

Data_Analysis_Project2

Parker Hicks

4/21/2021

Data

```
tit_data <- read.csv("/Users/parkerhicks/Desktop/Biostats_data/tit_fitness_data.csv"), stringsAsFactors = FALSE
tit_dataOmit <- na.omit(subset(tit_data, select=c("Weight", "wing", "broodSizeWhenSampled")))
#tit_dataOmit$broodSizeWhenSampled <- tit_dataOmit$broodSizeWhenSampled / 10
```

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

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

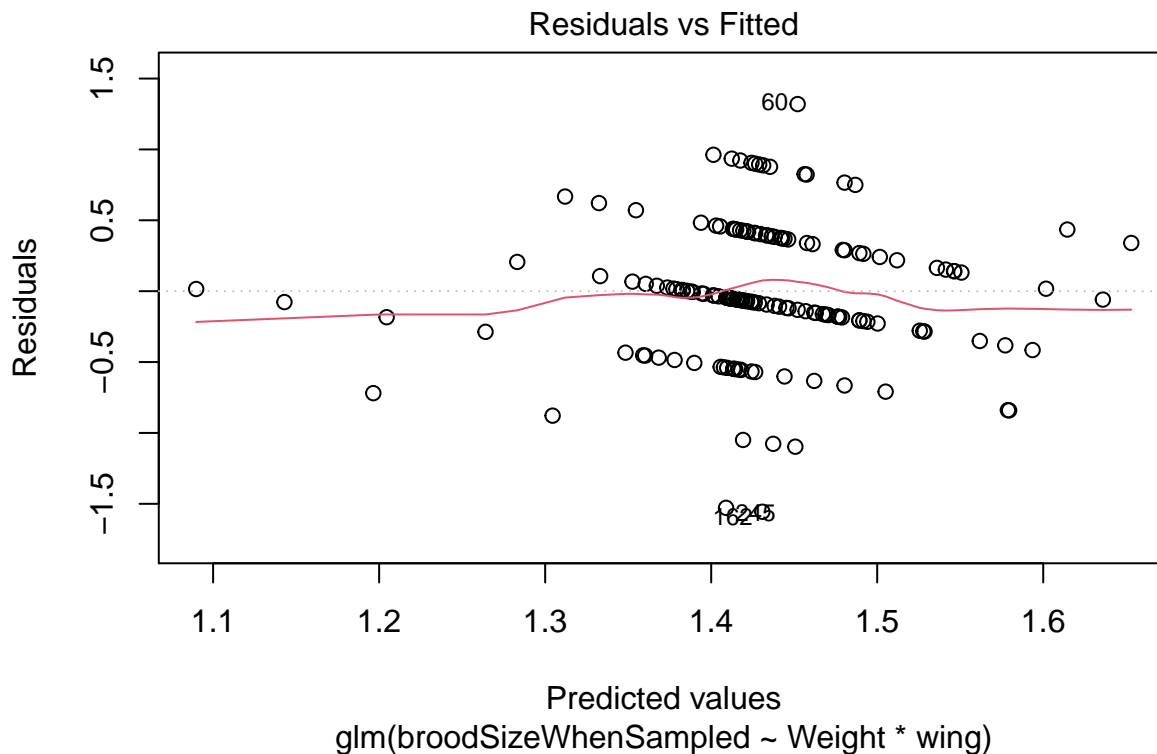
```
cov.tit_glm <- vcovHC(tit_glm, type="HC0")
std.err <- sqrt(diag(cov.tit_glm))
r.est <- cbind(Estimate= coef(tit_glm), "SE" = std.err,
"Pr(>|z|)" = 2 * pnorm(abs(coef(tit_glm)/std.err), lower.tail=FALSE),
Lower_CI = coef(tit_glm) - 1.96 * std.err,
Upper_CI = coef(tit_glm) + 1.96 * std.err)

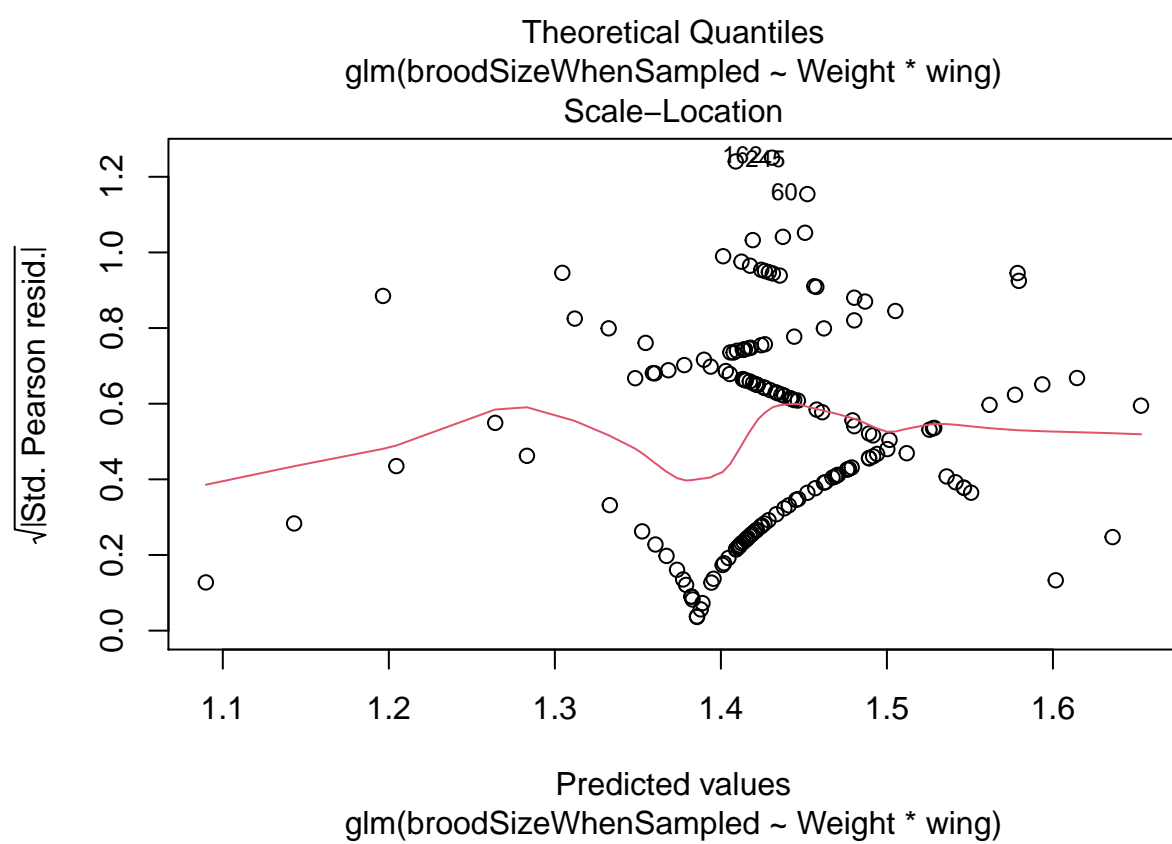
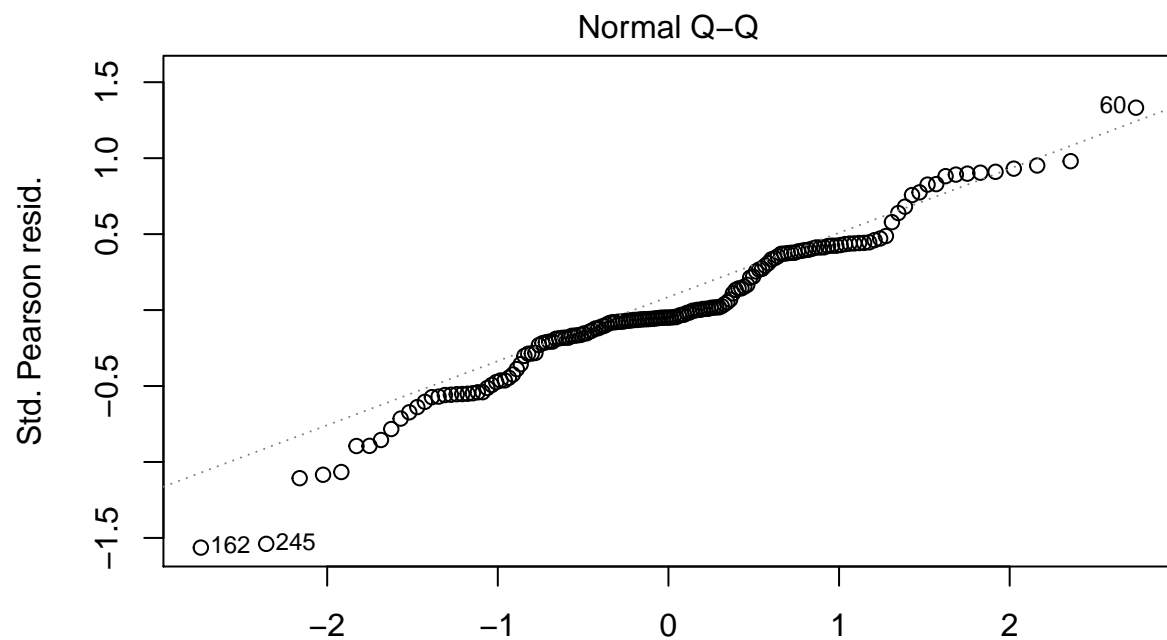
r.est
```

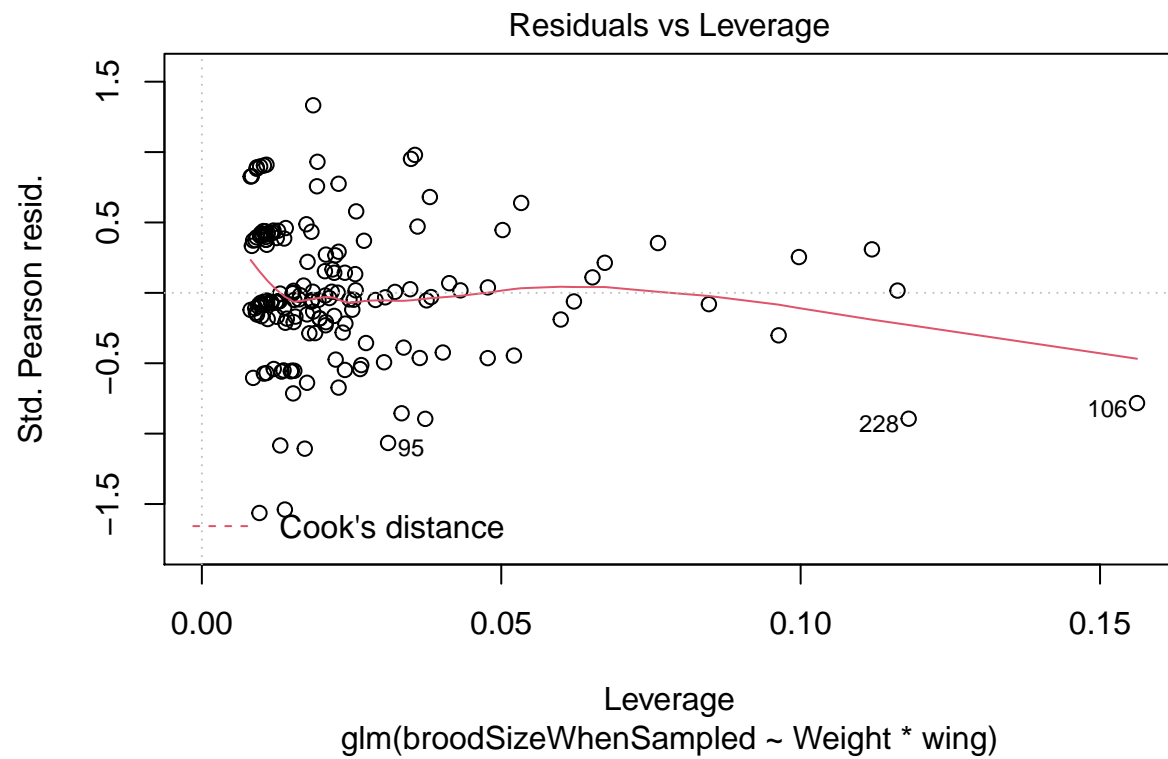
```
##           Estimate          SE   Pr(>|z|)   Lower_CI   Upper_CI
## (Intercept) -1.24854405 0.919561651 0.174539843 -3.050884887  0.553796785
## Weight      0.14680090 0.055743487 0.008450791  0.037543669  0.256058140
## wing        0.06117815 0.019014265 0.001293188  0.023910193  0.098446110
##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

#Assumption Tests

```
plot(tit_glm)
```



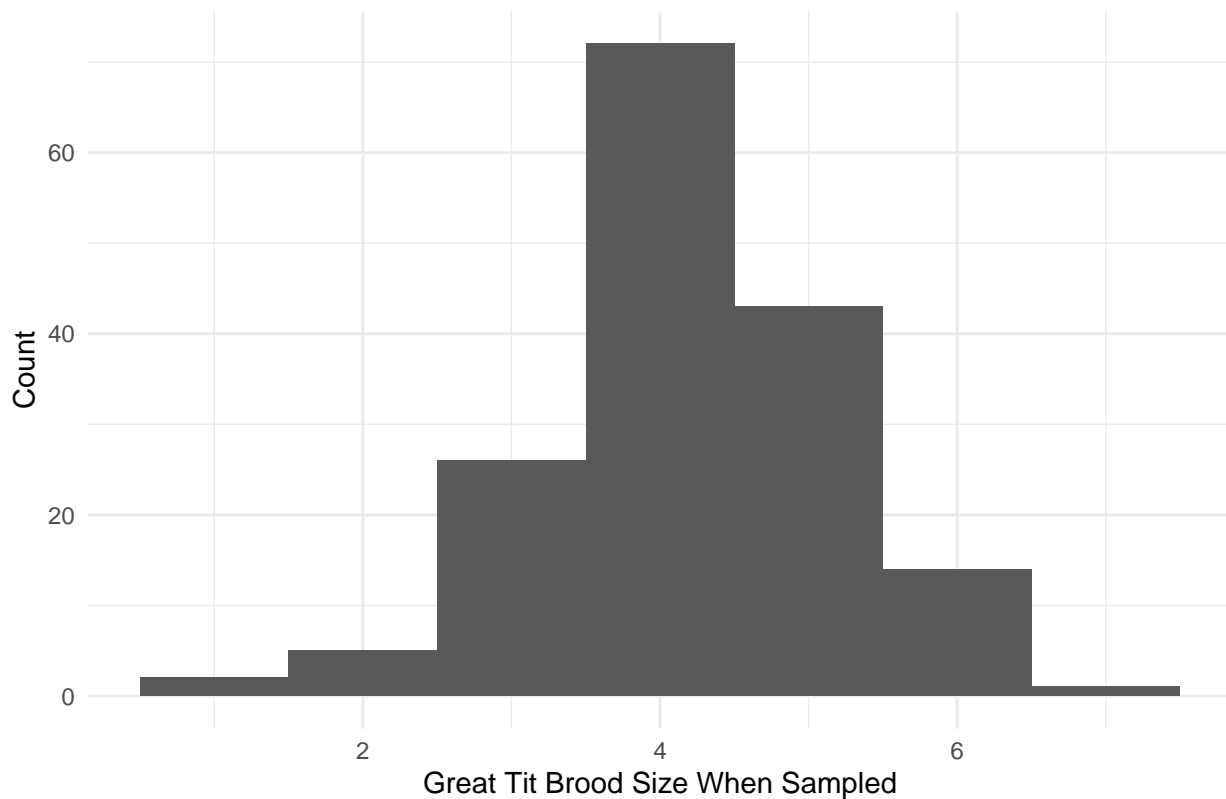




Graph

```
ggplot(tit_data0mit, 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



Permutation Independence Test

```
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

```
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.

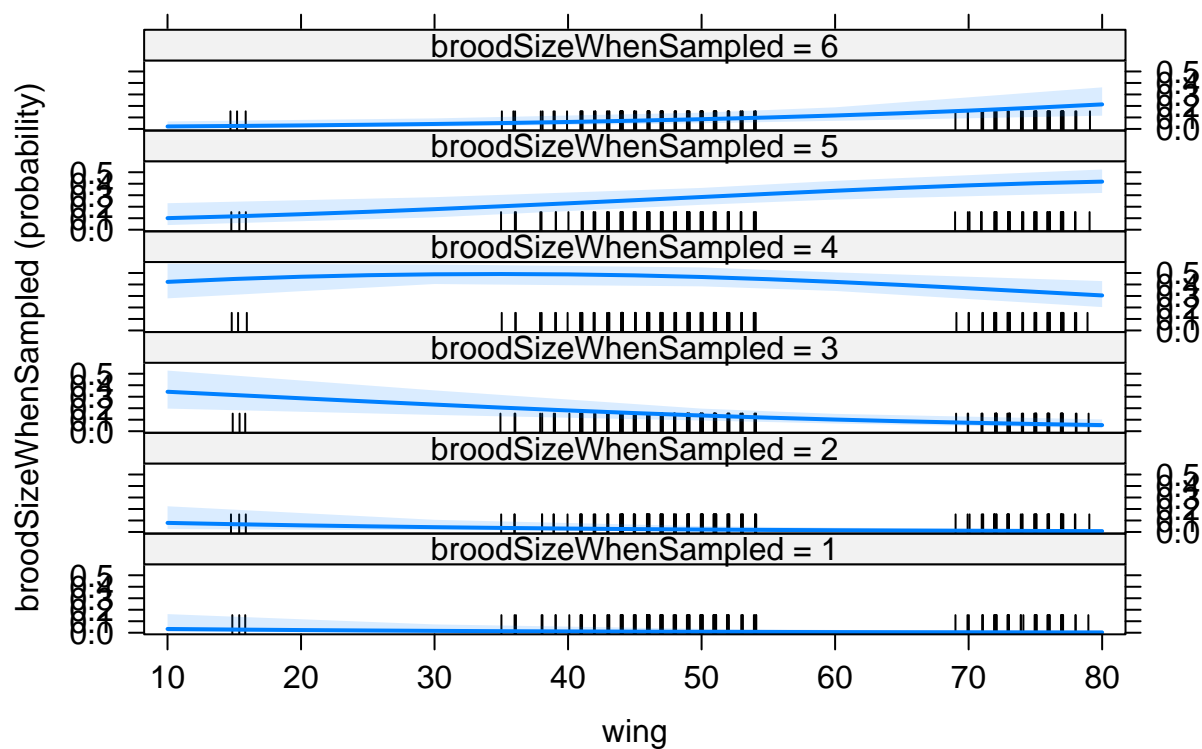
```
library("effects")
Effect(focal.predictors = "Weight", log_model)
```

```
##
## 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
```

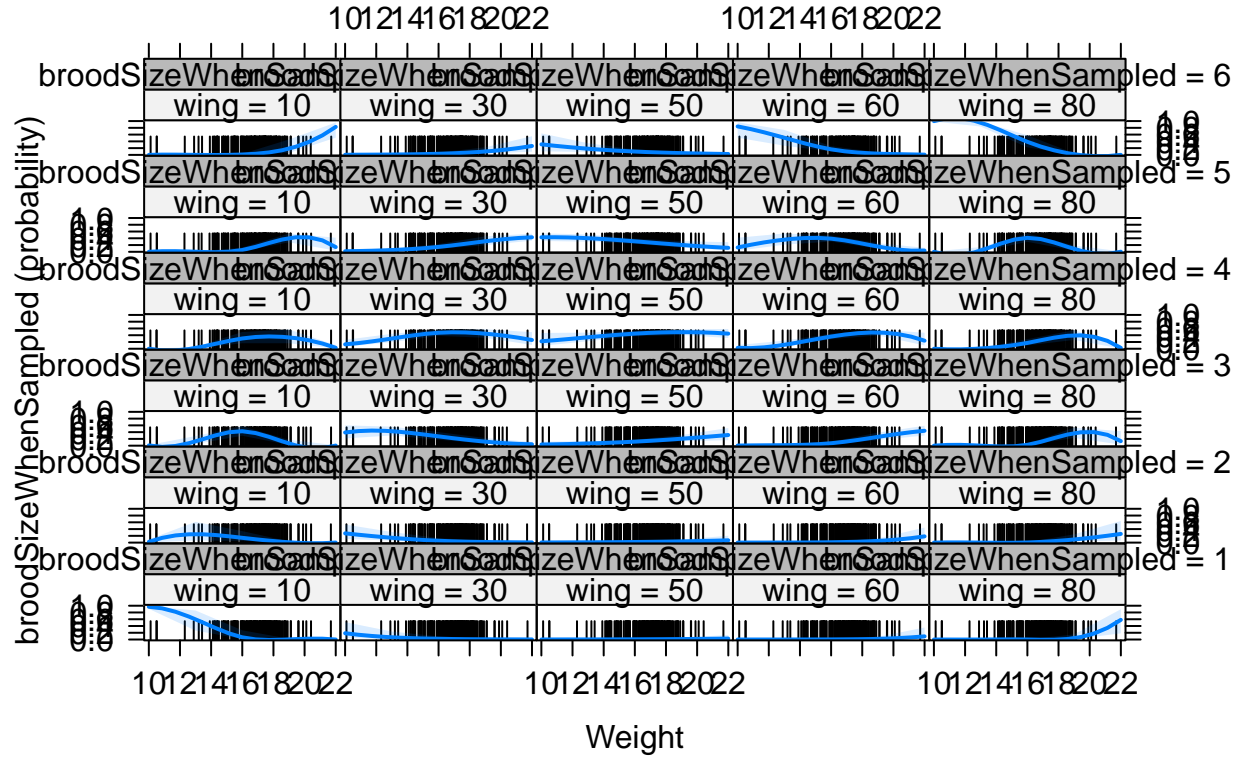
```
plot(Effect(focal.predictors = "wing", log_model))
```

wing effect plot



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

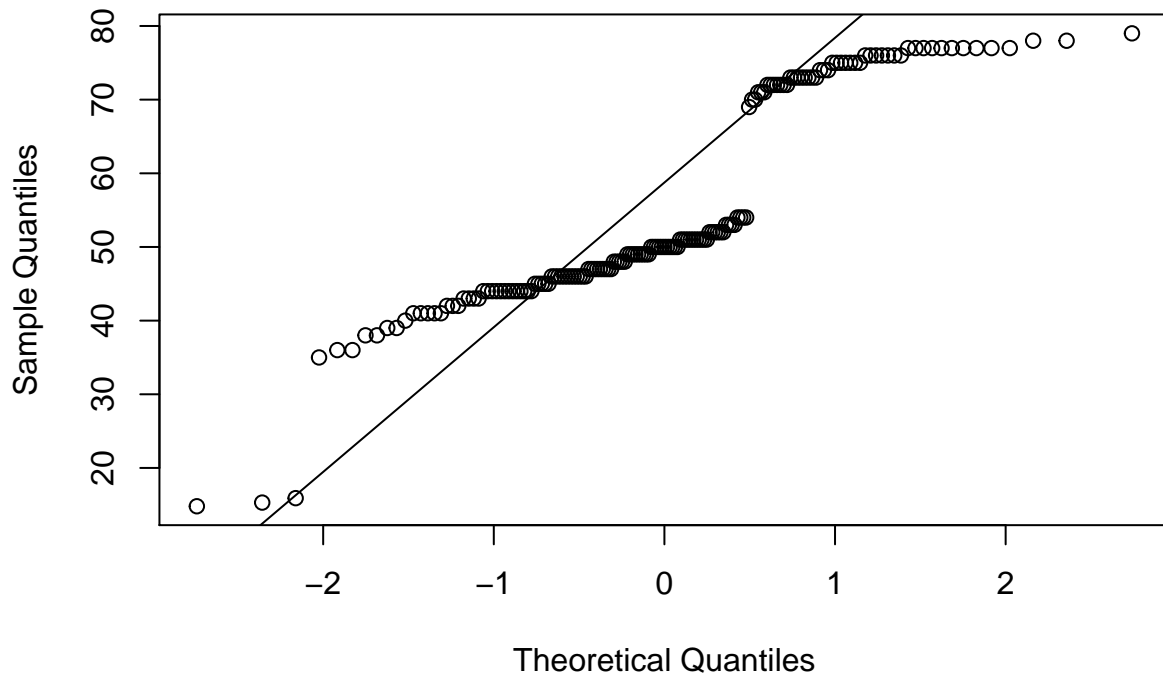
Weight*wing effect plot



Assumption Tests

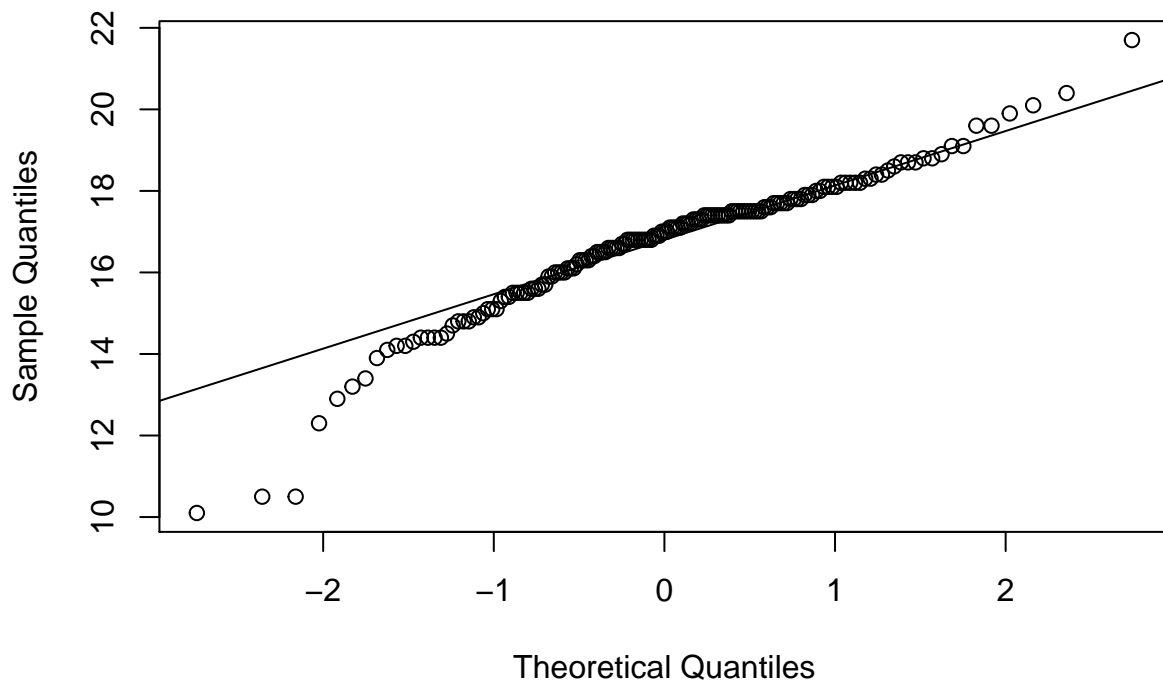
```
qqnorm(tit_dataOmit$wing)
qqline(tit_dataOmit$wing)
```


Normal Q-Q Plot



```
qqnorm(tit_data0mit$Weight)  
qqline(tit_data0mit$Weight)
```

Normal Q-Q Plot



```
shapiro.test(tit_data0mit$wing)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  tit_data0mit$wing  
## W = 0.86366, p-value = 5.376e-11
```

```
shapiro.test(tit_data0mit$Weight)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  tit_data0mit$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.