WU #21

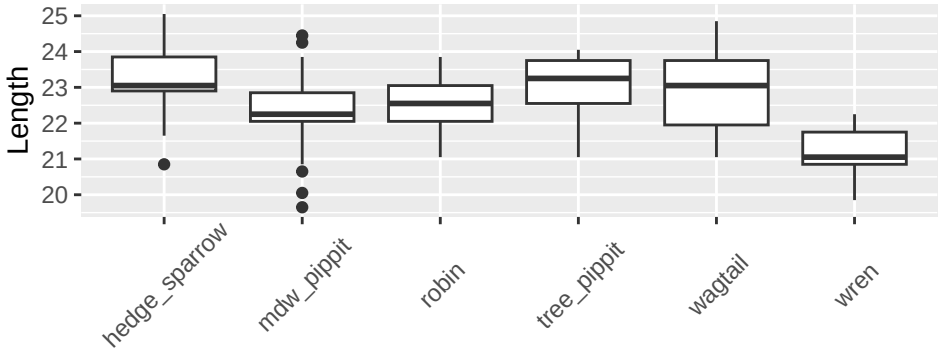
Math 58B, Spring 2023

Thursday, April 13, 2023

Your Name: _____

Names of people you worked with: _____

1. With your friends, would you rather play board games / cards or play video games?
2. Which models the size of the boxplot: MSG or MSE? How?
3. Using the model below, predict the average length of Cuckoo bird eggs for each of the 6 host birds.



Converting the `Bird` variable:

$$X_{mdw\_pippit} = \begin{cases} 1 & \text{if mdw\_pippit} \\ 0 & \text{otherwise} \end{cases} \quad X_{robin} = \begin{cases} 1 & \text{if robin} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{tree\_pippit} = \begin{cases} 1 & \text{if tree\_pippit} \\ 0 & \text{otherwise} \end{cases} \quad X_{wagtail} = \begin{cases} 1 & \text{if wagtail} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{wren} = \begin{cases} 1 & \text{if wren} \\ 0 & \text{otherwise} \end{cases}$$

So the model which describes the average egg length (denoted with the $\hat{Y}$ notation) can be written as the following:

$$\hat{Y} = 23.12 - 0.82 \cdot X_{mdw\_pippit} - 0.54 \cdot X_{robin} - 0.03 \cdot X_{tree\_pippit} + \\ - 0.21 \cdot X_{wagtail} - 1.99 \cdot X_{wren}$$

```
Cuckoo %>%
  lm(Length ~ Bird, data = .) %>%
  tidy()
```

```
## # A tibble: 6 x 5
##   term             estimate std.error statistic   p.value
##   <chr>               <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)        23.1       0.243    95.1    1.87e-110
## 2 Birdmdw_pippit     -0.823     0.278    -2.96   3.79e-  3
## 3 Birdrobin          -0.546     0.333    -1.64   1.03e-  1
## 4 Birdtree_pippit    -0.0314    0.338    -0.0930 9.26e-  1
## 5 Birdwagtail        -0.218     0.338    -0.645  5.20e-  1
## 6 Birdwren           -1.99      0.338    -5.89   3.91e-  8
```

Solution:

2. The size of the boxplot represents the variability of the data within each group. The variability is quantified by the value s_i which is used in the mean squared error.

$$MSE = \frac{\sum_{i=1}^{I} s_i^2(n_i - 1)}{N - I}$$

3. For each host bird, calculate the average predicted egg length by adding the relevant coefficient to the intercept. For the `hedge_sparrow`, use the intercept.

$$\begin{aligned}
\hat{Y}_{\texttt{hedge_sparrow}} &= 23.12 - 0.82 \cdot 0 - 0.54 \cdot 0 - 0.03 \cdot 0 - 0.21 \cdot 0 - 1.99 \cdot 0 = 23.12 \\
\hat{Y}_{\texttt{mdw_pippit}} &= 23.12 - 0.82 \cdot 1 - 0.54 \cdot 0 - 0.03 \cdot 0 - 0.21 \cdot 0 - 1.99 \cdot 0 = 22.30 \\
\hat{Y}_{\texttt{robin}} &= 23.12 - 0.82 \cdot 0 - 0.54 \cdot 1 - 0.03 \cdot 0 - 0.21 \cdot 0 - 1.99 \cdot 0 = 22.58 \\
\hat{Y}_{\texttt{tree_pippit}} &= 23.12 - 0.82 \cdot 0 - 0.54 \cdot 0 - 0.03 \cdot 1 - 0.21 \cdot 0 - 1.99 \cdot 0 = 23.09 \\
\hat{Y}_{\texttt{wagtail}} &= 23.12 - 0.82 \cdot 0 - 0.54 \cdot 0 - 0.03 \cdot 0 - 0.21 \cdot 1 - 1.99 \cdot 0 = 22.91 \\
\hat{Y}_{\texttt{wren}} &= 23.12 - 0.82 \cdot 0 - 0.54 \cdot 0 - 0.03 \cdot 0 - 0.21 \cdot 0 - 1.99 \cdot 1 = 21.13
\end{aligned}$$

```
Cuckoo %>%
  group_by(Bird) %>%
  summarize(mean_length = mean(Length))
```

```
## # A tibble: 6 x 2
##   Bird          mean_length
##   <fct>               <dbl>
## 1 hedge_sparrow        23.1
## 2 mdw_pippit           22.3
## 3 robin                22.6
## 4 tree_pippit          23.1
## 5 wagtail              22.9
## 6 wren                 21.1
```