Sparrows Dataset Analysis Report

Francisco Solis

Objectives:

1. Determining sample size and number of variables in the dataset
2. Examining strength, direction and relationship of variables in the scatterplot
3. Reporting of regression intercepts and slopes
4. Standard deviation and its degrees of freedom
5. Calculation of residuals and interpretation
6. Verifying model conditions via diagnostic plots
7. Identifying outliers and or influential points using the residuals and leverage diagnostic plots

*All content is referring to data in “Sparrows_Dataset_Analysis.pdf”

Discussion/Analysis

1. The total number of observations under consideration in the Sparrows dataset is 116. The total number of variables measured is 3 consisting of treatment, wing length and weight.

2. As seen in part b) the scatter plot of "Weight vs Wing length" the simple linear regression line follows a positive, moderately strong, linear relationship. A moderately strong, linear trend signals that there is a relationship between weight and wing length. Hence, for every unit increase in wing length, the weight is expected to increase as well.

3. After fitting a simple regression line to the data, the intercept is 1.36549 and the slope is 0.46740. Interpretation of each regression coefficient B_0 and B_1 is as follows,

   B_0: The expected weight of a sparrow given that the other predictor (Wing length) is 0. However, an expected weight of 1.36549g for a bird with a 0mm wing length is not practical so in this case, the intercept is not meaningful which is normal for some simple linear regression models.

   B_1: The expected weight is supposed to increase by 0.46740g per 1mm increase in wing length.

4. The standard deviation is 1.4 on 114 degrees of freedom. Meaning that on average the observation's weights deviate about 1.4 units from the mean.

5. Calculation of the residual when a sparrow's wing length is given as 26.0mm and its weight is 12.6g.

   The least squares regression equation is given by:

   Weight^= 1.36549 + 0.046740 (Wing Length) + error

   Residual= Y-Y^ (Actual value minus predicted value)
   =12.6 – (1.36549 + 0.046740*26.0)
   = -0.91789

A negative residual of -0.91789 indicates that the model's prediction was 0.91789g greater than the observed weight. Thus, the model slightly overestimated the weight given a sparrow with a wing length of 26.0 mm.

6. Based on the diagnostic plots in part d) the Residuals vs Fitted plot checks conditions of linearity and constant variance. In this case the data is relatively linear since the red line is along 0 so the linearity condition is satisfied. The data seems to create an inward fan shape indicating that the spread of data is heavier as the fitted values increase so constant variance is violated. Lastly the QQ-plot checks the normality of the data, in this case the data follows a relatively diagonal line indicating that the data is normally distributed. Thus, the normality condition of the data is satisfied. We check these conditions because further inference relies on these assumptions.

7. The Cook's distance plot checks for any influential points (influential points heavily affect the model), in this case there are no data points exceeding +/- 1 so there are no influential points. The Residuals vs Leverage plot checks for any outliers in the data, since no points exceed +/- 3, there are no outliers present in the data. It is important to check for outliers and influential points because unusual observations distort regression results.