**ANALYSIS OF AIRLINES DATA**

**What is our data?**

The data set which was given to us had 462 observations and 18 variables named accordingly. We have to fit a model taking independent values as regressor/independent variable and PriceRelative as a dependent variable.

Out of these 18 variables four of them were categorical, and rest of them were numerical (9 independent and 5 dependent). We divide the categorical variables into dummy indicator variables.

**Early Observations from Boxplots**

1. Out of the six airlines, Virgin, Singapore, Jet and British have higher relative prices than the others.
2. Boeing Planes have higher relative prices than AirBus.
3. International flights have higher relative prices.
4. Flights with shorter length have shown an increase in relative prices followed by a gradual decline when flights are longer than 9 hours.

**T-tests on Hypotheses**

**Hypothesis 1:** The average cost of a premium ticket is more than that of economy ticket.

**Assumptions:** The data sets is normally distributed. The data points are independent of each other. The vectors are numeric here.

**Result:** We reject that there average cost of premium and economy ticket is same.

**Hypothesis 2:** Boeing planes have higher relative ticket costs than AirBus carriers.

**Assumptions**: The data sets is normally distributed. The data points are independent of each other. Relative price is numeric and Aircraft is a dichotomous variable

**Result:** We reject the hypothesis that Boeing and AirBus planes have equal relative costs.

**Hypothesis 3:** International flights have higher relative ticket costs than domestic planes.

**Assumptions:** The data sets is normally distributed. The data points are independent of each other. Relative price is numeric and Flight type is a dichotomous variable.

**Result:** We reject that international and domestic planes have similar pricing.

**Regression model**

* 5 variables were not defined because of singularities. We have these things because data columns have shown perfect collinearity with some other data column. For example, the indicator columns for **IsInternational** column has just two variables whose corresponding indicator will show perfect collinearity.
* The relative price cannot be zero as then there would no difference between premium and economy seats and it would make no sense to have premium seat if there is no extra profit. The intercept is just an adjustment term.
* The multiple R-squared (0.4803) indicates that the model accounts for 48.03% of the variance in Relative Price.
* The residual standard error (0.3328) is the average error in predicting Relative Price using this model.
* The p-value is 2.2e-16. So the model has a 95% confidence interval.