Multiple Regression

Stat 230

## Example: Abalone age

Abalone are a type of mollusk that are harvested for their meat. Knowing the age of abalone can help set sustainable fishing quotas; however, the process is time consuming and labor intensive. Researchers collected various physical measurements, such as length, diameter, height, and weight, that are easy to collect to develop a method to predict abalone age. Below is R output for a multiple linear regression model predicting rings (a proxy for age) that uses length (mm), diameter (mm), height (mm), and whole\_weight (in grams) as predictor variables.

Call:  
lm(formula = rings ~ length + diameter + height + whole\_weight,   
 data = abalone)  
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 2.90861 0.30922 9.406 < 2e-16  
length -12.04857 2.10217 -5.731 1.07e-08  
diameter 25.64381 2.57384 9.963 < 2e-16  
height 20.24276 1.78180 11.361 < 2e-16  
whole\_weight 0.06589 0.22557 0.292 0.77   
---  
  
Residual standard error: 2.59 on 4172 degrees of freedom  
Multiple R-squared: 0.3556, Adjusted R-squared: 0.3549   
F-statistic: 575.4 on 4 and 4172 DF, p-value: < 2.2e-16

1. Report the estimated regression equation.

|  |
| --- |
| Solution |
| The estimated regression equation is: |

1. Using the model, predict the number of rings for an abalone with length 0.5 mm, diameter 0.4 mm, height 0.1 mm, and whole weight of 1 gram.

|  |
| --- |
| Solution |
| To make the prediction, we substitute the values into the regression equation: |

1. Interpret the slope coefficient for length in context.

|  |
| --- |
| Solution |
| The slope coefficient for length is -12.04857. This means that, holding all other predictor variables constant, for each additional millimeter increase in length, we expect the number of rings to decrease by approximately 12.05. |

1. State the hypotheses for the test of the slope coefficient for whole\_weight. What do you conclude based on the results of this test?

|  |
| --- |
| Solution |
| The hypotheses for the test of the slope coefficient for whole\_weight are:   * Null hypothesis (): (there is no association between whole weight and rings, controlling for other variables) * Alternative hypothesis (): (there is an association between whole weight and rings, controlling for other variables)   The p-value for the test is 0.77, which is very large. Therefore, we fail to reject the null hypothesis. There is no evidence to suggest that whole weight is associated with the number of rings, after accounting for length, diameter, and height. |

1. Calculate a 95% confidence interval for the slope coefficient for length. Interpret this interval in context.

|  |
| --- |
| Solution |
| To calculate the 95% confidence interval for the slope coefficient for length, we use the formula:  where is the estimated coefficient, is the standard error, and is the critical value from the t-distribution with 4172 degrees of freedom.  Using the output, we have:   * For a 95% confidence level and 4172 degrees of freedom, (using normal approximation for large df)   Thus, the confidence interval is:  We are 95% confident that the true slope coefficient for length is between -16.169 and -7.928. We are 95% confident that for each additional millimeter increase in length, the number of rings is expected to decrease by between 7.93 and 16.17, holding all other variables constant. |

1. What proportion of the variability in rings is explained by the model?

|  |
| --- |
| Solution |
| The proportion of variability in rings explained by the model is given by the R-squared value, which is 0.3556. This means that approximately 35.56% of the variability in the number of rings can be explained by the multiple regression model with the predictor variables length, diameter, height, and whole weight. |