Test 2

Trevor VanVeldhuisen

November 03, 2021

```
birds <- read_csv('https://sldr.netlify.app/data/seabird-mercury.csv')
  glimpse(birds)
```

```
## Rows: 229
## Columns: 8
## $ bird_ID         <dbl> 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6,~
## $ total_mercury   <dbl> 15.015, 15.641, 6.867, 6.949, 12.497, 12.501, 8.275, ~
## $ foraging_habitat <dbl> -20.07, -20.01, -19.63, -20.33, -20.22, -19.99, -18.6~
## $ trophic_level   <dbl> 13.82, 13.76, 11.50, 10.25, 13.56, 13.91, 13.27, 13.5~
## $ age             <dbl> 35, 35, 35, 35, 35, 35, 32, 32, 32, 32, 32, 32, 32, 3~
## $ sex             <chr> "M", "M", "M", "M", "M", "M", "F", "F", "F", "F", "F"~
## $ breeding_history <chr> "Failed_Not_seen", "Failed_Not_seen", "Failed_Not_see~
## $ breeding        <chr> "Failed", "Failed", "Failed", "Successful", "Successf~
```

The specific research question I am going to answer is if mercury levels are associated with the success of breeding in grey-headed albatrosses near Antarctica. I am going to answer this question by fitting a binary regression model since the outcome of interest here is binary. The model plan includes the response variable: breeding success and the following set of predictors: total_mercury, breeding_history, age, foraging_habitat, trophic_level, sex. Total Mercury is the main predictor variable that we want to see if it has an association with breeding success. However, I included the rest of the variables that describe features of the albatrosses to see if these other factors might also be associated with breeding, as total mercury may not be the only factor and others may provide insight as to the impact of breeding success. After model fitting, the breeding history predictor provided peculiar estimates. Since it was not part of the primary question, I decided to remove breeding history from the model.

```
birds <- birds %>%
  mutate(breeding = factor(breeding),
         breeding_history = factor(breeding_history))
```

```
gf_boxplot( breeding ~ total_mercury, data = birds) %>%
  gf_labs( y = 'Breeding',
           x = 'Total Mercury Levels')
```
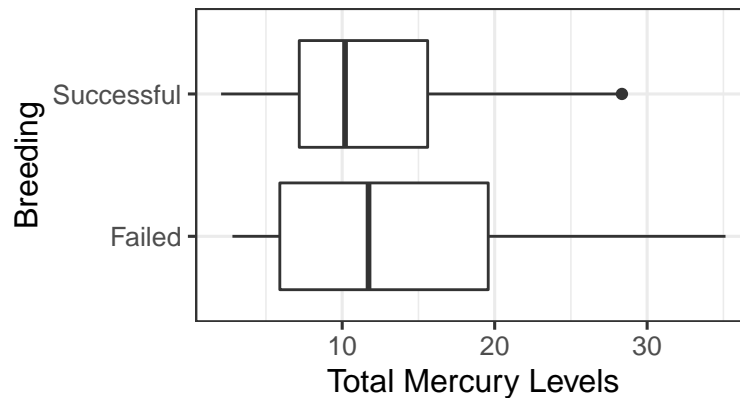
Data exploration/graphics: From the boxplot we learn that successful breeding tends to happen when there are lower levels of mercury. The median and the upper quartile of failed breeding are both greater than successful breeding. This graph supports the hypothesis that the total mercury levels may be associated with breeding success, that I will further assess by fitting a binary regression model to the data.

```r
birds.logr <- glm(factor(breeding) ~ total_mercury + foraging_habitat + trophic_level + age + sex,
                  data = birds,
                  family = binomial(link='logit'))
```

```r
summary(birds.logr)
```

```
##
## Call:
## glm(formula = factor(breeding) ~ total_mercury + foraging_habitat +
##     trophic_level + age + sex, family = binomial(link = "logit"),
##     data = birds)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2933  -0.8498  -0.7289   1.2358   1.9922
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       3.46859    3.93175   0.882   0.3777
## total_mercury    -0.04982    0.02195  -2.270   0.0232 *
## foraging_habitat  0.19921    0.13539   1.471   0.1412
## trophic_level     0.06085    0.13872   0.439   0.6609
## age              -0.03792    0.02808  -1.350   0.1769
## sexM              0.63694    0.32225   1.977   0.0481 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 278.58  on 228  degrees of freedom
## Residual deviance: 266.41  on 223  degrees of freedom
## AIC: 278.41
##
## Number of Fisher Scoring iterations: 4
```

$y_i \sim Binom(n_i, p_i)$ - the ith observation of the response trials. Here $n_i = 1$, I fit a binary regression model with the logit link. Here is the specific model: $logit(p_i) = 3.46859 -0.04982 total\_mercury + 0.19921 foraging\_habitat$

+ 0.06085trophic_level - 0.03792age + 0.63694sexM + E yi ~ Binom(n, pi)

I chose binary regression because I have a binary response variable: breeding. The logit function was also used because the outcome is binary. This logit function maps the probabilities from the range of 0 to 1.

```
require(DHARMa)
```

```
birds_sim <- (simulateResiduals(birds.logr))
```

```
gf_point(birds_sim$scaledResiduals ~ fitted(birds.logr),
         alpha = 0.2) %>%
  gf_labs(x = 'Fitted Values',
          y = 'Scaled Residuals')
```
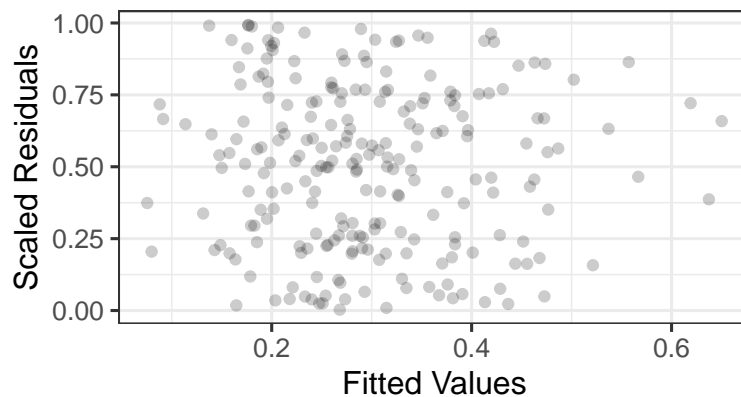

Through this scaled Residuals vs Fitted Values plot we see that there are no trends and the scaled residuals have uniform distribution showing it passes the mean-variance assessment.

```
car::Anova(birds.logr)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: factor(breeding)
##                  LR Chisq Df Pr(>Chisq)
## total_mercury     5.5309  1    0.01868 *
## foraging_habitat  2.1754  1    0.14023
## trophic_level     0.1941  1    0.65954
## age               1.8674  1    0.17178
## sex               4.0523  1    0.04411 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I used the hypothesis testing approach, with the null hypothesis being that there is no association between total mercury and breeding success (Odds ratio = 1). Through testing the null hypothesis, we have evidence to reject the null hypothesis since there is statistically significant association between total mercury and breeding success, with higher mercury levels associated with less breeding success. As part of this analysis that included other predictors, we also noted that the albatross sex was also associated with less breeding success.

```
require(s245)
```

```
get_fixed(birds.logr)
```

```
##   factor.breeding. total_mercury foraging_habitat trophic_level age sex
## 1          Failed        11.057          -20.08          13.19  22   M
```

```
pred_plot(birds.logr, 'total_mercury', data = birds) %>%
  gf_labs( y = 'Proportion Sucessful',
           x = 'Total Mercury')
```
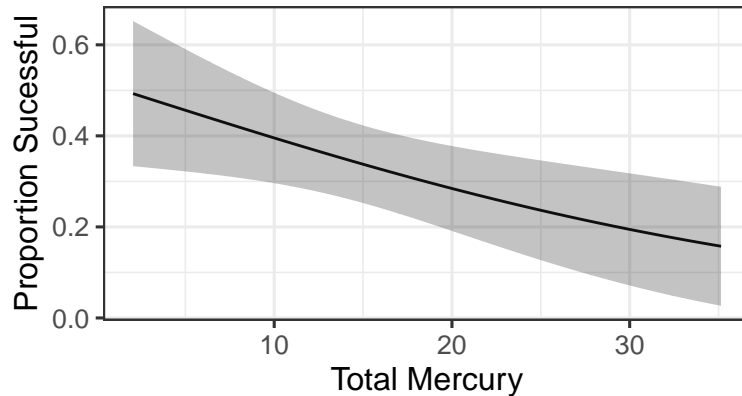

This prediction plot shows a negative correlation between successful breeding and total mercury levels. The more mercury content in a grey-headed albatross, the less likely they are able to successfully breed.

Looking at all of the data presented above, it seems that mercury content in grey-headed albatrosses does have a negative effect on the breeding success. First, looking at the boxplot, we can see the median mercury level is higher in albatrosses who were not successful in breeding. We can also observe the upper quartile extending beyond that of those who were successful at breeding. Second, from the binary regression model, the estimate of -0.04982 for mercury was statistically significant for breeding success (p-value = 0.0232), with the odds ratio of 0.95, indicating a 0.95 odds of success for every 1 unit increase in mercury. SexM was also a significant predictor with an estimate of 0.63694 (p-value = 0.0481) with an odds ratio of 1.8907. The rest of the predictors I chose: foraging_habitat, trophic_level, sex, and breeding_history, were not associated and had higher p-values > 0.05, suggesting they are not significant predictors. Third, the model passed the mean-variance test, as there were no patterns or trends, and the scaled residuals showed a uniform distribution, indicating a good fit to the model. Fourth, testing our null hypothesis of no age effect, we concluded to reject the null hypothesis since there is evidence of an association between total_mercury and breeding success. We also found that age and breeding success also have an association through and rejecting a null hypothesis of no age effect. Fifth, looking at the prediction plot we can see a negative correlation between total_mercury and the proportion of successful breeding. Lastly, looking at the odds ratio, the Beta for Mercury is -0.04982. The exponentiation of e^-0.04982 = 0.9514 is the odds ratio for a one unit change in mercury. The odds of breeding are 0.95 less for every one unit change in mercury showing a negative association between breeding success and total_mercury.