Linear Regression/Linear Fit

Linear regression is used to model the relationship between 2 variables, a predictor variable (x) and the outcome variable (y). It is used to predict the value of y, when only x is known.

The mathematical model used to make this prediction is: y=b0 + b1 \* x + e

b0= intercept of the regression line: the predicted value of y when x=0

b1= the slope of the regression line

e= error: the difference between y data points from the regression line.

Residual Standard Error (RSE) is the average variation of points around the fitted regression line. The lower the RSE the better quality of the fitted regression model

In R, the command is used to fit a linear regression model to the data is lm

The abline function is used to create line on plot.

To get the correlation coefficient value, which measures the level of the association between x and y, use the R function cor()

-1 being a perfect negative correlation and 1 being a perfect positive correlation. 0 suggests very little relationship between variables.

To get a summary of the stats, use summary(), which includes:

Call= shows lm function used to fit the linear regression model

Residual= shows distribution of residuals

Coefficients=shows regression beta coefficients b0 and b1 and if they are significant

RSE, R-squared, and F-stat all show how the well the model fits the data