Data_Analysis_Project2

Parker Hicks

4/21/2021

Data

```
tit_data <- read.csv(("/Users/parkerhicks/Desktop/Biostats_data/tit_fitness_data.csv"), stringsAsFactor
tit_dataOmit <- na.omit(subset(tit_data, select=c("Weight", "wing", "broodSizeWhenSampled")))
```

Hypothesis

Null: There is no difference in the brood size of wild great tits on the basis of weight. There is no difference in the brood size of wild great tits on the basis of wing size. There is no difference in the brood size of wild great tits on the basis of weight and wing size

Alternative: There is a difference in the brood size of wild great tits on the basis of weight. There is a difference in the brood size of wild great tits on the basis of wing size. There is a difference in the brood size of wild great tits on the basis of weight and wing size

Generalized Linear Model

```
tit_glm <- glm(broodSizeWhenSampled ~ Weight*wing, family="poisson", data = tit_dataOmit)
summary(tit_glm)
```

```
##
## Call:
## glm(formula = broodSizeWhenSampled ~ Weight * wing, family = "poisson",
##     data = tit_dataOmit)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.87156  -0.19867  -0.04929   0.35734   1.20731
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.248544   2.016328  -0.619    0.536
## Weight       0.146801   0.120275   1.221    0.222
## wing         0.061178   0.041186   1.485    0.137
## Weight:wing -0.003377   0.002421  -1.395    0.163
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 42.072  on 162  degrees of freedom
```

```
## Residual deviance: 38.277  on 159  degrees of freedom
## AIC: 580.95
##
## Number of Fisher Scoring iterations: 4
```

Confidence Intervals (glm)
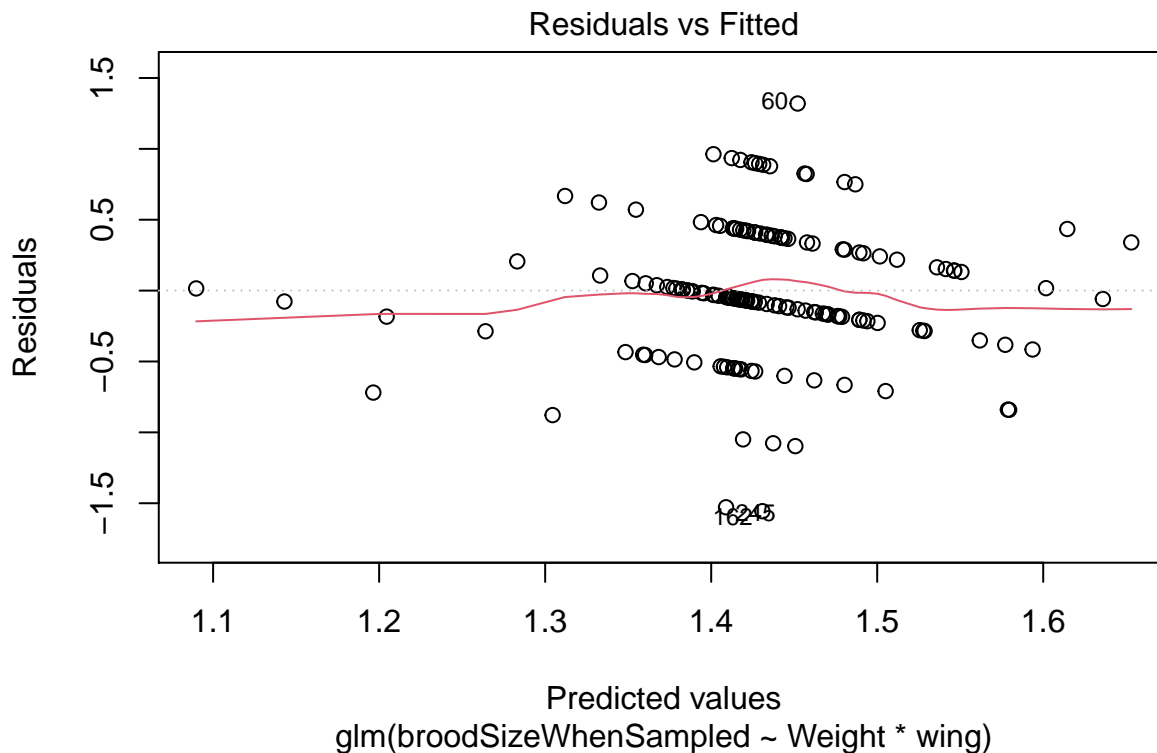
```
cov.tit_glm <- vcovHC(tit_glm, type="HC0")
std.err <- sqrt(diag(cov.tit_glm))
r.est <- cbind(Estimate= coef(tit_glm), "SE" = std.err,
"Pr(>|z|)" = 2 * pnorm(abs(coef(tit_glm)/std.err), lower.tail=FALSE),
Lower_CI = coef(tit_glm) - 1.96 * std.err,
Upper_CI = coef(tit_glm) + 1.96 * std.err)

r.est
```
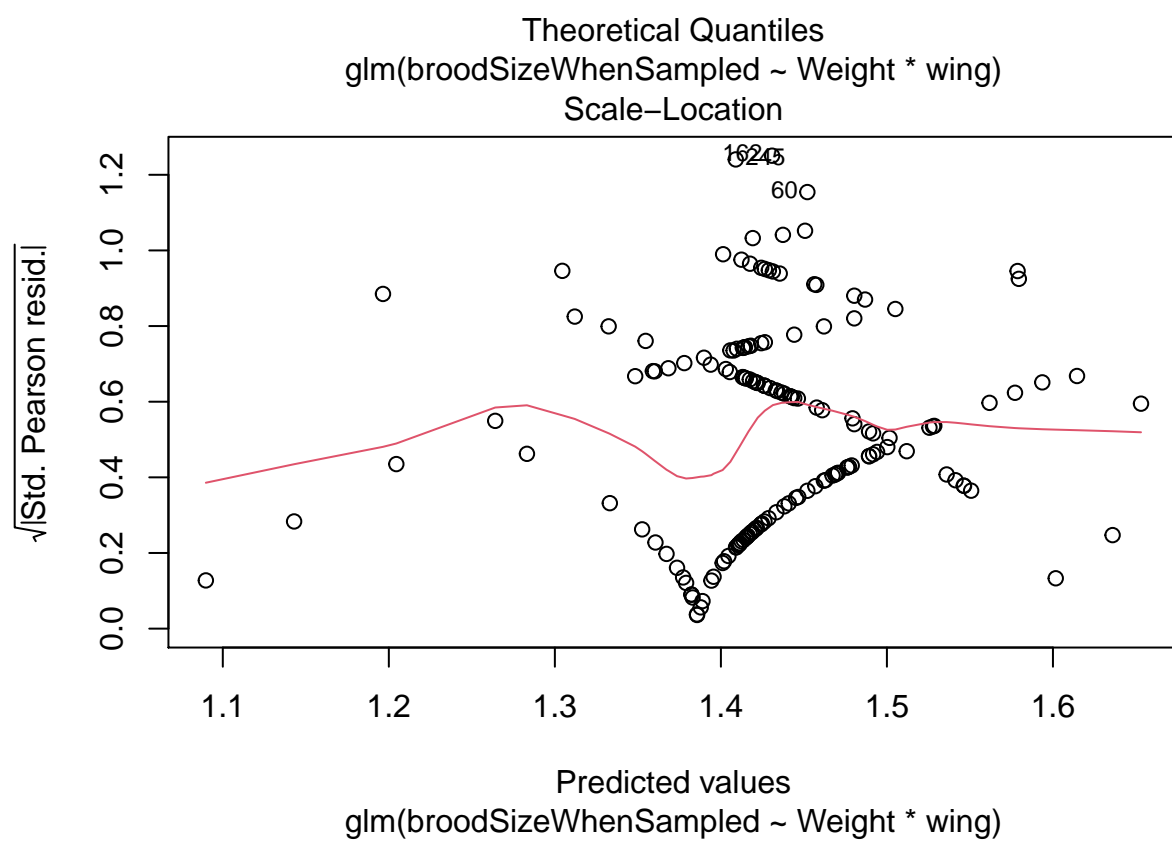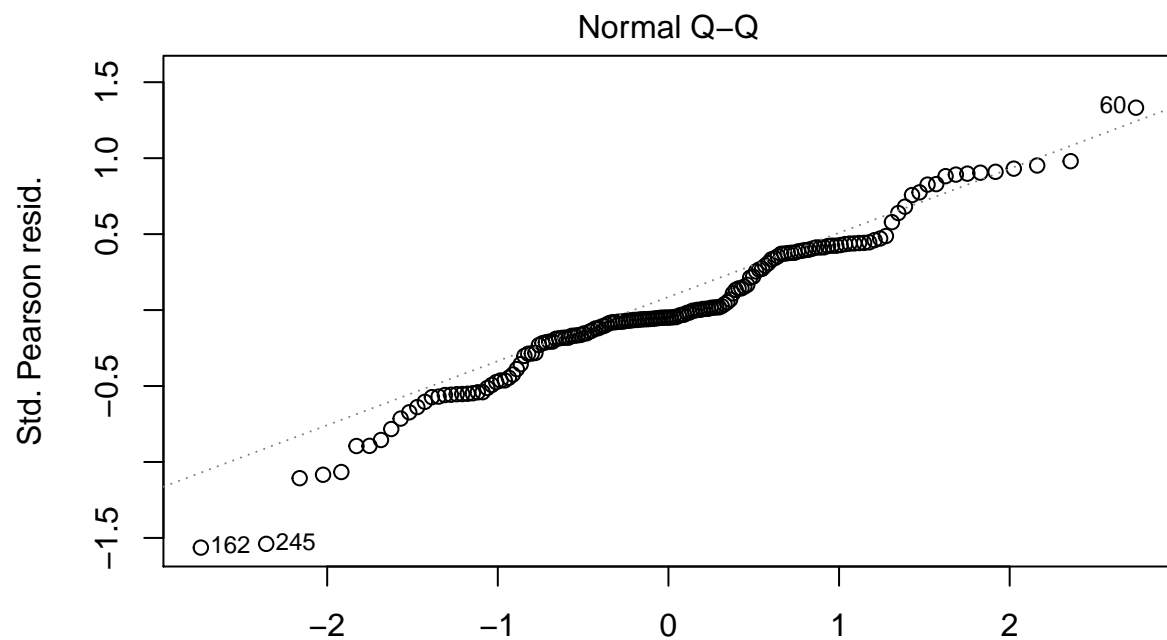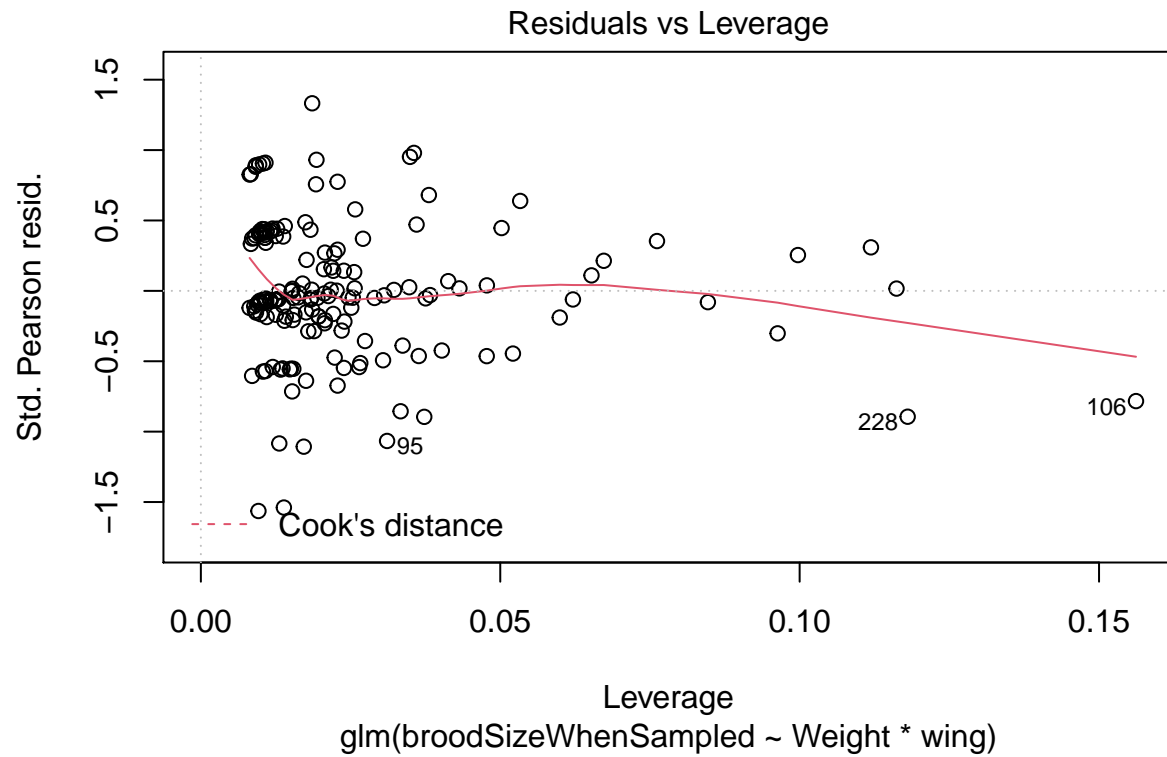
```
##                  Estimate          SE    Pr(>|z|)     Lower_CI      Upper_CI
## (Intercept)  -1.24854405  0.919561651  0.174539843  -3.050884887   0.553796785
## Weight        0.14680090  0.055743487  0.008450791   0.037543669   0.256058140
## wing          0.06117815  0.019014265  0.001293188   0.023910193   0.098446110
## Weight:wing  -0.00337726  0.001137251  0.002981202  -0.005606273  -0.001148248
```

##Assumption Tests (glm)

```
plot(tit_glm)
```


Residuals vs Fitted

Predicted values
glm(broodSizeWhenSampled ~ Weight * wing)

Normal Q–Q

Std. Pearson resid.

60

162 245

Theoretical Quantiles
glm(broodSizeWhenSampled ~ Weight * wing)

Scale–Location

$\sqrt{|\text{Std. Pearson resid.}|}$

162 245

60

Predicted values
glm(broodSizeWhenSampled ~ Weight * wing)

Residuals vs Leverage
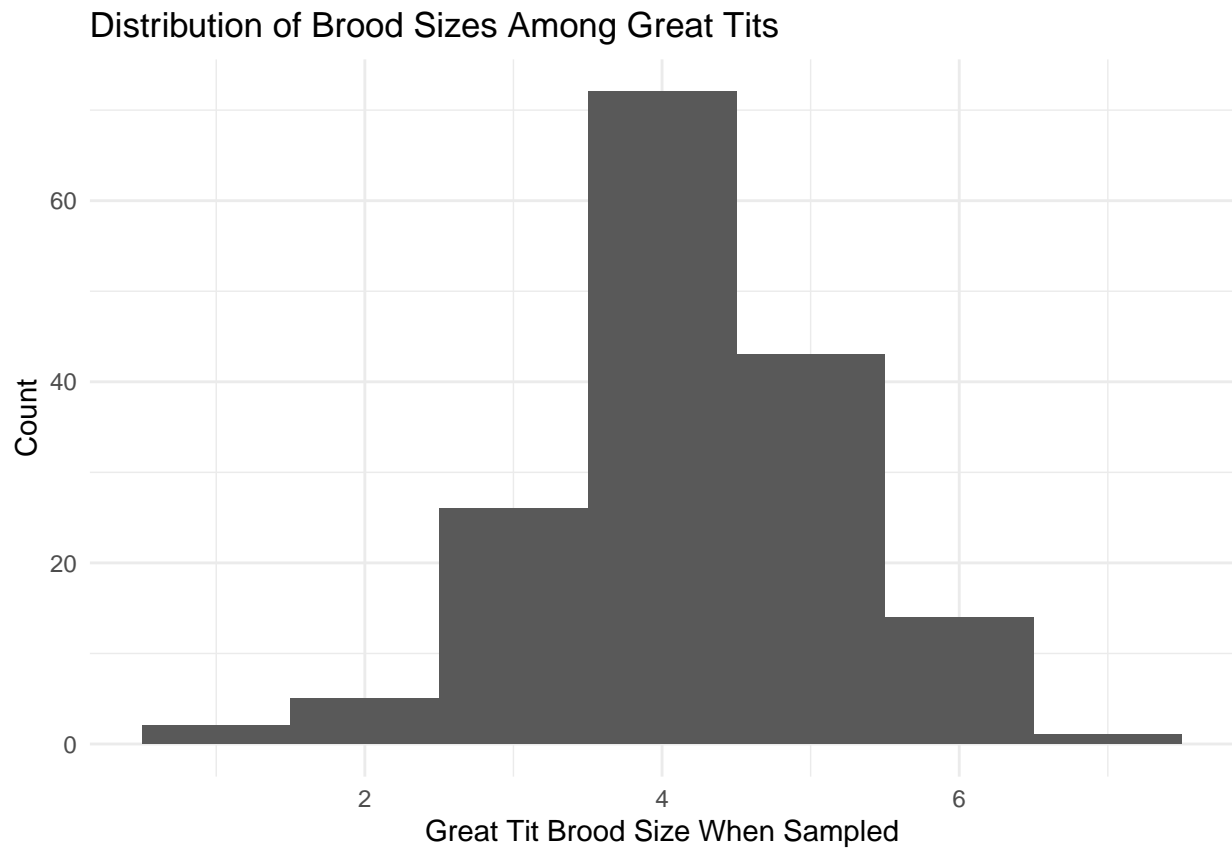
glm(broodSizeWhenSampled ~ Weight * wing)

Graphs

Brood Size Distribution

```
ggplot(tit_dataOmit, aes(x= broodSizeWhenSampled)) +
  geom_histogram(binwidth = 1) +
  theme_minimal() +
  xlab("Great Tit Brood Size When Sampled") +
  ylab("Count") +
  ggtitle("Distribution of Brood Sizes Among Great Tits ")
```

Distribution of Brood Sizes Among Great Tits

Wing and Weight

```
ggplot(tit_dataOmit, aes(x=wing, y=Weight)) +
  geom_point() +
  theme_minimal() +
  ggtitle("Wing and Weight Measurements of Great Tits") +
  xlab("Wing Size (mm)") +
  ylab("Weight (g)")
```

Wing and Weight Measurements of Great Tits
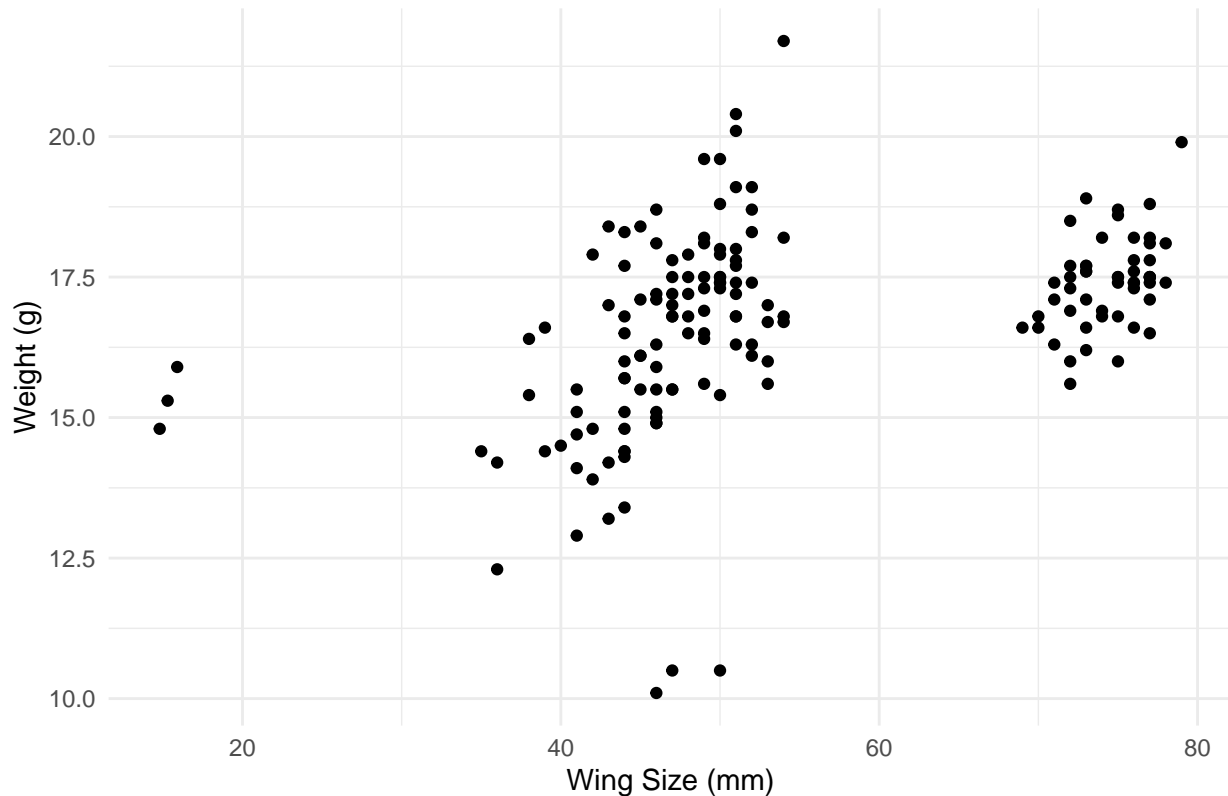
Summary Statistics

Distribution

```
summary(tit_dataOmit)
```

```
##      Weight           wing       broodSizeWhenSampled
##  Min.   :10.10   Min.   :14.8   Min.   :1.000
##  1st Qu.:15.90   1st Qu.:45.5   1st Qu.:4.000
##  Median :17.00   Median :50.0   Median :4.000
##  Mean   :16.72   Mean   :54.9   Mean   :4.196
##  3rd Qu.:17.70   3rd Qu.:72.0   3rd Qu.:5.000
##  Max.   :21.70   Max.   :79.0   Max.   :7.000
```

Data means

```
mean_wing <- mean(tit_dataOmit$wing)
mean_weight <- mean(tit_dataOmit$Weight)

mean_wing
```

```
## [1] 54.89571
```

```
mean_weight
```

```
## [1] 16.71963
```

Standard Error

```r
ser <- function(x){
  st_dev <- sd(x)
  n <- length(x)
  st_er <- st_dev / sqrt(n)

  return(st_er)
}

st_er_wing <- ser(tit_dataOmit$wing)
st_er_weight <- ser(tit_dataOmit$Weight)

st_er_weight
```

```
## [1] 0.1346556
```

```r
st_er_wing
```

```
## [1] 1.124782
```

Permutation Independence Test

```r
independence_test(broodSizeWhenSampled ~ Weight*wing,
                  data = tit_dataOmit)
```

```
##
##   Asymptotic General Independence Test
##
## data:  broodSizeWhenSampled by Weight, wing
## maxT = 2.2954, p-value = 0.04146
## alternative hypothesis: two.sided
```

Post hoc (Ordinal Logistic Regression)

```r
tit_dataOmit$broodSizeWhenSampled <- factor(tit_dataOmit$broodSizeWhenSampled, levels = c(1, 2, 3, 4, 5

log_model <- polr(broodSizeWhenSampled ~ Weight*wing , data = tit_dataOmit, Hess = TRUE)
summary(log_model)
```

```
## Call:
## polr(formula = broodSizeWhenSampled ~ Weight * wing, data = tit_dataOmit,
##     Hess = TRUE)
##
## Coefficients:
##               Value Std. Error t value
## Weight      1.33297    0.05652   23.58
## wing        0.56058    0.03792   14.78
## Weight:wing -0.03139    0.00219  -14.33
##
## Intercepts:
##     Value    Std. Error t value
## 1|2   19.2494    0.0097  1981.2869
```

```
## 2|3   20.5648    0.6131    33.5429
## 3|4   22.4573    0.7227    31.0741
## 4|5   24.6100    0.7625    32.2734
## 5|6   26.4539    0.7901    33.4815
##
## Residual Deviance: 427.8131
## AIC: 443.8131
## (1 observation deleted due to missingness)
```

Confusion Matrix

```
predictbrood = predict(log_model, tit_dataOmit)
table(tit_dataOmit$broodSizeWhenSampled, predictbrood)
```

```
##    predictbrood
##      1  2  3  4  5  6
##    1  0  0  0  2  0  0
##    2  0  0  1  4  0  0
##    3  0  0  3 21  2  0
##    4  0  0  0 69  3  0
##    5  0  0  0 35  8  0
##    6  0  0  0 12  2  0
```

```
mean(as.character(tit_dataOmit$broodSizeWhenSampled) != as.character(predictbrood))
```

```
## [1] NA
```

The ordinal logistic model identifies low and high brood sizes poorly. This is because of inadequate representation of low and high brood sizes in the data. Shown in the histogram above.

Post hoc Graphs

Probability of Brood Size (Wing as continuous variable)

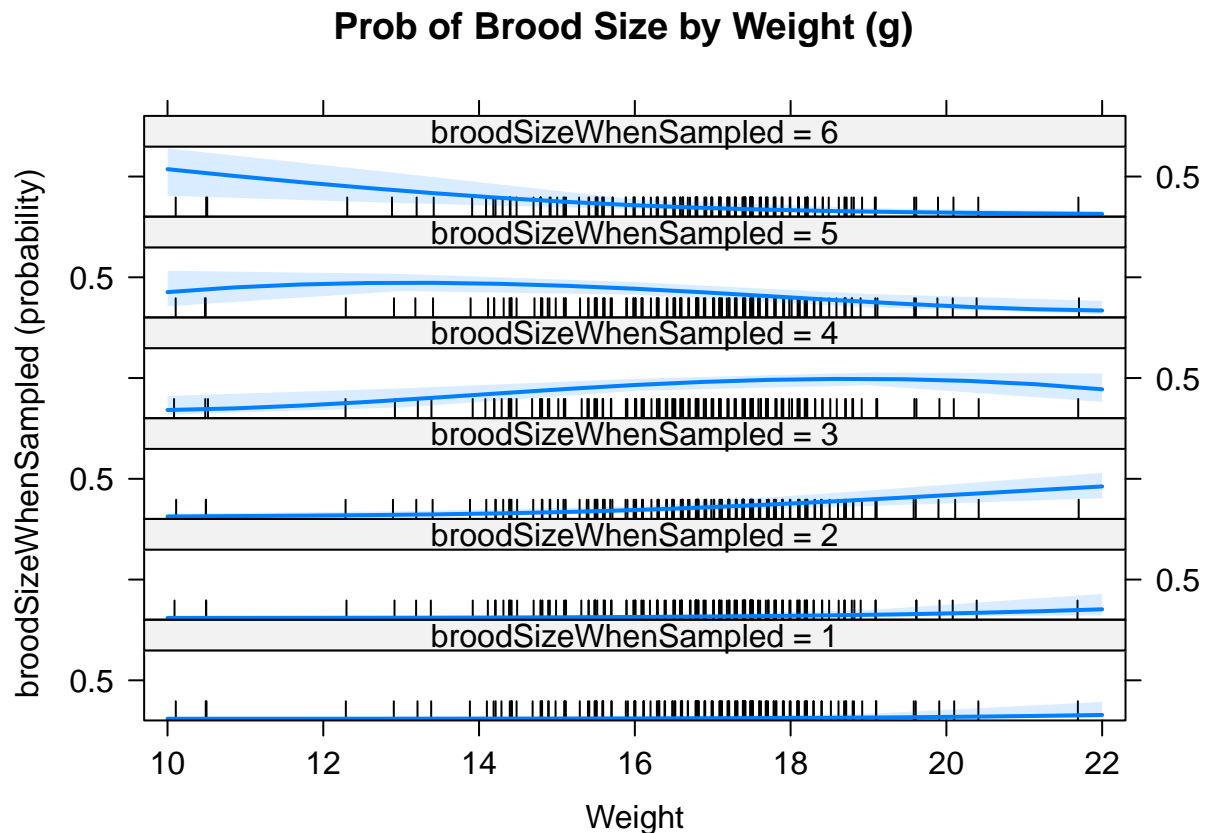```
library("effects")
Effect(focal.predictors = "Weight",log_model)
```

```
##
## Weight effect (probability) for 1
## Weight
##          10          13          16          19          22
## 0.000503763 0.001606005 0.005107644 0.016120781 0.049694385
##
## Weight effect (probability) for 2
## Weight
##          10          13          16          19          22
## 0.001370860 0.004352443 0.013663835 0.041421859 0.113389090
##
## Weight effect (probability) for 3
## Weight
##          10          13          16          19          22
## 0.01043547 0.03229759 0.09387945 0.23080099 0.40082998
##
## Weight effect (probability) for 4
## Weight
```

```
##              10           13           16           19           22
## 0.08458112 0.21681402 0.40952211 0.48882684 0.35365470
##
## Weight effect (probability) for 5
## Weight
##          10          13          16          19          22
## 0.3072151 0.4289088 0.3513661 0.1794377 0.0684184
##
## Weight effect (probability) for 6
## Weight
##              10           13           16           19           22
## 0.59589367 0.31602110 0.12646083 0.04339183 0.01401345
```

```
plot(Effect(focal.predictors = "Weight",log_model), ticks=list(at=c(0.5)), main = "Prob of Brood Size by
```

Prob of Brood Size by Weight (g)



```
Effect(focal.predictors = "wing",log_model)
```

```
##
## wing effect (probability) for 1
## wing
##           10          30          50          60          80
## 0.032664826 0.016203565 0.007969523 0.005579283 0.002729125
##
## wing effect (probability) for 2
## wing
##           10          30          50          60          80
## 0.079102281 0.041622069 0.021096186 0.014899581 0.007365443
##
```

```
## wing effect (probability) for 3
## wing
##          10         30         50         60         80
## 0.34328098 0.23158750 0.13666700 0.10135812 0.05328778
##
## wing effect (probability) for 4
## wing
##         10        30        50        60        80
## 0.4228206 0.4886576 0.4652692 0.4224334 0.3047134
##
## wing effect (probability) for 5
## wing
##         10        30        50        60        80
## 0.1005956 0.1787530 0.2843176 0.3387570 0.4183235
##
## wing effect (probability) for 6
## wing
##          10         30         50         60         80
## 0.02153574 0.04317624 0.08468053 0.11697259 0.21358074
```

```
plot(Effect(focal.predictors = "wing",log_model), ticks=list(at=c(0.5)), main = "Prob of Brood Size by l
```


Prob of Brood Size by Wing size (mm)
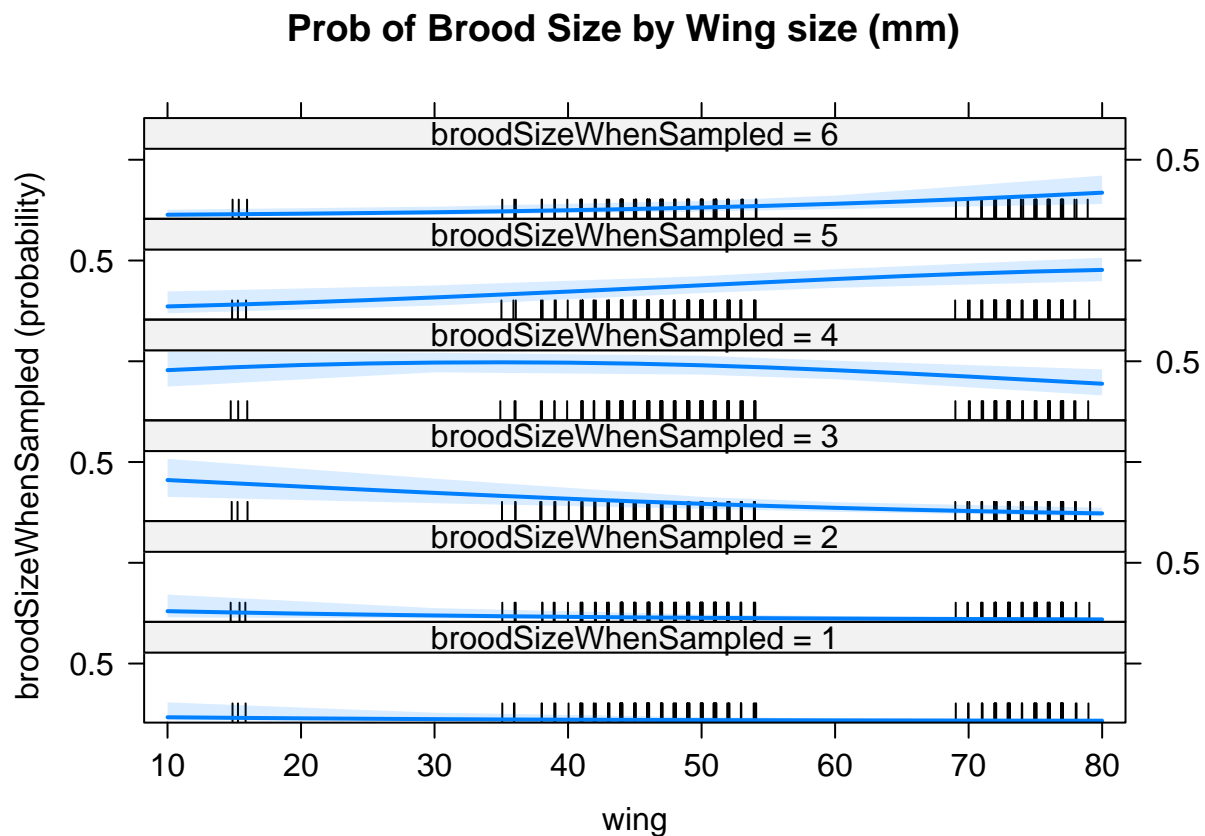
Probability of Brood Size (Wing as categorical variable)

```
tit_data_test <- read.csv(("/Users/parkerhicks/Desktop/Biostats_data/tit_fitness_data_cat.csv"), string

tit_data_test$broodSizeWhenSampled <- factor(tit_data_test$broodSizeWhenSampled, levels = c(1, 2, 3, 4,
```
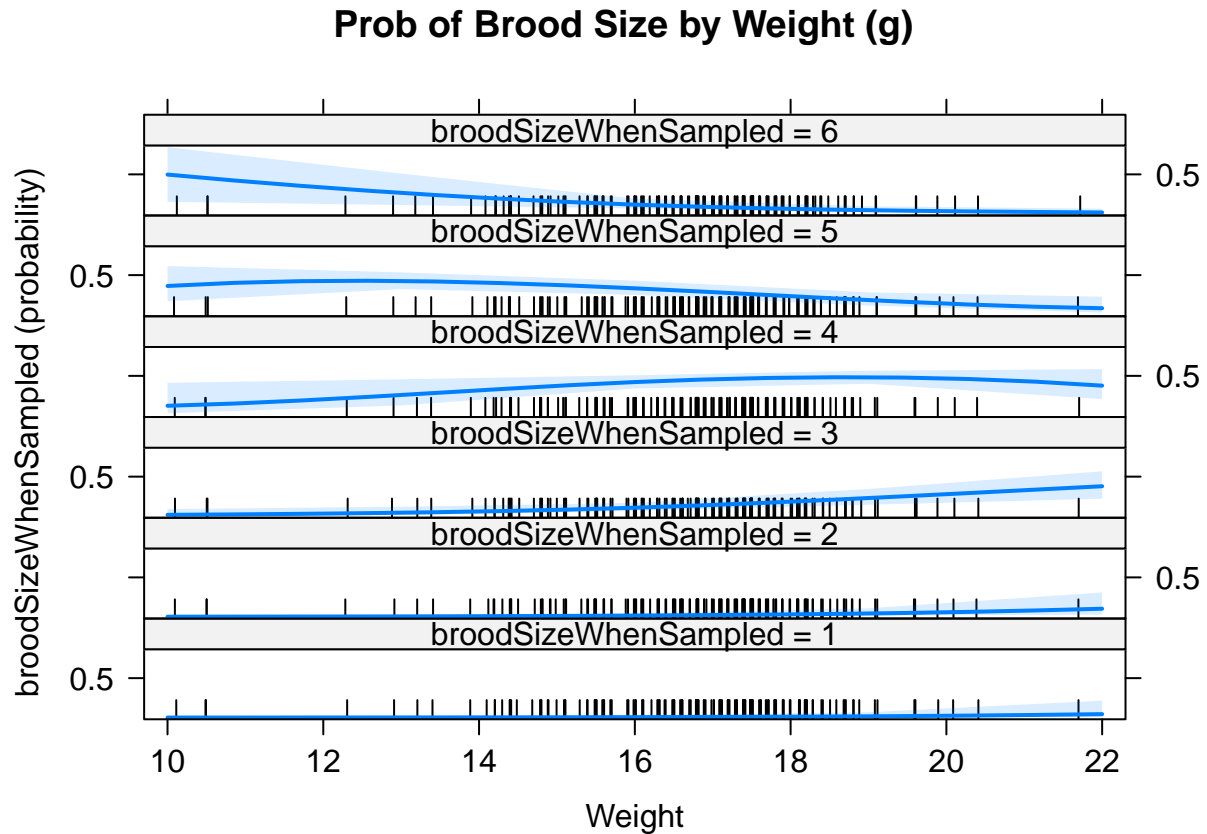
```
log_model_test <- polr(broodSizeWhenSampled ~ Weight*wing , data = tit_data_test, Hess = TRUE)
summary(log_model_test)
```

```
## Call:
## polr(formula = broodSizeWhenSampled ~ Weight * wing, data = tit_data_test,
##      Hess = TRUE)
##
## Coefficients:
##                     Value Std. Error t value
## Weight             -0.9523     0.3451  -2.759
## wingMedium        -14.6606     6.1919  -2.368
## wingSmall         -22.2632     9.0230  -2.467
## Weight:wingMedium   0.8307     0.3558   2.335
## Weight:wingSmall    1.2279     0.5614   2.187
##
## Intercepts:
##      Value    Std. Error t value
## 1|2 -21.3810   6.1438     -3.4801
## 2|3 -20.0645   6.0937     -3.2926
## 3|4 -18.2054   6.0492     -3.0095
## 4|5 -16.0910   6.0156     -2.6749
## 5|6 -14.2683   5.9904     -2.3818
##
## Residual Deviance: 432.0816
## AIC: 452.0816
## (1 observation deleted due to missingness)
```

```
Effect(focal.predictors = "Weight",log_model_test)
```

```
##
## Weight effect (probability) for 1
## Weight
##            10           13           16           19           22
## 0.0008237131 0.0022515935 0.0061394505 0.0166286390 0.0442407771
##
## Weight effect (probability) for 2
## Weight
##           10           13           16           19           22
## 0.002242143 0.006096354 0.016385364 0.042708240 0.103006940
##
## Weight effect (probability) for 3
## Weight
##          10          13          16          19          22
## 0.01628794 0.04290730 0.10630775 0.22882680 0.37840024
##
## Weight effect (probability) for 4
## Weight
##         10          13          16          19          22
## 0.1211801 0.2579450 0.4217786 0.4821603 0.3761305
##
## Weight effect (probability) for 5
## Weight
##          10          13          16          19          22
## 0.36242162 0.42554318 0.33287083 0.18371127 0.08092549
```

```
##
## Weight effect (probability) for 6
## Weight
##          10          13          16          19          22
## 0.49704444 0.26525660 0.11651802 0.04596473 0.01729609
```

```
plot(Effect(focal.predictors = "Weight",log_model_test), ticks=list(at=c(0.5)), main = "Prob of Brood S
```

Prob of Brood Size by Weight (g)
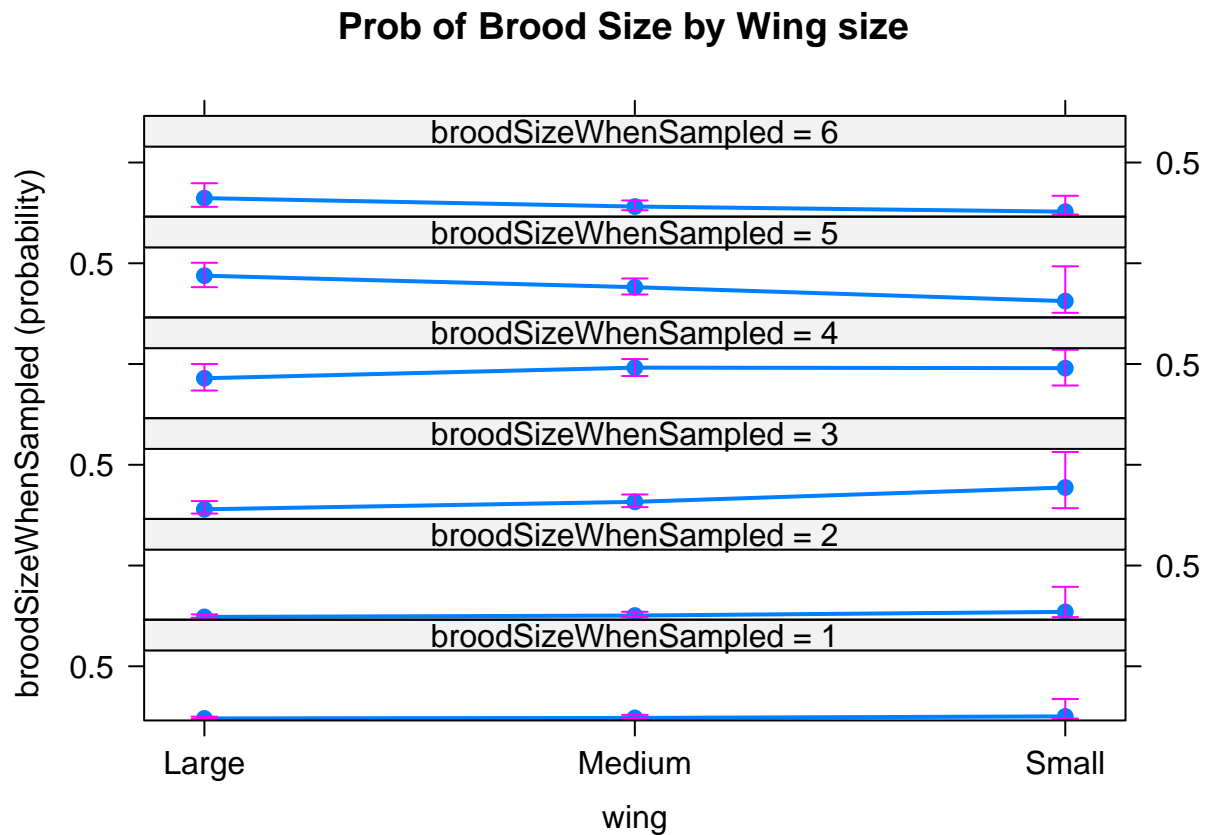

Weight

```
Effect(focal.predictors = "wing",log_model)
```

```
##
## wing effect (probability) for 1
## wing
##          10          30          50          60          80
## 0.032664826 0.016203565 0.007969523 0.005579283 0.002729125
##
## wing effect (probability) for 2
## wing
##          10          30          50          60          80
## 0.079102281 0.041622069 0.021096186 0.014899581 0.007365443
##
## wing effect (probability) for 3
## wing
##          10          30          50          60          80
## 0.34328098 0.23158750 0.13666700 0.10135812 0.05328778
##
## wing effect (probability) for 4
## wing
```

```
##          10          30          50          60          80
## 0.4228206 0.4886576 0.4652692 0.4224334 0.3047134
##
## wing effect (probability) for 5
## wing
##          10          30          50          60          80
## 0.1005956 0.1787530 0.2843176 0.3387570 0.4183235
##
## wing effect (probability) for 6
## wing
##            10          30          50          60          80
## 0.02153574 0.04317624 0.08468053 0.11697259 0.21358074
```

```
plot(Effect(focal.predictors = "wing",log_model_test), ticks=list(at=c(0.5)), main = "Prob of Brood Size
```
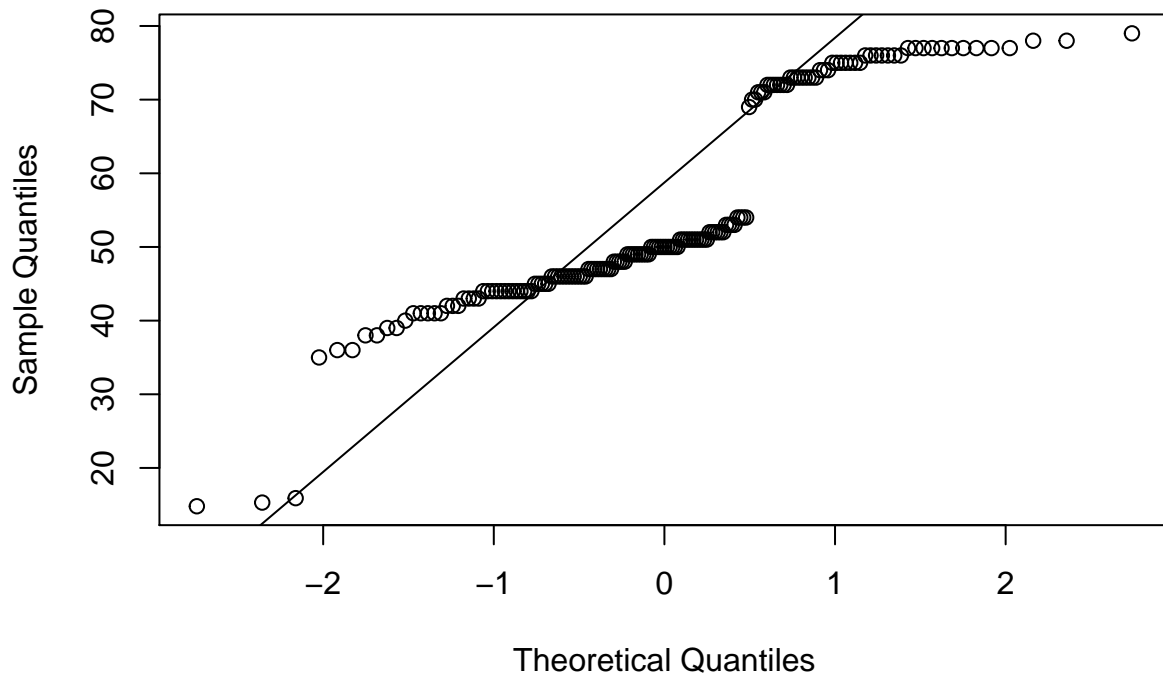
Prob of Brood Size by Wing size

Small: 10 - 40 Medium: 41 - 70 Large: 71 - 85
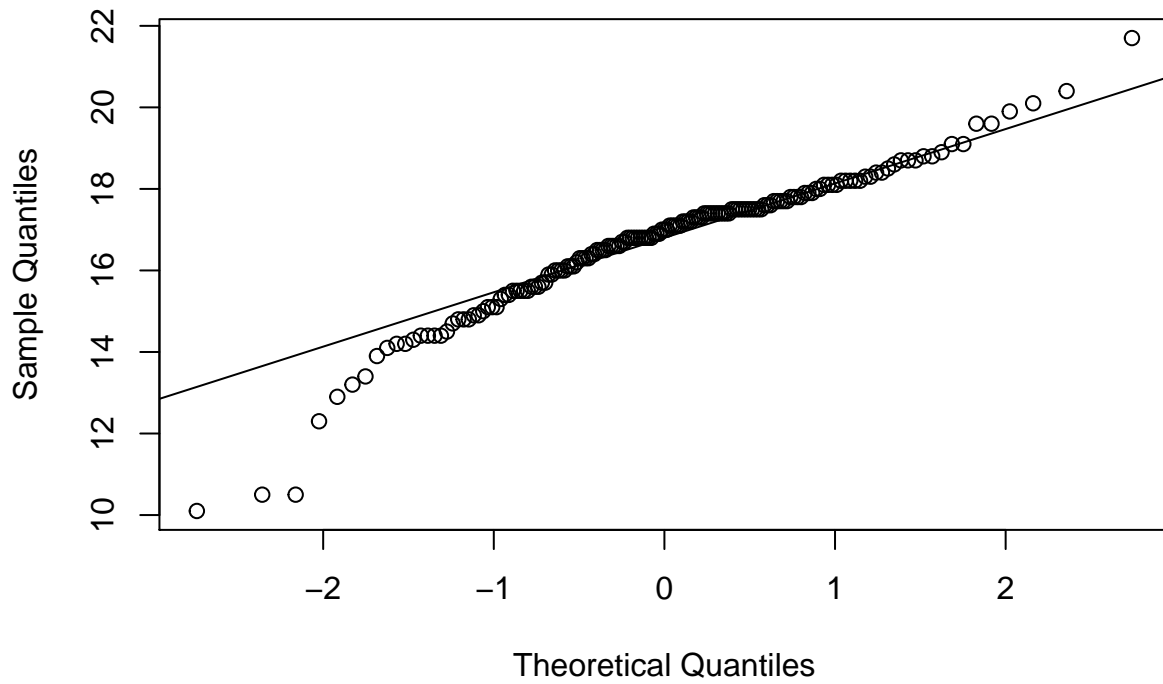
Assumption Tests

```
qqnorm(tit_dataOmit$wing)
qqline(tit_dataOmit$wing)
```

Normal Q–Q Plot


```
qqnorm(tit_dataOmit$Weight)
qqline(tit_dataOmit$Weight)
```

Normal Q–Q Plot

```
shapiro.test(tit_dataOmit$wing)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  tit_dataOmit$wing
## W = 0.86366, p-value = 5.376e-11
```

```
shapiro.test(tit_dataOmit$Weight)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  tit_dataOmit$Weight
## W = 0.94038, p-value = 2.41e-06
```

The numerical data do not fit a normal distribution as the great tit wing and weights failed the Shapiro-Wilk test and are visibly different on the Q-Q plots.