The data set is imported using appropriate R commands and the summary of the data set is carried out.

Each variable (column) is analyzed and their means, SD’s are notes to find out their behavior.

Boxplots and bar plots are used for each of the columns present in the data set. Corrgram is also drawn.

Analyses and inferences regarding their distribution can be drawn.

Then the variables are analyzed pair-wise to note any correlation present which can explain the price difference between economy and premium economy.

The flight duration along with the price (economy and premium economy) shows a relatively high correlation (>0.55)

**HYPOTHESIS 1:**

**The flight duration and the Economy price is linearly related.**

The flight duration and the Economy price are observed to have a correlation of 0.5666404.

Therefore a regression model is formulated based on simple linear regression regarding dependency of Economy price on flight duration. The “lm” command is used for this purpose.

A scatter plot is drawn for better visual representation of the

Correlated variables.

Then based on our model, we predict the Economy Prices

based on the duration of the flight.

fitted command is used to view the predicted values according

to the formula used in our model.

Residuals command is used to view the residual values in the

fitted model.

confint method provides confidence interval for model

parameters.

**Regression model formulated is**

**PriceEconomy= 158.10 (FlightDuration) + 129.03**

Some inferences can be drawn from the above regression model.

1. The regression coefficient of 158.10

Significantly greater than 0(p value << 0.001) and there is an expected increase Of 158.10 units of price for every 1 hour

increase in flight duration.

1. Multiple R squared indicates model accounts for 32.1%

variance in Economy prices.

1. The residual standard error(815.2) in predicting the Economy prices from the flight duration
2. The F statistic test predict whether the predictor variables taken together is able to predict the response variables or not. In this case as there is only one predictor variable, the F test is similar to the T test.

A T test is later conducted with the following conditions.

*H0 is that the difference between the means of the flight duration and the Economy Price should be EQUAL to zero.*

*H1 says the difference between the means of the Economy Price and the flight duration is NOT EQUAL to zero.*

**A very low p value due to which we reject the null hypothesis and accept the alternate hypothesis.**

**HYPOTHESIS 2:**

**The flight duration and the Premium price is linearly related.**

The flight duration and the Premium price are observed to have a correlation of 0.6487398

Therefore a regression model is formulated based on simple linear regression regarding dependency of Premium price on flight duration. The “lm” command is used for this purpose.

A scatter plot is drawn for better visual representation of the

Correlated variables.

Then based on our model, we predict the Premium Prices

based on the duration of the flight.

fitted command is used to view the predicted values according

to the formula used in our model.

Residuals command is used to view the residual values in the

fitted model.

confint method provides confidence interval for model

parameters.

**Regression model formulated is**

**PricePremium=** 235.93 **(FlightDuration) + 57.45**

Some inferences can be drawn from the above regression model.

1. The regression coefficient of 235.93

Significantly greater than 0(p value << 0.001) and there is an expected increase Of 235.93 units of price for every 1 hour

increase in flight duration.

1. Multiple R squared indicates model accounts for 42.09%

variance in Premium prices.

1. The residual standard error(981.4) in predicting the Premium prices from the flight duration
2. The F statistic test predict whether the predictor variables taken together is able to predict the response variables or not. In this case as there is only one predictor variable, the F test is similar to the T test.

A T test is later conducted with the following conditions.

*H0 is that the difference between the means of the flight duration and the Premium Price should be EQUAL to zero.*

*H1 says the difference between the means of the Premium Price and the flight duration is NOT EQUAL to zero.*

**A very low p value due to which we reject the null hypothesis and accept the alternate hypothesis.**