Assignment 2

DANA 4810  
Alejandro Cervantes Nassar - 100388874

1. [1] List the variable(s) of interest in this study along with their types, list of categories, unit of measurements when appropriate. (Note: To identify how many categories and what categories the categorical variable(s) have, use an R code.)

* **Diet** – The type of diet that each gosling was subjected to. Categorical variable with two categories:
  + Plant diet with 33 observations.
  + Purina Duck Chow diet with 9 observations.
  + R code and output: Text

    Description automatically generated
* **WtChange** – Change in the weight of the gosling after 2.5 hours, in percentage of initial weight. Numerical and continuous variable. Response variable.
* **DigEff** – Digestion efficiency, measured as a percentage. Numerical and continuous variable. Explanatory variable.
* **ADFiber** – Amount of acid-detergent fiber in the digestive tract, measured as a percentage. Numerical and continuous variable. Explanatory variable.

1. [1] Create a scatterplot and describe a possible relationship between weight change and digestion

efficiency.

Chart, scatter chart

Description automatically generated

There seems to be a positive linear relationship between these two variables. The scatterplot indicates that the more efficient the digestion is, the more weight the goslings will gain.

1. [1] Calculate the coefficient of correlation relating weight change to digestion efficiency. Interpret this value.

With a correlation value of 0.61, we can say that there is a positive relationship between these categories. However, the relationship is not very strong.

1. [7] Conduct a test to determine whether weight change is correlated with a digestion efficiency. Use *α* = *.*01. (Use the critical value approach.)

Pearson’s test:

Text

Description automatically generated

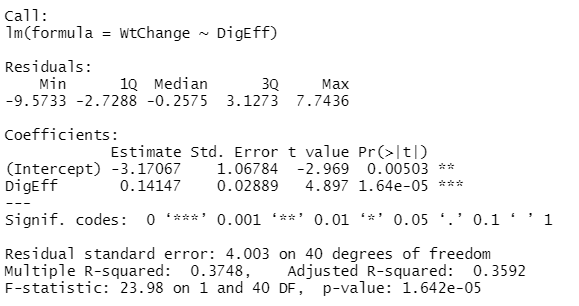
The t-value for alpha/2 = 0.10/2 and 40 degrees of freedom, is 1.684. Because the critical value of t in the Pearson test is 4.897, it falls in the H0’s rejection region. In other words, there is sufficient statistical evidence to conclude that the true correlation between these two variables is not equal to 0.

1. [4] Assuming the relationship between the variables mentioned in part (2) is best described by a straight line, use the method of least squares to estimate the *y*-intercept and slope of the line. (Also, provide the Least Squares Regression Line in the context of the question.)

y-intercept = -3.171

slope = 0.141

Least Regression Line: **estimated(Weight Change) = -3.171 + 0.141(Digestion efficiency)**



1. [2] Provide an interpretation of the slope estimate and intercept of the Least Square Regression Line in the context of the question.

The y-intercept of -3.171 means that if there is no change in the digestion efficiency (x=0), the goslings’ weight change would be on average -3.171.

The slope of 0.141 means that for every percent increase in the digestion efficiency, the goslings would increase on average 0.141% of their original weight.

1. [4] Find a 90% confidence interval for the slope and provide an interpretation for the calculated

interval.

With 90% confidence, we can say that the Least Square Regression line for the population will have a value between the following interval:

A picture containing text

Description automatically generated

1. [4] Find a 90% confidence interval for the mean weight change of all goslings with digestion efficiency of 17%. Interpret the result.

We are 90% confident that the average value of the weight change for all goslings when the digestion efficiency is equal to 17% will be between -1.987% and 0.455%:

Text

Description automatically generated

1. [4] Find a 90% prediction interval for the weight change of a gosling with digestion efficiency of 17%. Interpret the result.

We can predict with 90% confidence that the value of a single gosling’s weight change will be between -7.615% and 6.084% if its digestion efficiency is of 17%

Text

Description automatically generated with medium confidence

1. [5] Plot the least squares line on your scatterplot and depict 90% confidence intervals for mean values and prediction intervals for new values over the entire range of the regression line.

Chart, scatter chart

Description automatically generated

1. [1] Find the coefficient of determination from the R-Studio output and interpret the value.

About 36% of the sample’s variation in the goslings’ weight change can be explained by using their digestion efficiency in the straight-line model