## Report on Airline Pricing Analysis

### Pratyusha Maiti

#### Email: [pratyusha.maiti@gmail.com](mailto:pratyusha.maiti@gmail.com)

#### Jadavpur University

## Comparative study using boxplots

*par(mfrow = c(3,2))*

*boxplot(airlines$PRICE\_PREMIUM, horizontal = TRUE, xlab = "Premium price")*

*boxplot(airlines$PRICE\_ECONOMY, horizontal = TRUE, xlab = "Economy price")*

*boxplot(airlines$WIDTH\_PREMIUM, horizontal = TRUE, xlab = "Width of premium seats")*

*boxplot(airlines$WIDTH\_ECONOMY, horizontal = TRUE, xlab = "Width of economy seats")*

*boxplot(airlines$PITCH\_PREMIUM, horizontal = TRUE, xlab = "Pitch of premium seats")*

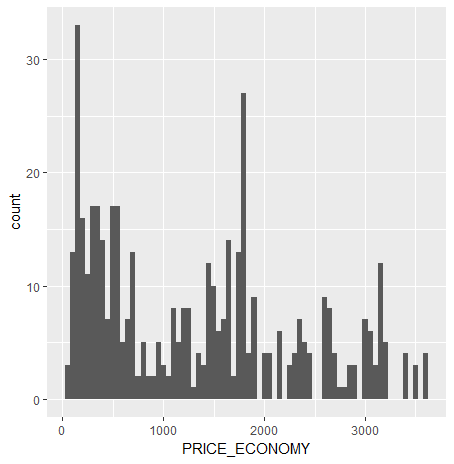
*boxplot(airlines$PITCH\_ECONOMY, horizontal = TRUE, xlab = "Pitch of economy seats")*

The above code allowed a side-by-side representation of prices of premium and economy seats with respect to the seat width and pitch.

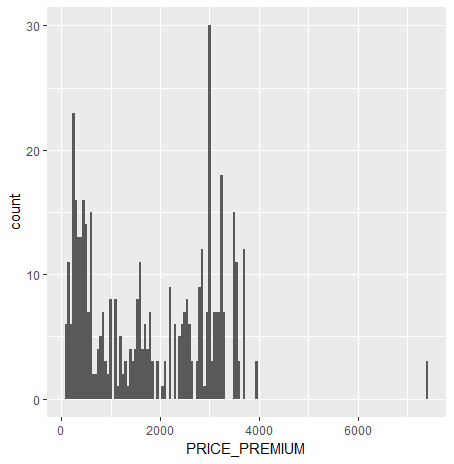
The resulting plot showed that a major difference in price points of the two was **due to seat dimensions**.

## Histogram showing seats vs prices

*ggplot(airlines, aes(x = PRICE\_ECONOMY, fill = SEATS\_ECONOMY)) + geom\_histogram(binwidth = 50)*



*ggplot(airlines, aes(x = PRICE\_PREMIUM, fill = SEATS\_PREMIUM)) + geom\_histogram(binwidth = 50)*



The plots show that there is no fixed dependence of seat price with number of seats. A closer look at the data shows that for different airlines, prices differ with flight duration irrespective of the number of seats.

This leads to the question whether the number of seats is a completely redundant term then?

Hence, we perform correlation tests over both economy and premium data sets.

*#****Hypothesis 1a: Number of seats does not affect pricing of economy seats***

*cor.test(PRICE\_ECONOMY,SEATS\_ECONOMY+PITCH\_ECONOMY+WIDTH\_ECONOMY, data = airlines)*

*cor.test(PRICE\_ECONOMY,PITCH\_ECONOMY+WIDTH\_ECONOMY, data = airlines)*

*#****Hypothesis 1b: Number of seats affects pricing of premium seats***

*cor.test(PRICE\_PREMIUM,SEATS\_PREMIUM+PITCH\_PREMIUM+WIDTH\_PREMIUM, data = airlines)*

*cor.test(PRICE\_PREMIUM,PITCH\_PREMIUM+WIDTH\_PREMIUM, data = airlines)*

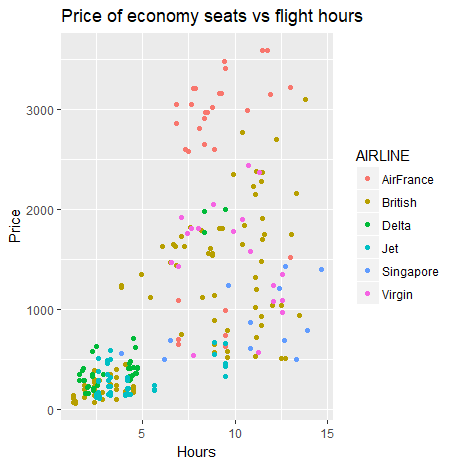
The first result shows that on including number of seats in the formula, the correlation factor **decreases** for economy seat pricing.

However, on running the same for premium, it seems that the number of seats DOES have an effect on its price.

## Scatterplot to study the effect of flight duration on pricing wrt individual airlines

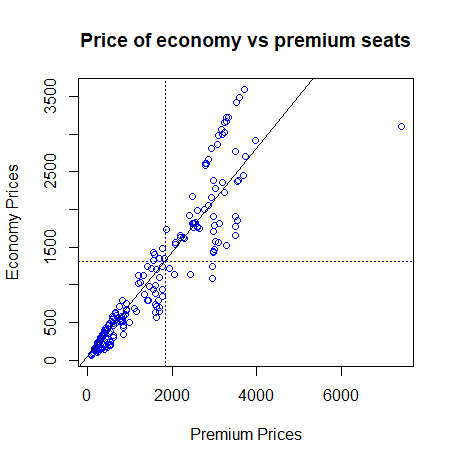
Having confirmed that the seat dimensions are more important in pricing difference, we now proceed to show how different companies have different price points.

For example, the economy seats pricing model shows the following:

This clearly shows that AirFrance has a higher price point compared to say, Delta or Jet. British Airways are meanwhile, distributed over the entire duration graph.

On running the same test on premium seats data, we have a similar result. Again, AirFrance has peak-pricing, British Airways have a general spread-over pricing and Delta and Jet being cheap.

This led to the intuitive conclusion that there is some linear relation between the pricing of economy and premium seats for all these companies. We run a regression model and plot a regression line to see a pretty well-fitted regression line depicting linear relationship.



## Regression Model with all factors affecting seat pricing

*regression\_economy<-lm(PRICE\_ECONOMY~FLIGHT\_DURATION+SEATS\_PREMIUM+PITCH\_PREMIUM+WIDTH\_PREMIUM+QUALITY+MONTH+AIRCRAFT+AIRLINE,*

*data=airlines)*

The regression model is built using factors:

* Flight duration
* Number of seats
* Seat dimensions (width and pitch)
* Flight quality
* Airline company

We obtain **p-value: < 2.2e-16** for both economy and premium models.This successfully eliminates any null hypothesis scenario. We can safely conclude that all of the above factors are important in pricing of a seat.