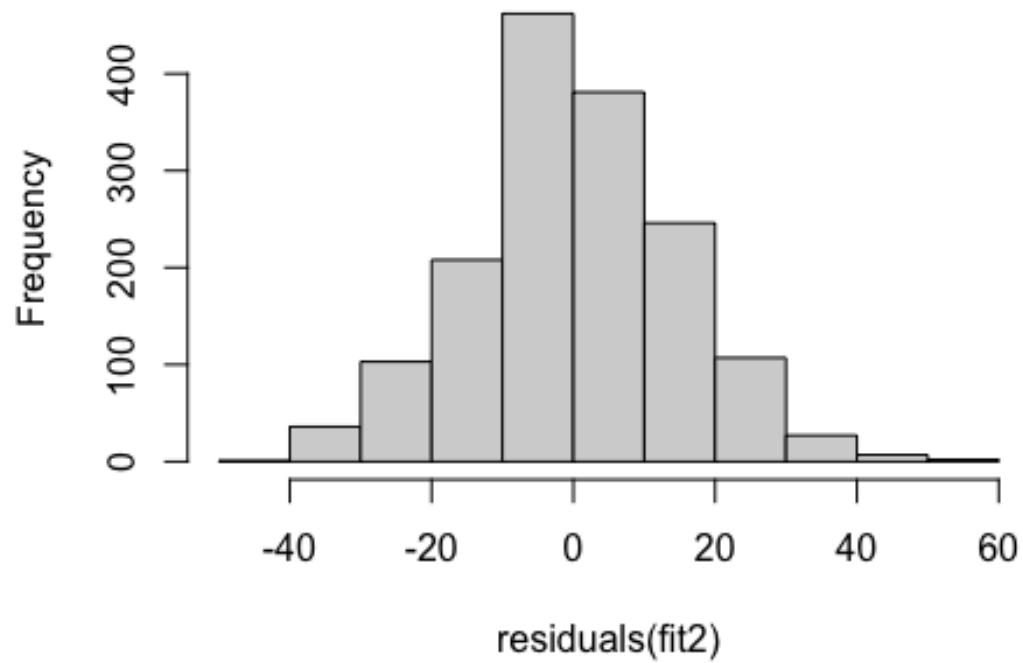
```
fit1 = lm(body_length ~ body_mass, data = delomys)
fit2 = lm(body_mass ~ sex, data = delomys)
fit3 = lm(body_mass ~ binomial, data = delomys)
fit4 = lm(body_mass ~ sex + binomial, data = delomys)
fit5 = lm(body_mass ~ sex * binomial, data = delomys)
```

```
hist(residuals(fit1))
```



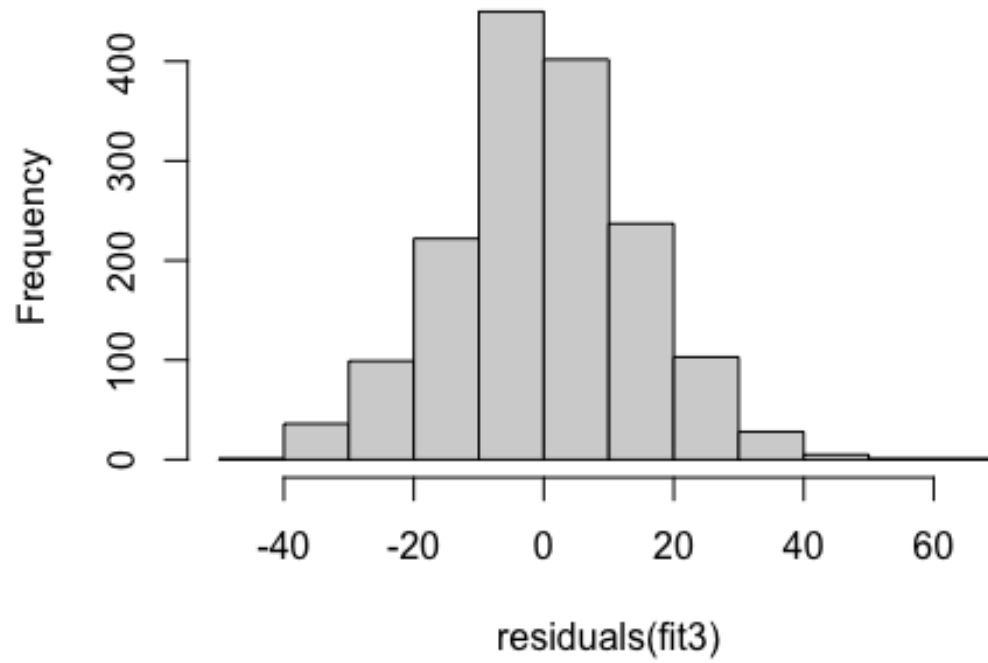
```
hist(residuals(fit2))
```


Histogram of residuals(fit2)



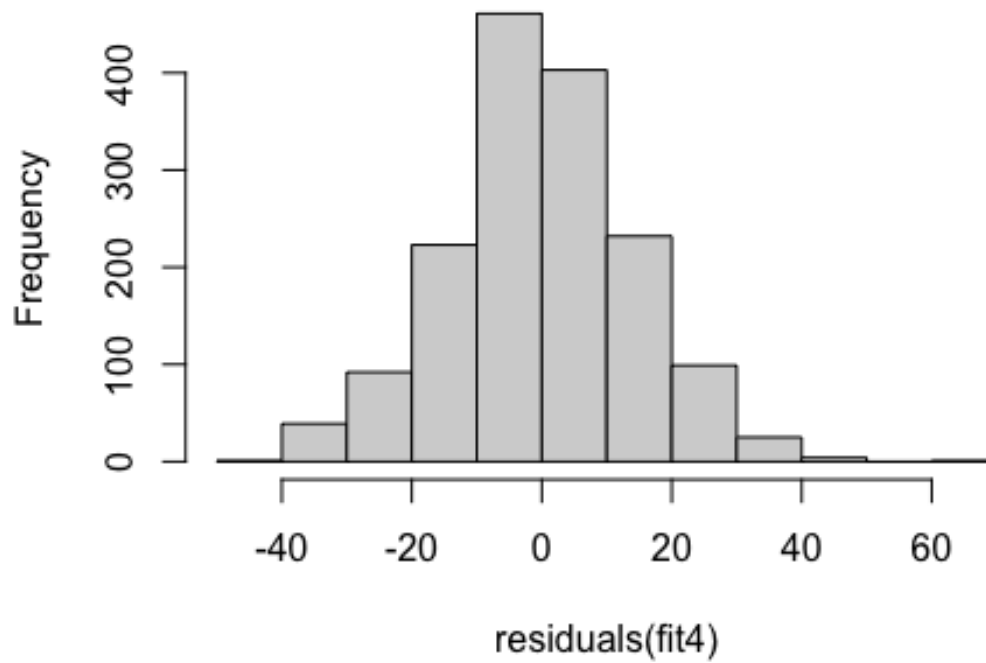
```
hist(residuals(fit3))
```

Histogram of residuals(fit3)



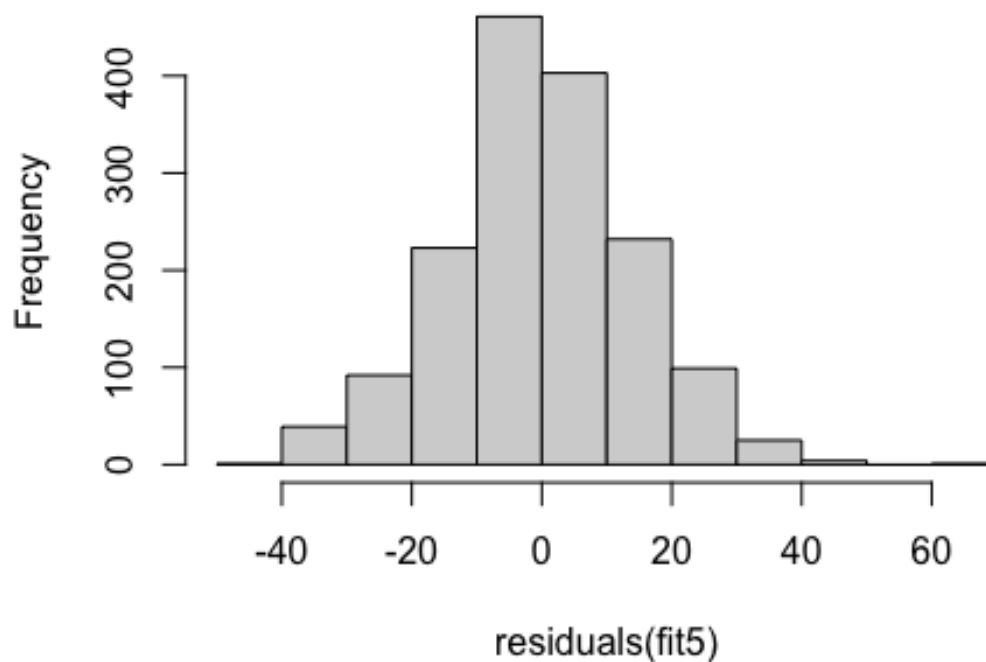
```
hist(residuals(fit4))
```

Histogram of residuals(fit4)



```
hist(residuals(fit5))
```

Histogram of residuals(fit5)



```
shapiro.test(residuals(fit1))  
  
##  
##  Shapiro-Wilk normality test  
##  
## data:  residuals(fit1)  
## W = 0.73462, p-value < 2.2e-16  
  
shapiro.test(residuals(fit2))  
  
##  
##  Shapiro-Wilk normality test  
##  
## data:  residuals(fit2)  
## W = 0.99563, p-value = 0.0001541  
  
shapiro.test(residuals(fit3))  
  
##  
##  Shapiro-Wilk normality test  
##  
## data:  residuals(fit3)  
## W = 0.99535, p-value = 8.103e-05
```

```
shapiro.test(residuals(fit4))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(fit4)
## W = 0.99525, p-value = 6.805e-05
```

```
shapiro.test(residuals(fit5))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(fit5)
## W = 0.99526, p-value = 6.816e-05
```

#Q5 Based on the numerical and graphical diagnostics, all models fail to fulfill the assumption of normality of the residuals.
#Q6 Violations of the normality assumption are not equally severe for all the models. The violation is much more severe in model 1.

```
knitr::kable(coef(summary(fit1)))
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	76.1246565	0.9158120	83.12258	0
body_mass	0.8754988	0.0196905	44.46298	0

#Q7 0.8754988

#Q8 163.6745

*76.1246565+(0.8754988*100)*

```
## [1] 163.6745
```

#Q9 76.12466

*76.1246565+(0.8754988*0)*

```
## [1] 76.12466
```

```
knitr::kable(coef(summary(fit2)))
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	42.711465	0.5288929	80.756355	0.0000000
sexmale	2.784133	0.7456117	3.734024	0.0001951

```
knitr::kable(coef(summary(fit3)))
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	46.752427	0.4499933	103.89582	0
binomialDelomys sublineatus	-7.683058	0.7604562	-10.10322	0

```
knitr::kable(coef(summary(fit4)))
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	45.070355	0.5556429	81.113891	0e+00
sexmale	3.795395	0.7259609	5.228099	2e-07
binomialDelomys sublineatus	-8.193492	0.7609985	-10.766764	0e+00
knitr::kable(coef(summary(fit5)))				

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	45.0554562	0.6052973	74.4352459	0.0000000
sexmale	3.8280908	0.8966667	4.2692460	0.0000208
binomialDelomys sublineatus	-8.1417394	1.1281046	-7.2171847	0.0000000
sexmale:binomialDelomys sublineatus	-0.0950187	1.5285856	-0.0621612	0.9504424
#Q10 female				
#Q11 Delomys dorsalis				
#Q12 male				
#Q13 Delomys dorsalis				
knitr::kable(anova(fit2))				

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sex	1	3061.678	3061.6779	13.94294	0.0001951
Residuals	1578	346507.131	219.5863	NA	NA
knitr::kable(anova(fit3))					

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
binomial	1	21289.68	21289.6805	102.0751	0
Residuals	1583	330164.45	208.5688	NA	NA
knitr::kable(anova(fit4))					

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sex	1	3061.678	3061.6779	14.95838	0.0001144
binomial	1	23727.134	23727.1342	115.92320	0.0000000
Residuals	1577	322779.997	204.6798	NA	NA
knitr::kable(anova(fit5))					

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sex	1	3.061678e+03	3.061678e+03	14.948932	0.0001150
binomial	1	2.372713e+04	2.372713e+04	115.849977	0.0000000
sex:binomial	1	7.913846e-01	7.913846e-01	0.003864	0.9504424
Residuals	1576	3.227792e+05	2.048091e+02	NA	NA
#Q14 Sex and species are significant predictors for body mass.					
#Q15 There is not a significant interaction.					

#Q16 The significance level of sex and the significance level of species do not change much among the different models. All corresponding p-values are very small.

```
AIC(fit2)
```

```
## [1] 13006.8
```

```
AIC(fit3)
```

```
## [1] 12966.36
```

```
AIC(fit4)
```

```
## [1] 12896.73
```

```
AIC(fit5)
```

```
## [1] 12898.72
```

#Q17 model 4 and 5

#Q18 I would select model 4 because it is the best fit (lowest AIC), and an additive model is easier to understand and explain than an interactive model.