**RegionEx Case**

**Question1:**

Compute the mean, median, 90th percentile, and standard deviation of the

arrival delay minutes of RegionEx’s flights. Do the same for MDA’s flights. How do the

two airlines compare? What’s your conclusion from this quick analysis?

**Statistic summary obtained from R**

**Airline Mean Median 90th**

**percentile**

**Standard**

**Deviation n( )**

RegionEx : Mean:15.66, Median: 9, 90th Percentile:21, Standard Deviation:27.65

MDA: Mean:10.92,Median:13, 90th Percentile:16.4, Standard Deviation:6.42

**Comparison and conclusion:**

According to the information above, the arrival delay distribution of RegionEx is

right-skewed while that of MDA Airline is left-skewed. On average, MDA has a lower

arrival delay in minutes when compared to RegionEx.

Also, the distribution of MDA Airline Arrival delay clusters much closer than that of

RegionEx Airline based on the comparison of standard deviation.

**Question2:**

Inspect the distribution of RegionEx’s arrival delays by constructing a histogram (or a

frequency distribution) of the arrival delay minutes of RegionEx’s flights. Do the same

for MDA’s flights. How do these two distributions compare? See if your conclusions from

comparison of descriptive statistics (in Question 1) is consistent with these histograms.

What, if any, additional information do the histograms provide?

**Frequency Distribution of Flight Arrival delay minutes (RegionEx)**

**Frequency Distribution of Flight Arrival delay minutes (MDA)**

**Comparison and Additional information:**

These histograms reconfirm the skewness mentioned at answer of Question1 and

provide more explanation for the large difference in standard deviation.

We could tell from the histogram that the distribution of arrival delay minutes for

RegionEx is definitely not a normal distribution but a right-skewed distribution. We could

observe from the histogram of RegionEx that most data are clustered around -12.5

minutes to 27.5 minutes, but there are small amounts of extremely large data spread

out from 57.5 minutes to 157.5 minutes, which explains the large standard deviation.

As for MDA's flight, it seems to be relatively normal distribution when compared to

RegionEx, but we can still see a little left-skewed tail from the graph. When we compare

two histograms, the distribution of RegionEx's flight spread much wider than MDA's

flight.

Also, the histogram also indicates that both airlines had flights that arrived earlier than

the scheduled arrival time.

**Question3:**

Using the FAA definition of a “late” flight, what percentage of RegionEx’s September

flights were “late”? What percentage of MDA’s September flights were “late”? What

percentage were “on-time” for each airline, according to the FAA definition?

(Note: The data already incorporates the FAA definition and calls it “Delay indicator”).

**Result obtained from R calculation**

**Airline “Late” Percentage “On-time” Percentage**

RegionEx 26.25% 73.75%

MDA 26.50% 73.50%

**Question4:**

Compare the performance of the two airlines on each flight leg by calculating the

descriptive statistics (mean, median, 90th percentile, standard deviation) of delay

minutes for each of the four routes. Calculate the % delay flights for each of the route.

Do any of the comparisons change? If so, why?

There are four flight route: DFW-MSY, MSY-DFW, MSY-PNS and PNS-MSY. We

subset the original dataset into these four routes in order to obtain the output below.

**DFW to MSY Route**

**MSY to DFW Route**

For the first two route, although the average delay time of RegionEx Airline is longer

than that of MDA Airline, MDA Airline's passengers have a higher chance to be on a

delayed flight compared to Region Ex's.

Also, we notice from the standard deviation and 90th percentile that the mean arrival

delay data of RegionEx is influenced by some extreme data.

**MSY to PNS Route**

**PNS to MSY Route**

For MSY-PNS and PNS-MSY route, flights of both airlines have the same chance to

delay based on the identical percentage of delay.Therefore, it would be better for the

passengers to go on MDA's flights because it has less delay time than RegionEx's

flights on average.

To conclude, the minute delayed of RegionEx's flight on both 4 routes are generally

larger than that of MDA's flight. Especially from DFW to MSY and MSY to DFW route.

As for MSY to PNE route, the mean of delayed time of both airlines are really close.

Also, by checking the median of delayed minute for this route, RegionEx is right skewed

distribution but the distribution of MDA for this route is slightly left skewed. It indicate

that the delayed minutes for MDA’s flight tends to be

smaller than the mean for RegionEx flight on this route.

**Question5:**

Consider only the RegionEx flights. Prepare a scatter plot of arrival delay minutes

versus number of passengers. Your scatter plot should consist of 240 data points, one

for each flight in the data set where the vertical coordinate is arrival delay minutes of

that flight and the horizontal coordinate is the number of passengers. What is the

correlation coefficient between arrival delay minutes and number of passengers for

RegionEx’s flights? Interpret your results.

**Scatter plot of Arrival Delay Minutes versus Number of Passengers**

The correlation coefficient between arrival delay minutes and number of passengers for

RegionEx's flights is 0.485. The arrival delay time in minutes shows a moderate rise

with the increase in no. of passengers as observed in the scatter plot.

It shows that the arrival delay minute and number of passengers with moderate positive

linear relationship. When the number of passengers increase, the arrival delay minutes

would possibly increase too, but we do not know if there is a causal relationship

between these two variables.

**Question 6:**

Compare the scheduled flight durations for the 2 airlines on each of their 4

routes.Compare actual flight durations.What do you notice?

If two airlines had the same scheduled durations,what impact would it have on their

delay records?

**DFW to MSY Route**

Mean of scheduled duration and Actual duration for both airlines:

If scheduled flight duration is made equal for both airlines, the delay times for RegionEx

flight will be reduced by increase in flight duration which is 10 mins

**Original:**

Actual arrival - Scheduled arrival = Minutes delayed

**New(Made equal):**

Actual arrival - Scheduled arrival -10= Minutes delayed -10

new\_delay\_R1Rx shows delay times for RegionEx airlines on route1, in case

Scheduled durations for both airlines were made equal.

Based on the FAA definition of late flights find out revised number of delayed flight for

RegionEx on Route 1:

Total no. of delayed flights for RegionEx on this route after revised schedule travel time:

**5**

Total no. of flights delayed on R1 with same scheduled durations for MDA airlines= **3.**

Delay ratio for RegionEx after revised schedule time: 5.555%

Delay ratio for MDA : 10.344%

**MSY to DFW Route**

Mean of scheduled duration and Actual duration of both airlines:

If scheduled flight duration is made equal for both airlines, the delay times for RegionEx

flight will be reduced by increase in flight duration which is 10 mins

Based on the FAA definition of late flights find out revised number of delayed flight for

RegionEx

Here,RegionEx on an average, actually saves (10-4.6)= 5.4 mins as compared to MDA.

Total no. of delayed flights for RegionEx on this route after revised schedule travel time:

= **8** .

Total no. of delayed flights for MDA on this route = **3** .

Delay ratio for RegionEx after revised schedule time: 8.8889%

Delay ratio for MDA : 10.348%

**MSY to PNS Route**

Mean of scheduled duration and Actual duration of both airlines:

If scheduled flight duration is made equal for both airlines, the delay times for RegionEx

flight will be reduced by increase in flight duration which is 5 mins.

Based on the FAA definition of late flights find out revised number of delayed flight for

RegionEx on Route 3

Total no. of delayed flights for RegionEx on this route after revised schedule travel time

= **2** .

Total no. of delayed flights for MDA on this route = **6.**

Delay ratio for RegionEx after revised schedule time: 6.6%

Delay ratio for MDA : 20%

**PNS to MSY Route**

Mean of scheduled duration and Actual duration of both airlines:

If scheduled flight duration is made equal for both airlines, the delay minutes for

RegionEx flight will be reduced by increase in flight duration which is 5 mins

Total no. of delayed flights for RegionEx on this route after revised schedule travel time

= **2.**

Total no. of delayed flights for MDA on this route : **6**

Delay ratio for RegionEx after revised schedule time: 6.666%

Delay ratio for MDA : 20%

**Delay Ratio for RegionEx after revised schedule flight time (4 route combine) :**

**(17/240)\*100%= 7.083%**

**Delay Ratio for MDA (4 route combine) :**

**(18/117)\*100%= 15.384%**

To conclude, after revising the scheduled flight time, it indicates that the delay ratio of

RegionEx is smaller than that of MDA. If the scheduled flight duration is the same, it

seems that there is a higher probability that MDA’s flight will be delayed.