

# Archosaurs: Linear Regression Example

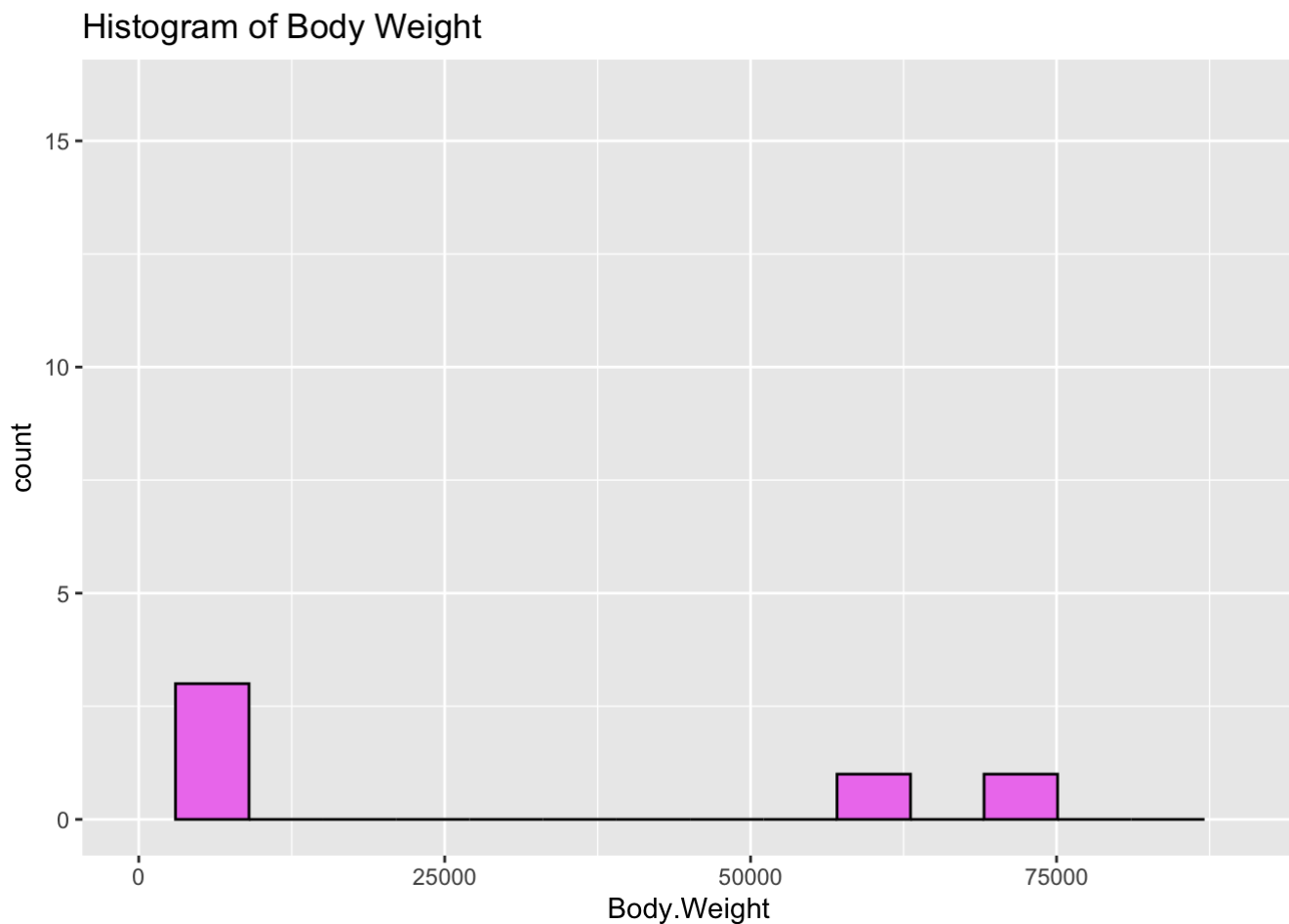
Katie, Rita, and Chang

10/6/23

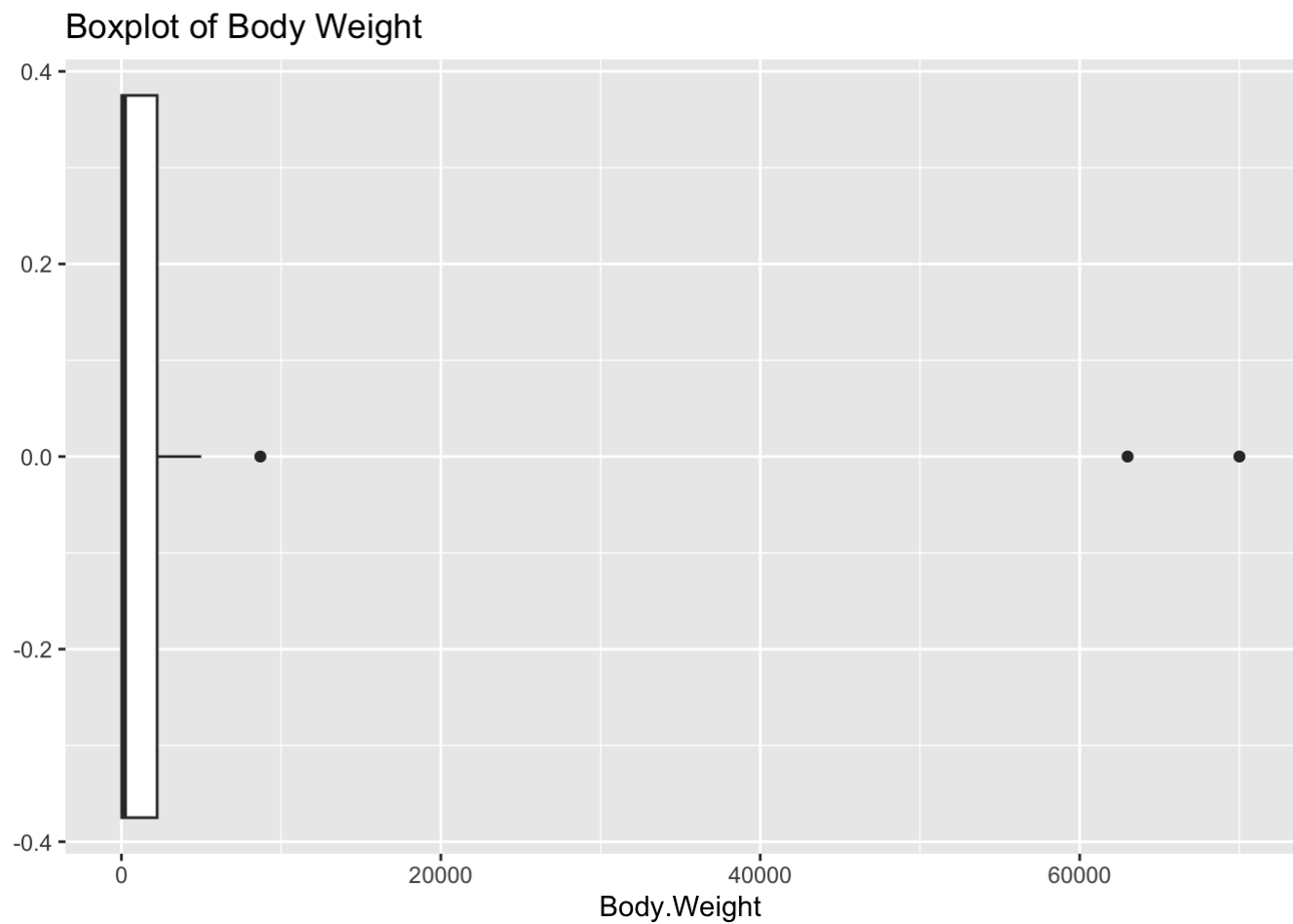
## Distribution of body weight

```
library("ggplot2")  
ggplot(brain, aes(x = Body.Weight)) + geom_histogram(bins = 16, color = "black", fill =  
"violet") + ggtitle("Histogram of Body Weight") + scale_x_continuous(limits = c(-100,900  
00))
```

```
## Warning: Removed 2 rows containing missing values (`geom_bar()`).
```



```
ggplot(brain, aes(x = Body.Weight)) + geom_boxplot() + ggtitle("Boxplot of Body Weight")
```



```
xbar <- mean(brain$Body.Weight)
xbar
```

```
## [1] 7472.371
```

```
sd <- sd(brain$Body.Weight)
sd
```

```
## [1] 19770.46
```

```
n = 21
standard_error_mean <- sd/sqrt(n)
standard_error_mean
```

```
## [1] 4314.268
```

```
margin <- qt(0.975,df=n-1)*sd/sqrt(n)
lowerinterval <- xbar - margin
lowerinterval
```

```
## [1] -1527.033
```

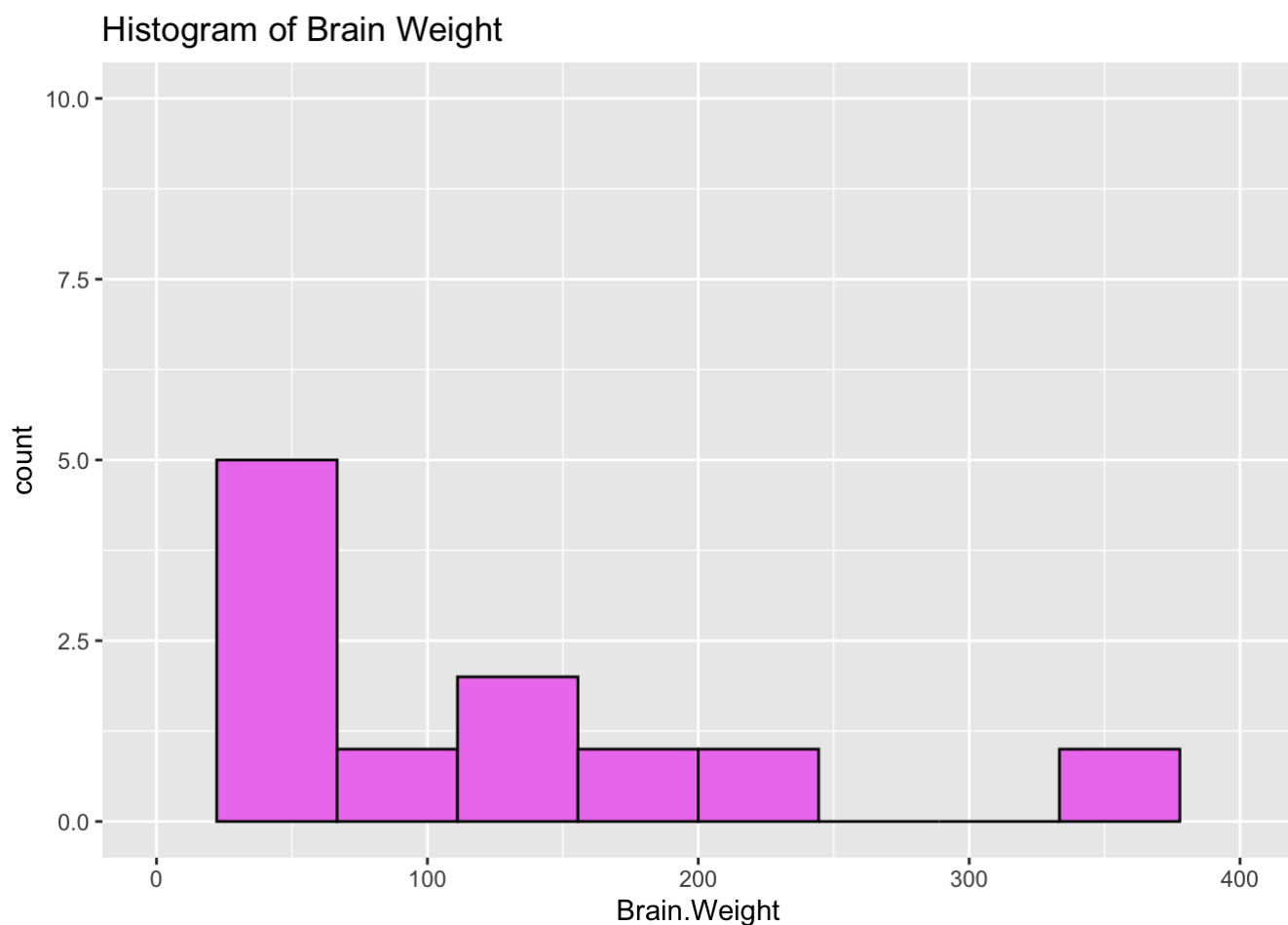
```
upperinterval <- xbar + margin  
upperinterval
```

```
## [1] 16471.78
```

## Distribution of brain weight

```
library("ggplot2")  
ggplot(brain, aes(x = Brain.Weight)) + geom_histogram(bins=10, color = "black", fill =  
"violet") + ggtitle("Histogram of Brain Weight") + scale_x_continuous(limits = c(0,400))
```

```
## Warning: Removed 2 rows containing missing values (`geom_bar()`).
```



```
ggplot(brain, aes(x = Brain.Weight)) + geom_boxplot() + ggtitle("Boxplot of Brain Weight")
```



```
xbar <- mean(brain$Brain.Weight)
xbar
```

```
## [1] 64.94086
```

```
sd <- sd(brain$Brain.Weight)
sd
```

```
## [1] 90.15867
```

```
n = 21
standard_error_mean <- sd/sqrt(n)
standard_error_mean
```

```
## [1] 19.67423
```

```
margin <- qt(0.975,df=n-1)*sd/sqrt(n)
lowerinterval <- xbar - margin
lowerinterval
```

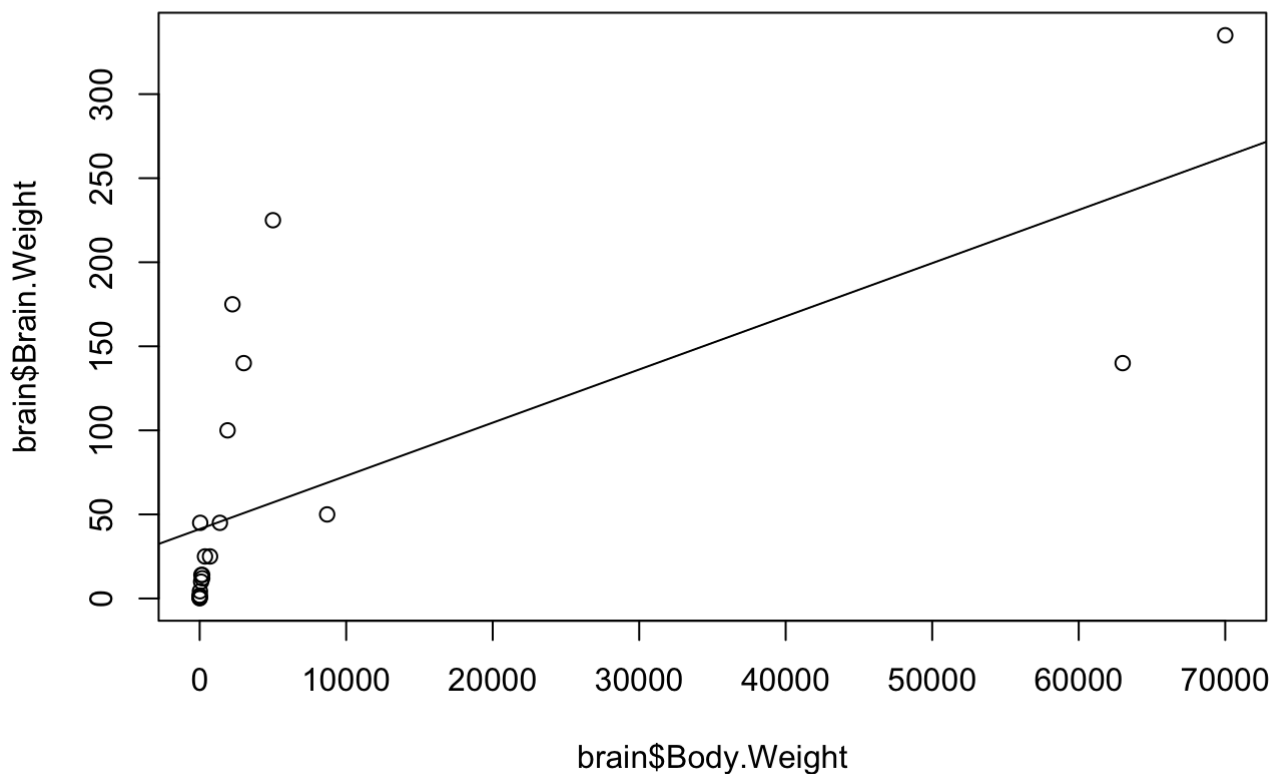
```
## [1] 23.90112
```

```
upperinterval <- xbar + margin  
upperinterval
```

```
## [1] 105.9806
```

## Regression with body weight and brain weight prior to transformation

```
plot(brain$Brain.Weight~brain$Body.Weight)  
abline(lm(brain$Brain.Weight~brain$Body.Weight))
```

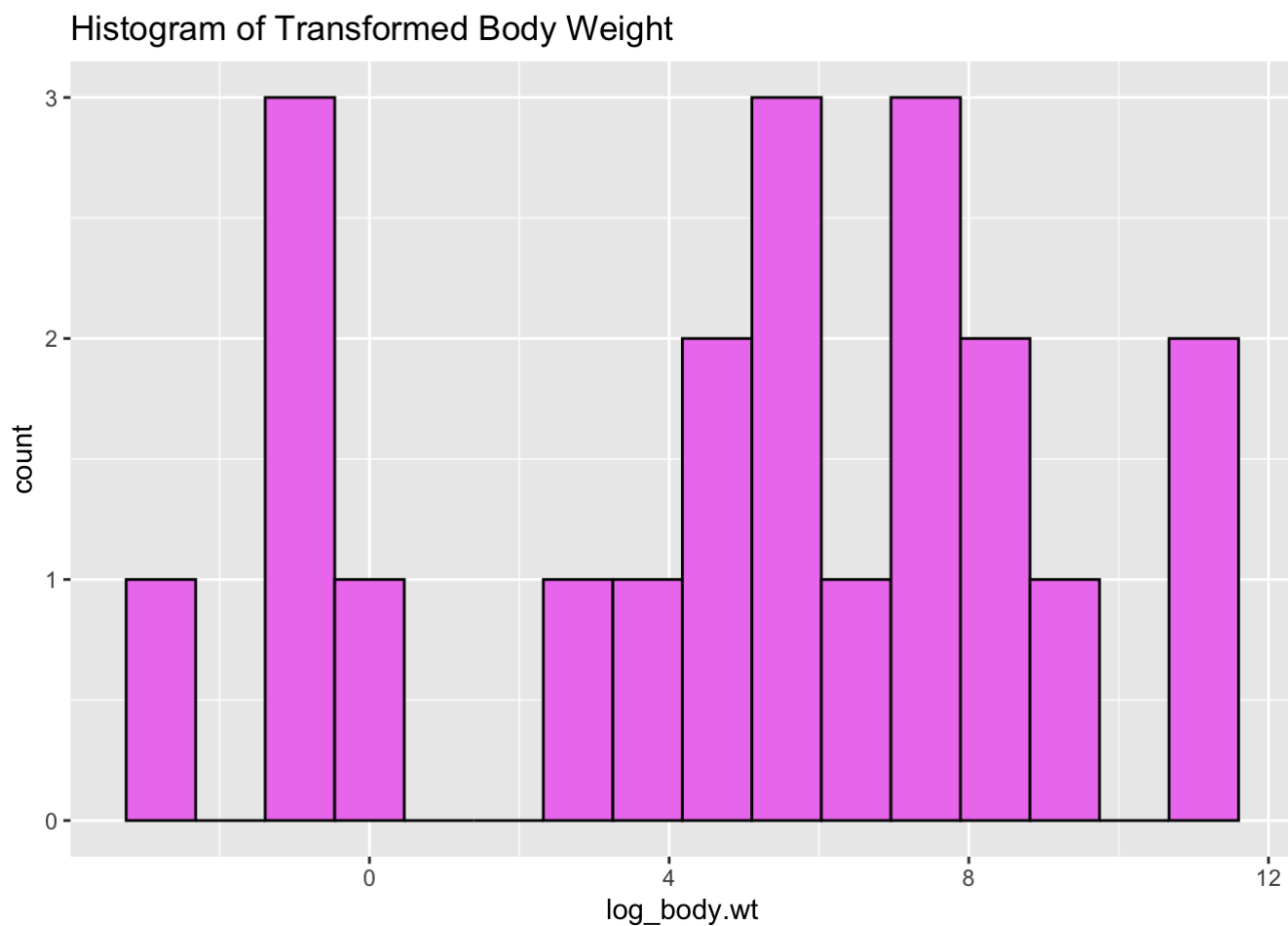


## Transform body weight and brain weight

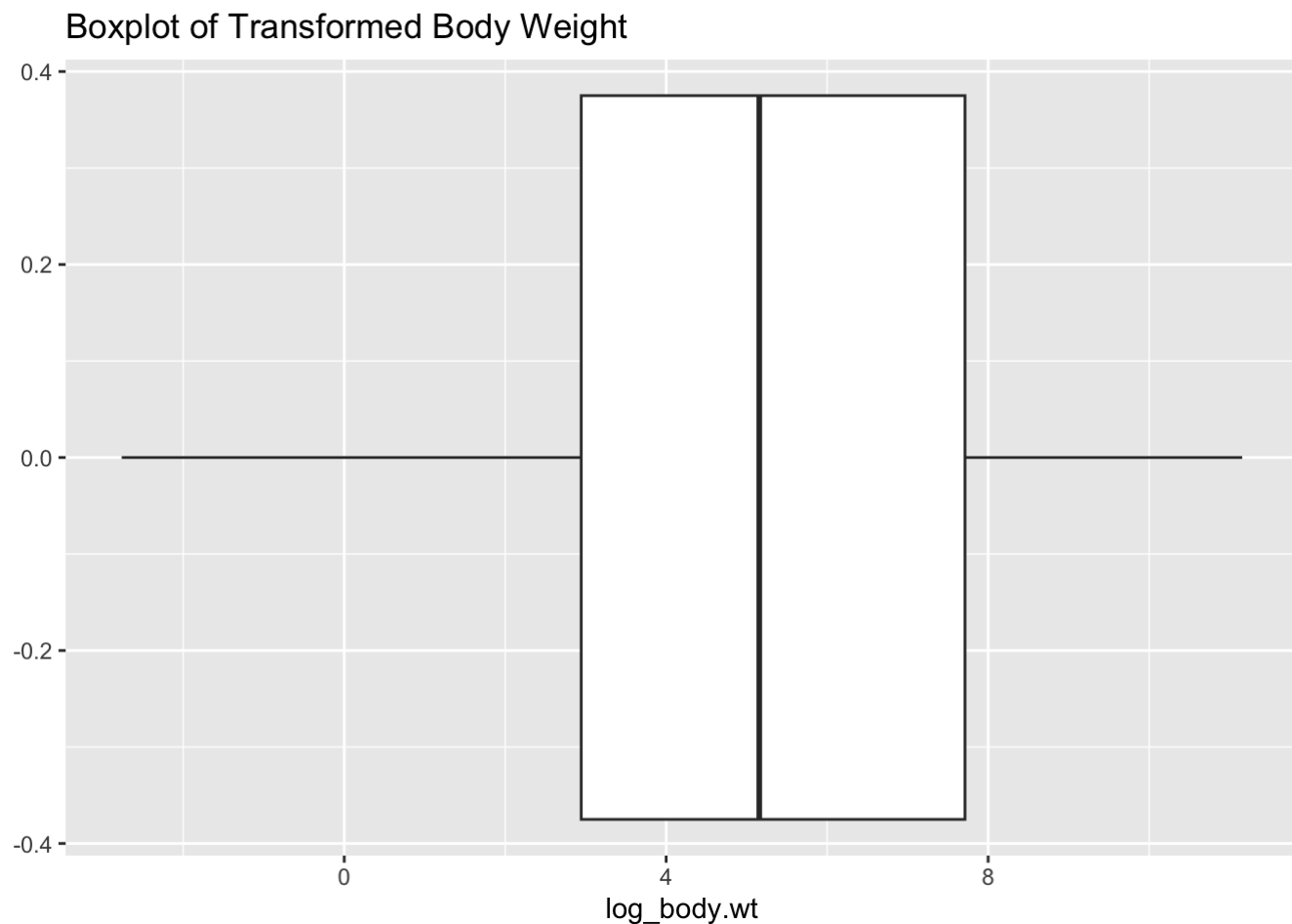
```
brain$log_body.wt = log(brain$Body.Weight)  
brain$log_brain.wt = log(brain$Brain.Weight)
```

# Distribution of transformed body weight

```
library("ggplot2")  
ggplot(brain, aes(x = log_body.wt)) + geom_histogram(bins = 16, color = "black", fill =  
"violet") + ggtitle("Histogram of Transformed Body Weight")
```



```
ggplot(brain, aes(x = log_body.wt)) + geom_boxplot() + ggtitle("Boxplot of Transformed B  
ody Weight")
```



```
xbar <- mean(brain$log_body.wt)
xbar
```

```
## [1] 4.946397
```

```
sd <- sd(brain$log_body.wt)
sd
```

```
## [1] 4.044355
```

```
n = 21
standard_error_mean <- sd/sqrt(n)
standard_error_mean
```

```
## [1] 0.8825506
```

```
margin <- qt(0.975,df=n-1)*sd/sqrt(n)
lowerinterval <- xbar - margin
lowerinterval
```

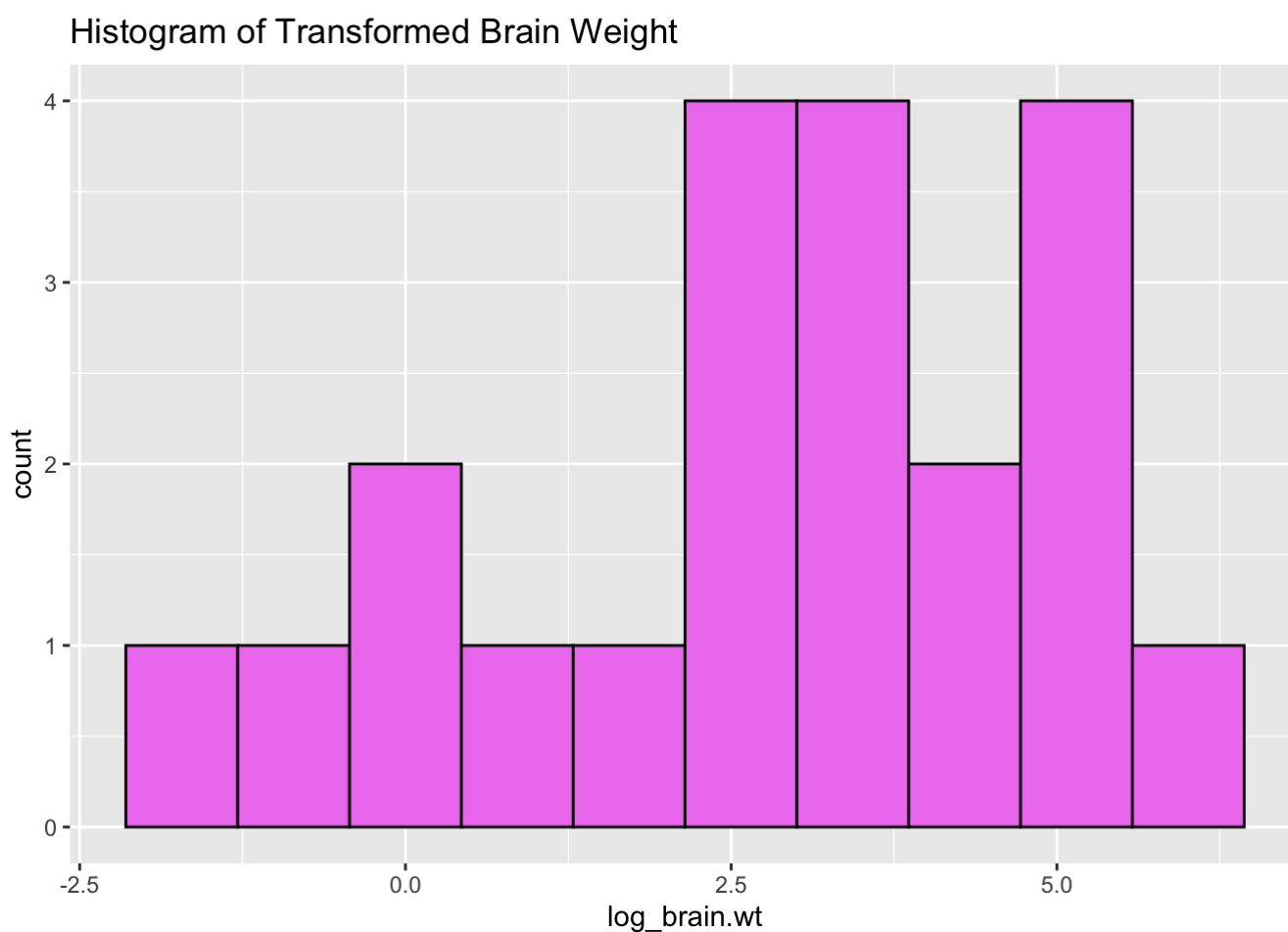
```
## [1] 3.105429
```

```
upperinterval <- xbar + margin  
upperinterval
```

```
## [1] 6.787366
```

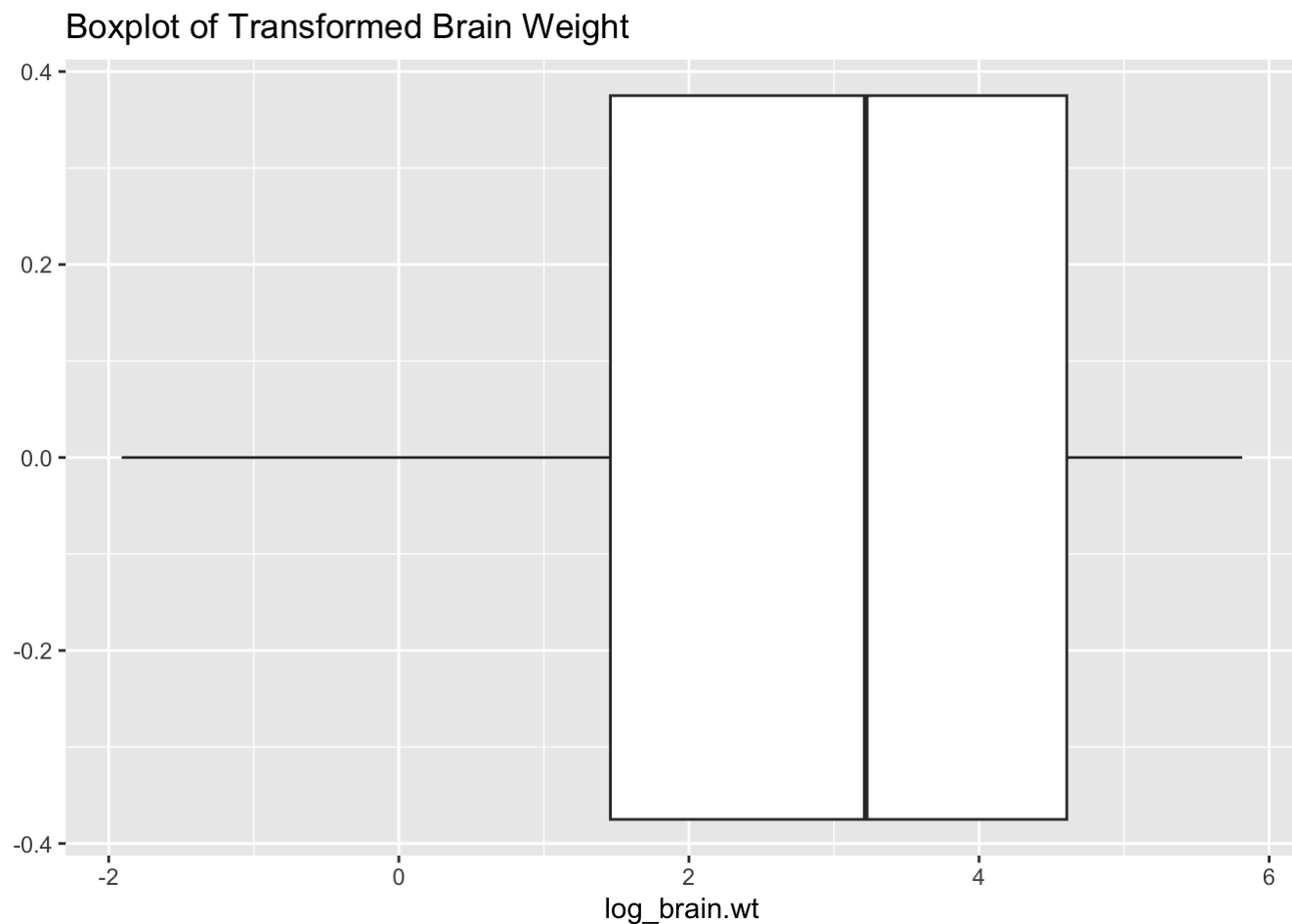
## Distribution of transformed brain weight

```
library("ggplot2")  
ggplot(brain, aes(x = log_brain.wt)) + geom_histogram(bins=10, color = "black", fill =  
"violet") + ggtitle("Histogram of Transformed Brain Weight")
```



```
ggplot(brain, aes(x = log_brain.wt)) + geom_boxplot() + ggtitle("Boxplot of Transformed  
Brain Weight")
```





```
xbar <- mean(brain$log_brain.wt)
xbar
```

```
## [1] 2.768455
```

```
sd <- sd(brain$log_brain.wt)
sd
```

```
## [1] 2.196627
```

```
n = 21
standard_error_mean <- sd/sqrt(n)
standard_error_mean
```

```
## [1] 0.4793432
```

```
margin <- qt(0.975,df=n-1)*sd/sqrt(n)
lowerinterval <- xbar - margin
lowerinterval
```

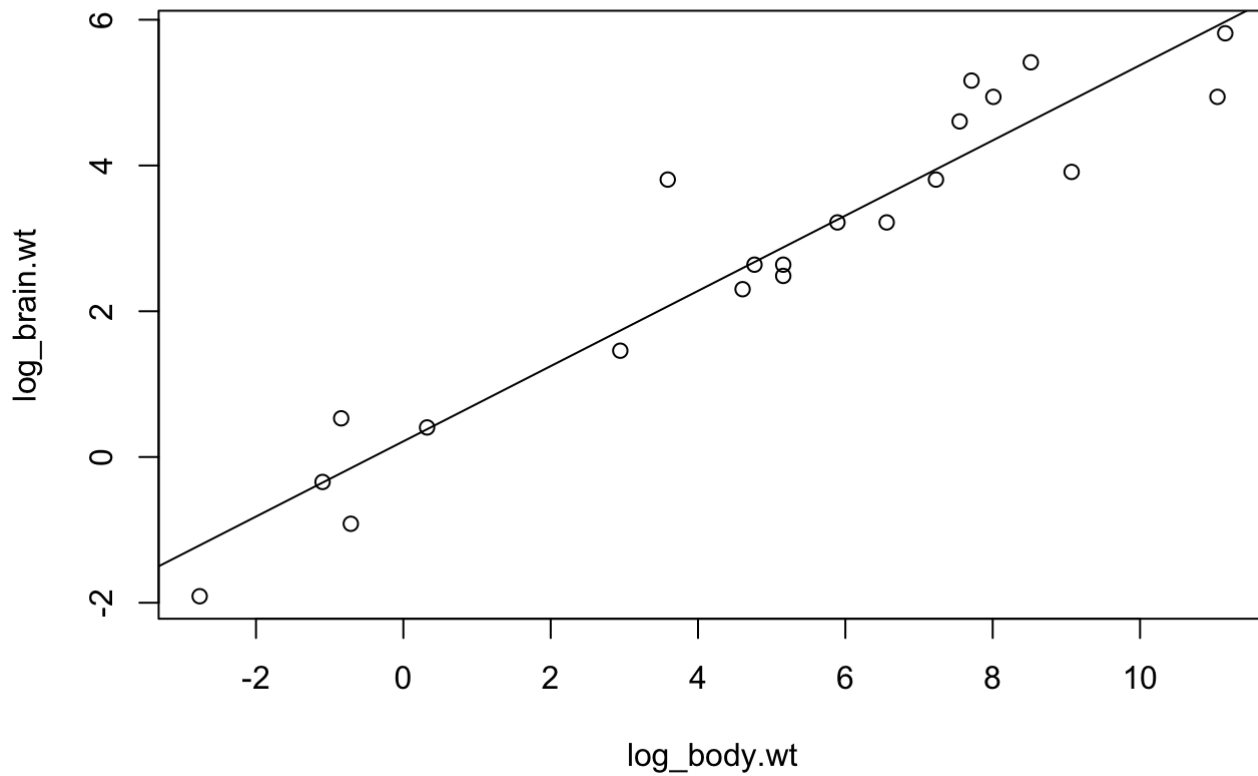
```
## [1] 1.768563
```

```
upperinterval <- xbar + margin  
upperinterval
```

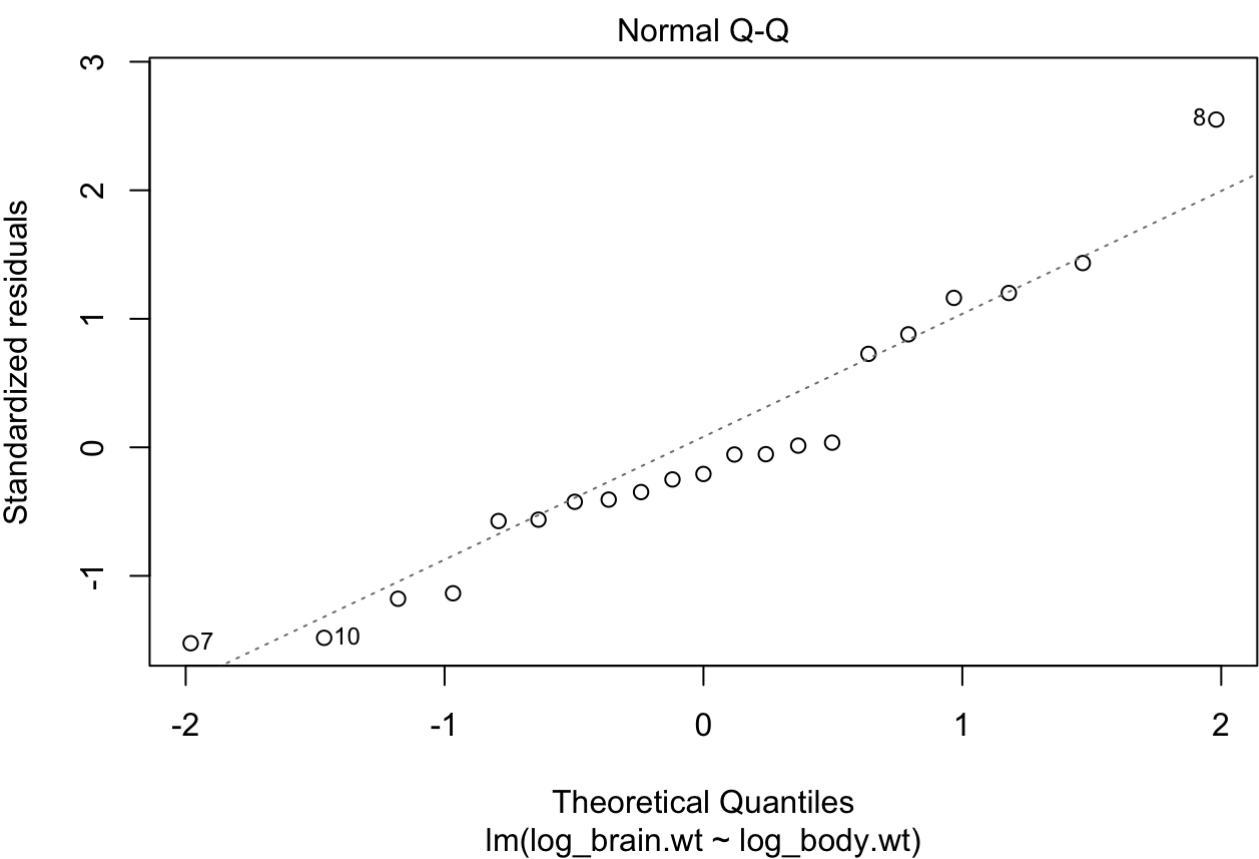
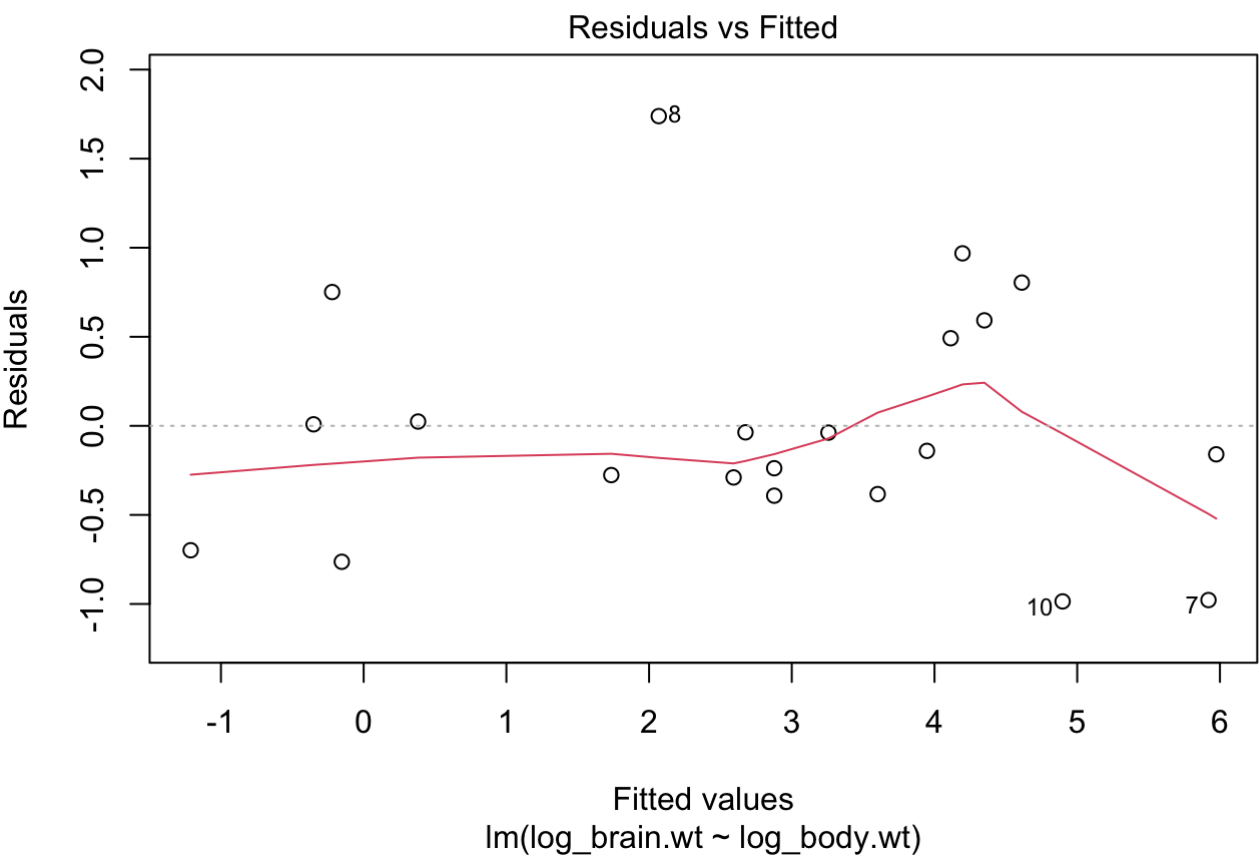
```
## [1] 3.768348
```

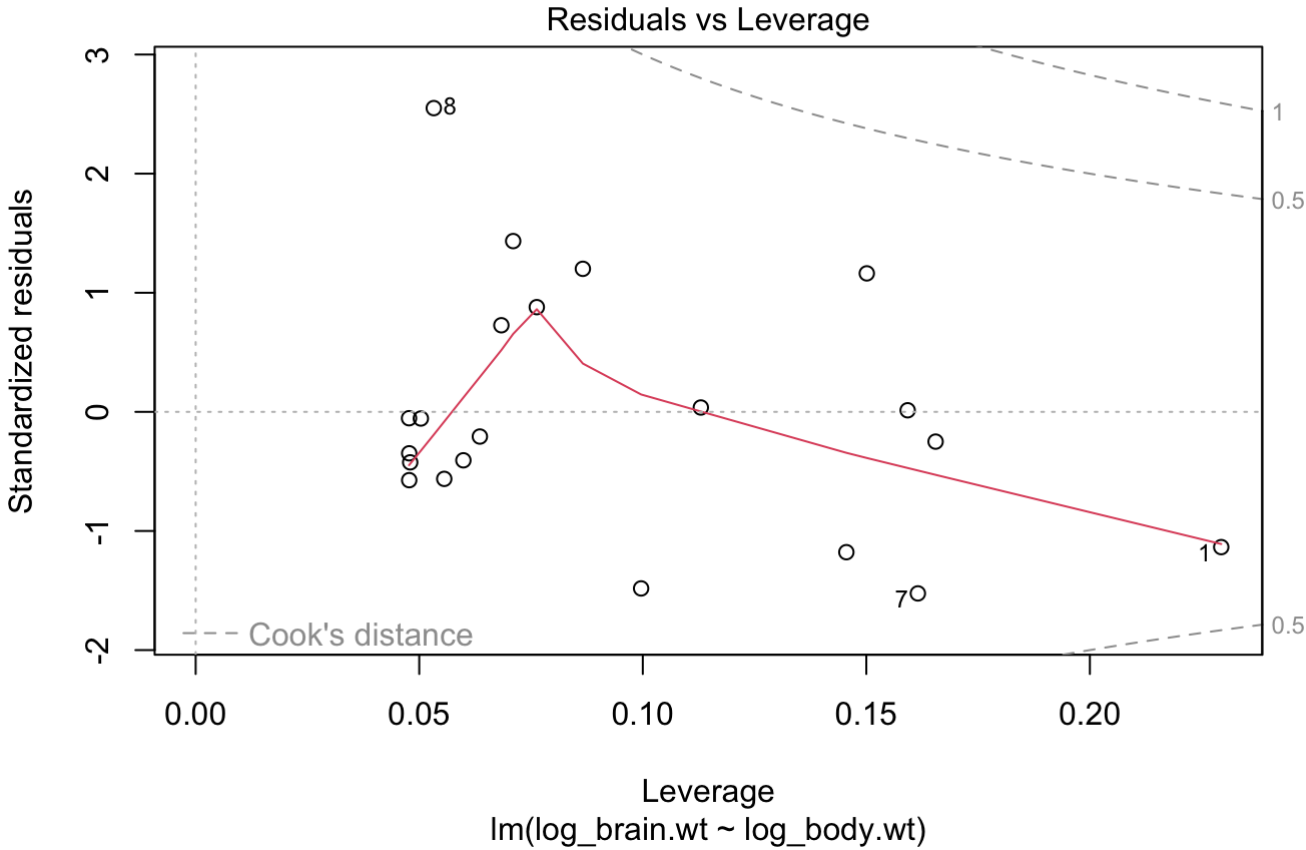
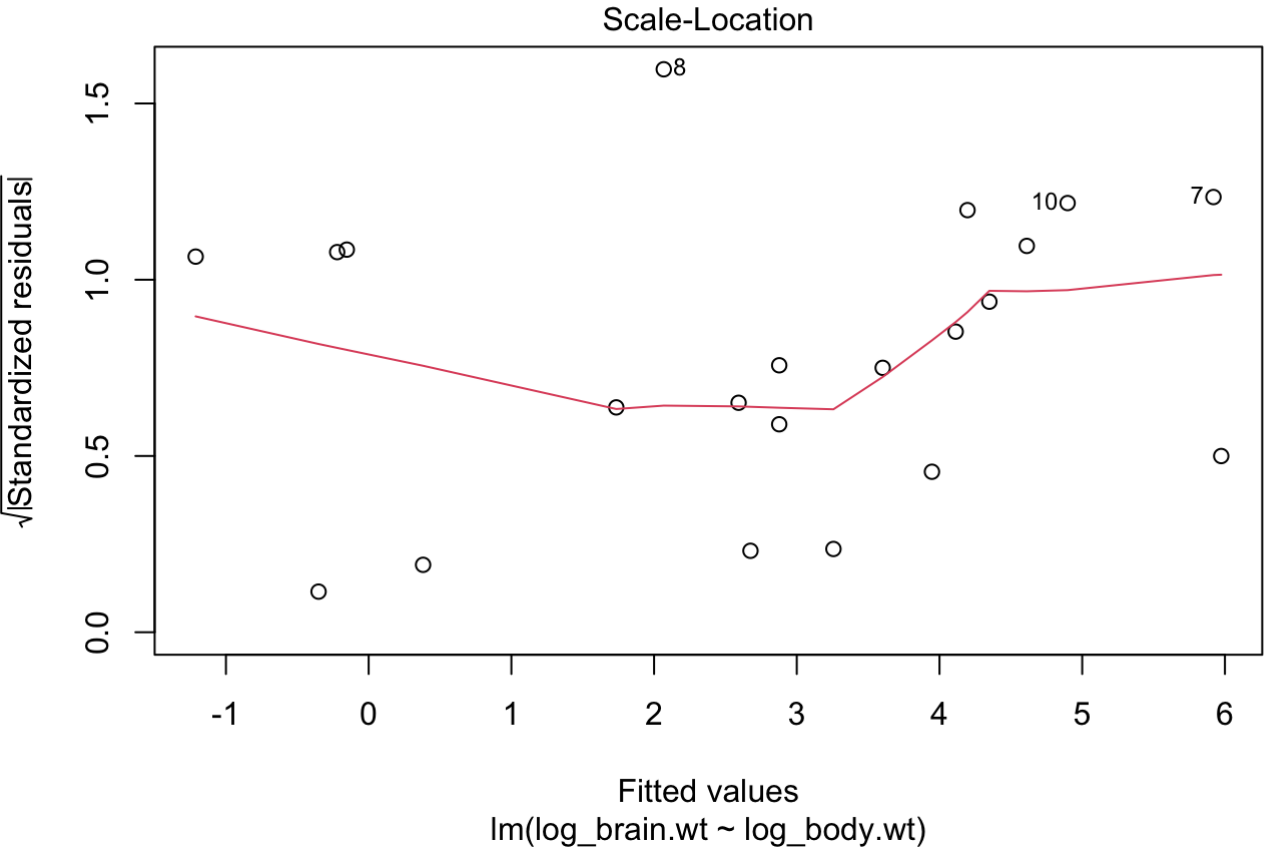
## Regression plot of transformed data

```
plot(log_brain.wt~log_body.wt, dat = brain)  
abline(lm(log_brain.wt~log_body.wt, dat = brain))
```



```
result<-lm(log_brain.wt~log_body.wt, dat = brain)  
plot(result)
```





# Lack of fit test

```
library("olsrr")
```

```
##
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:datasets':
##
##     rivers
```

```
fit <- lm(log_brain.wt~log_body.wt, dat = brain)
ols_pure_error_anova(fit)
```

```
##      Lack of Fit F Test
## -----
## Response :    log_brain.wt
## Predictor:    log_body.wt
##
##              Analysis of Variance Table
## -----
```

	DF	Sum Sq	Mean Sq	F Value	Pr(>F)
log_body.wt	1	87.17316	87.17316	7337.057	4.703166e-26
Residual	19	9.330217	0.4910641		
Lack of fit	18	9.318336	0.5176853	43.57175	0.1187295
Pure Error	1	0.01188122	0.01188122		

```
## -----
```

## Parameter estimates

```
result<-lm(log_brain.wt~log_body.wt, dat = brain)
summary(result)
```

```
##
## Call:
## lm(formula = log_brain.wt ~ log_body.wt, data = brain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9856 -0.3831 -0.1405  0.4919  1.7389
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.21507    0.24518   0.877   0.391
## log_body.wt  0.51621    0.03874  13.324 4.34e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7008 on 19 degrees of freedom
## Multiple R-squared:  0.9033, Adjusted R-squared:  0.8982
## F-statistic: 177.5 on 1 and 19 DF,  p-value: 4.341e-11
```

```
confint(result)
```

```
##              2.5 %      97.5 %
## (Intercept) -0.2980876 0.7282302
## log_body.wt  0.4351187 0.5973031
```

## Predict brain weight if body weight is 20

```
x <- data.frame(20, log(20))
names(x) <- c("Body.Weight", "log_body.wt")
pred <- predict(result, x)
pred
```

```
##      1
## 1.761501
```

```
exp(pred)
```

```
##      1
## 5.821168
```

## Residual Plot

```
result<-lm(log_brain.wt~log_body.wt, dat = brain)
res <- resid(result)
plot(fitted(result), res)
abline(0,0)
```

