BS 803 Final Project
Nicholas Cvercko

The written function will allow the inputs of the data to be used for linear regression analysis, which column name is identified as the outcome variable, or which columns are identified as the predictor variables. The options for data input are as follows: data may be entered as either a data frame or a matrix, and the identification for both the outcome and predictor variables may be in the form of either a character list, numeric list, or an integer. If the incorrect objects are used as inputs for the function, or the column indices are outside of the data frame or matrix input, or there are mistakes in the spellings of column names, an error output or informative error return will alert the user of such mistakes. The default setting is for the function to remove missing values, however, this may be turned off if desired.

In addition to calculating the parameter estimates for all the predictors, standard errors for all predictors, and the R-Squared goodness of fit measure, the function will also return the t-test values and the associated p-values for each parameter estimate.

The following variables were used for testing functionality (the gala.txt data set is included in the .zip file):

gala is a data frame object for the data
g3 is the matrix representation of the gala data frame
ex is a character array variable which contains the column names for Area and Scruz
ex.n is a numeric array variable which contains the column indices for the same columns named Area and Scruz
gala2 is similar to the gala data frame, but with missing values

The function was named linear and the inputs are linear(x, outcome, predictors, na.rm = TRUE)
x is the argument for the data frame or matrix input
outcome is the argument for the name or index of the outcome variable
predictors is the argument for the names or indices of the predictor variables

The code displayed below to document the functionality and error messages is included in the .zip file

FUNCTIONALITY AND ACCURACY

R's built in lm() function output

```
> test <- lm(Species ~ Area + Scruz, data = gala)
> summary(test)

Call:
lm(formula = Species ~ Area + Scruz, data = gala)

Residuals:
    Min      1Q  Median      3Q     Max
-98.365 -57.515 -18.836   0.211 296.522

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 74.72730   23.17158   3.225  0.00329 **
Area         0.08049    0.01998   4.028  0.00041 ***
Scruz       -0.18533    0.25379  -0.730  0.47152
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 92.51 on 27 degrees of freedom
Multiple R-squared:  0.3937,    Adjusted R-squared:  0.3488
F-statistic: 8.766 on 2 and 27 DF,  p-value: 0.001165
```

Testing the outputs by using different R objects options as inputs:

```
> # Testing all the different types of R objects that do work as inputs
> linear(gala, 'Species', ex)
$`Parameter Characteristics`
          Parameter estimate Standard Error    t-value   Pr(>|t|)
Intercept        74.72730016    23.17158058  3.2249548 0.00328720
Area              0.08049258     0.01998102  4.0284520 0.00041046
Scruz            -0.18533154     0.25378828 -0.7302604 0.47152029

$`R-Squared Goodness of Fit Measure`
[1] 0.3937051

> linear(gala, 'Species', ex.n)
$`Parameter Characteristics`
          Parameter estimate Standard Error    t-value   Pr(>|t|)
Intercept        74.72730016    23.17158058  3.2249548 0.00328720
Area              0.08049258     0.01998102  4.0284520 0.00041046
Scruz            -0.18533154     0.25378828 -0.7302604 0.47152029

$`R-Squared Goodness of Fit Measure`
[1] 0.3937051

> linear(gala, 1, ex)
$`Parameter Characteristics`
          Parameter estimate Standard Error    t-value   Pr(>|t|)
Intercept        74.72730016    23.17158058  3.2249548 0.00328720
Area              0.08049258     0.01998102  4.0284520 0.00041046
Scruz            -0.18533154     0.25378828 -0.7302604 0.47152029

$`R-Squared Goodness of Fit Measure`
[1] 0.3937051
```

```
> linear(gala, 1, ex.n)
$`Parameter Characteristics`
          Parameter estimate Standard Error    t-value   Pr(>|t|)
Intercept        74.72730016    23.17158058  3.2249548 0.00328720
Area              0.08049258     0.01998102  4.0284520 0.00041046
Scruz            -0.18533154     0.25378828 -0.7302604 0.47152029

$`R-Squared Goodness of Fit Measure`
[1] 0.3937051

> linear(g3, 'Species', ex)
$`Parameter Characteristics`
          Parameter estimate Standard Error    t-value   Pr(>|t|)
Intercept        74.72730016    23.17158058  3.2249548 0.00328720
Area              0.08049258     0.01998102  4.0284520 0.00041046
Scruz            -0.18533154     0.25378828 -0.7302604 0.47152029

$`R-Squared Goodness of Fit Measure`
[1] 0.3937051

> linear(g3, 'Species', ex.n)
$`Parameter Characteristics`
          Parameter estimate Standard Error    t-value   Pr(>|t|)
Intercept        74.72730016    23.17158058  3.2249548 0.00328720
Area              0.08049258     0.01998102  4.0284520 0.00041046
Scruz            -0.18533154     0.25378828 -0.7302604 0.47152029

$`R-Squared Goodness of Fit Measure`
[1] 0.3937051

> linear(g3, 1, ex)
$`Parameter Characteristics`
          Parameter estimate Standard Error    t-value   Pr(>|t|)
Intercept        74.72730016    23.17158058  3.2249548 0.00328720
Area              0.08049258     0.01998102  4.0284520 0.00041046
Scruz            -0.18533154     0.25378828 -0.7302604 0.47152029

$`R-Squared Goodness of Fit Measure`
[1] 0.3937051

> linear(g3, 1, ex.n)
$`Parameter Characteristics`
          Parameter estimate Standard Error    t-value   Pr(>|t|)
Intercept        74.72730016    23.17158058  3.2249548 0.00328720
Area              0.08049258     0.01998102  4.0284520 0.00041046
Scruz            -0.18533154     0.25378828 -0.7302604 0.47152029

$`R-Squared Goodness of Fit Measure`
[1] 0.3937051
```

For every combination of input, the outputs are identical to the lm() function in R.

ERROR OUTPUTS:

If numeric inputs for outcome or predictor variables are outside of x indices:

```
> # If numeric inputs for outcome variable or predictor variables are outside of indices for matrix or data frame columns
> linear(gala, 20, ex)
```

```
Error in linear(gala, 20, ex) :
  The numeric input for the argument outcome is outside of the column indices of the input for the argument of x
```

Show Traceback
Rerun with Debug

```
> linear(gala, -1, ex)
```

```
Error in linear(gala, -1, ex) :
  The numeric input for the argument outcome is outside of the column indices of the input for the argument of x
```

Show Traceback
Rerun with Debug

```
> linear(gala, 1, c(2, 3, 20))
```

```
Error in linear(gala, 1, c(2, 3, 20)) :
  At least one of the numeric inputs for the argument predictors is outside of the column indices of the input for the argument of x
```

Show Traceback
Rerun with Debug

```
> linear(gala, 1, c(2, 3, -1))
```

```
Error in linear(gala, 1, c(2, 3, -1)) :
  At least one of the numeric inputs for the argument predictors is outside of the column indices of the input for the argument of x
```

Show Traceback
Rerun with Debug

If character inputs are inaccurate:

```
> # If character inputs for outcome variable or predictor variables are inaccurate
> linear(gala, 'Species', c('a', 'b'))
[1] "The following inputs for arguments for predictors are not column names in the argument input for x: "
[2] "a"
[3] "b"
> linear(gala, 'Species', c('a'))
```

```
Error in linear(gala, "Species", c("a")) : The variable a is not in x
```

Show Traceback
Rerun with Debug

```
> linear(g3, 'Species', c('b'))
```

```
Error in linear(g3, "Species", c("b")) : The variable b is not in x
```

Show Traceback
Rerun with Debug

```
> linear(gala, 'a', ex)
```

```
Error in linear(gala, "a", ex) :
  The input for argument outcome is not an existing column name in the input for argument x
```

Show Traceback
Rerun with Debug

```
> linear(g3, 'a', ex)
```

```
Error in linear(g3, "a", ex) :
  The input for argument outcome is not an existing column name in the input for argument x
```

Show Traceback
Rerun with Debug

If incorrect R objects are used for x:

```
> # If you do not use the correct R object for input arguments
> linear(c(1, 2, 3), 'Species', ex)
$`The function encountered the following errors`
[1] "Incorrect object input for the argument x, for this function it must be the data for the linear regression model, containing both the outcome variable and predictor variable values
either a matrix or a dataframe object in R."

> linear(gala, gala, ex)
$`The function encountered the following errors`
[1] "Incorrect object input for the argument outcome, for this function it must be the identifier for the desired outcome, this could be a character object of the column name or a numeri
c value of the column index in the data frame or matrix"

> linear(gala, 'Species', gala)
$`The function encountered the following errors`
[1] "Incorrect object input for the argument ppredictors, for this function it must be the identifiers for the desired predictor variables, this could be a character list of the column n
ames or numeric values of the column indeces in the data frame or matrix"

> linear(gala, gala, gala)
$`The function encountered the following errors`
[1] "Incorrect object input for the argument outcome, for this function it must be the identifier for the desired outcome, this could be a character object of the column name or a numeri
c value of the column index in the data frame or matrix"
[2] "Incorrect object input for the argument ppredictors, for this function it must be the identifiers for the desired predictor variables, this could be a character list of the column n
ames or numeric values of the column indeces in the data frame or matrix"

> linear(c(1, 2, 3), gala, gala)
$`The function encountered the following errors`
[1] "Incorrect object input for the argument x, for this function it must be the data for the linear regression model, containing both the outcome variable and predictor variable values
either a matrix or a dataframe object in R."
[2] "Incorrect object input for the argument outcome, for this function it must be the identifier for the desired outcome, this could be a character object of the column name or a numeri
c value of the column index in the data frame or matrix"
[3] "Incorrect object input for the argument ppredictors, for this function it must be the identifiers for the desired predictor variables, this could be a character list of the column n
ames or numeric values of the column indeces in the data frame or matrix"
```

Removing or Leaving in NAs

```
> # If you input an na value, they will be removed unless requested
> linear(gala, 'Species', ex, na.rm = FALSE)
$`Parameter Characteristics`
          Parameter estimate Standard Error    t-value   Pr(>|t|)
Intercept        74.72730016    23.17158058  3.2249548 0.00328720
Area              0.08049258     0.01998102  4.0284520 0.00041046
Scruz            -0.18533154     0.25378828 -0.7302604 0.47152029

$`R-Squared Goodness of Fit Measure`
[1] 0.3937051

> linear(gala2, 'Species', ex)
$`Parameter Characteristics`
          Parameter estimate Standard Error    t-value   Pr(>|t|)
Intercept        75.96007734    24.35403030  3.1189941 0.00440043
Area              0.08020653     0.02039349  3.9329477 0.00055671
Scruz            -0.19403766     0.26190837 -0.7408609 0.46541672

$`R-Squared Goodness of Fit Measure`
[1] 0.3934549

> linear(gala2, 'Species', ex, na.rm = FALSE)
$`Parameter Characteristics`
          Parameter estimate Standard Error t-value Pr(>|t|)
Intercept                 NA             NA      NA       NA
Area                      NA             NA      NA       NA
Scruz                     NA             NA      NA       NA

$`R-Squared Goodness of Fit Measure`
[1] NA
```