MINI EXERCISE

1. By creating a histogram for all the variables, we have the following.

Chart, box and whisker chart

Description automatically generated

There are no missing values for any other variables. There is a severe outlier for height. Height is reported in each, and it is highly unlikely that man is 29.5 inches and weighs 205 pounds. There was likely a typographical error, so this observation should be healing and excluded from the analysis period from the remainder of the exercise this outlier is headed and excluded there are other outlines but the range of values for all the variables look reasonable

1. Creating a scatterplot for all the variables, we have the following

A picture containing table

Description automatically generated

From the matrix ports above chest and abdomen is strongly correlated with %Fat. There are many correlations between the predictors. Some points seem to be outliers for many other variables.

1. By fitting a full model with % Fat as the response in all the other variables as predictors, we have the following;

Table

Description automatically generated

It appears that Abdomen and Wrist are the most important variables.

1. By using VIF to investigate multicollinearity in the model we have the following.

A close-up of a document

Description automatically generated with low confidence

From the parameter estimates above, it can be realized that the variance inflation factor for weight is 52.6643 which indicates that there is a multicollinearity. We fit this model by removing weight from the equation. we have the reduce model below;

Table

Description automatically generated

From this we considered the issue of multicollinearity is fixed.

1. We analyze the residuals, and we have the following plots

Chart, scatter chart

Description automatically generated

Constance variance is seen here.

Chart, histogram

Description automatically generated

Normally distributed here.

1. Cook D influence

Chart

Description automatically generated

It appears that there are two observations that have extreme cooks D values relating to the other observations but none of the Cook D values is greater than one.

1. ANOVA TABLE

Table

Description automatically generated

Text

Description automatically generated with medium confidence

The model is significant. they adjusted Rsquared is 0.732877 and the root mean square error RMSE is 4.325296.

1. By using the effect summary table to slowly remove nonsignificant terms from the model using a cutoff P value of 0.05 we have the following,

Graphical user interface, text, table

Description automatically generated

This is better.