Bangabandhu Sheikh Mujibur Rahman Aviation and Aerospace University

Statistical Analysis of United Airlines Aircraft Operating Statistics

Probability and Statistics

MAT 4509

Md. Rafi

ID: 22024010

Instructor: Assoc Prof Dr. M Siddikur Rahman

Submission Date: 12/11/2024

Table of Contents

Question 01: Frequency Distribution of “Salaries and Wages”

The data included information for the Total fleet, Large narrowbodies, Small narrowbodies, and wide bodies. B2 to W158 were the relevant data points. Using the following code, the data was loaded and analyzed.

Table I: Code for the Q1

```
library(readxl)
library(fdth)

######################################
########## Data Loading ############
######################################
data <- read_excel("B:/BSMRAAU 22024010 5th/MAT_4509/Programming/United Airlines Aircraft Operating Statistics- Cost Per Block Hour (Unadjusted).xls", range="b2:w158")
data <- as.data.frame(data)

fleets <- c("Small Narrowbodies", "Large Narrowbodies", "Widebodies", "Total Fleet")
wages_rows <- c(8, 47, 86, 125) - 2

######################################
########## Functions ###############
######################################
get_row <- function(row, data) {
  na.omit(as.numeric(data[row, -1]))[1:10]
}

# Function to create and print frequency distribution tables
create_table <- function(data, fleet_name) {
  interval <- as.integer((max(data) - min(data)) / (log2(length(data)) + 1))
  table <- fdt(data, start = min(data), end = max(data), h = interval)
  print(paste("Frequency Distribution Table for", fleet_name))
  print(as.data.frame(table$table))  # Convert to data frame and print for readability
}

######################################
############ Main code #############
######################################
for (i in seq_along(wages_rows)) {
  pilots_wages <- get_row(wages_rows[i], data)
  create_table(pilots_wages, fleets[i])
}
```

The output of this code provides following frequency distribution tables for small narrowbodies, large narrowbodies, wide bodies and total fleet.

Table I.I: Small Narrowbodies

Class Intervals	Frequency
326.982-381.982	2
381.982-436.982	2
436.982-491.982	4
491.982-546.982	0

Table I.II: Large Narrowbodies

Class Intervals	Frequency
392.86-452.86	2
452.86-512.86	1
512.86-572.86	3
572.86-632.86	3

Table I.III: Widebodies

Class Intervals	Frequency
652.931-762.931	3
762.931-872.931	1
872.931-982.931	2

Table I.IV: Total Fleet

Class Intervals	Frequency
429.84-504.84	2
504.84-579.84	2
579.84-654.84	4
654.84-729.84	1

Question 02: Central Tendency of “Salaries and Wages”

Mean, Median, Mode, Standard Deviation, Variance, Quartiles, 9th Decile, 10th Percentile and Range of “Salaries and Wages”

Table II: Code for the Q2

```
library(readxl)

# Load data
data <- read_excel("B:/BSMRAAU 22024010 5th/MAT_4509/Programming/United Airlines Aircraft Operating Statistics- Cost Per Block Hour (Unadjusted).xls", range = "b2:w158")
data <- as.data.frame(data)

# Define rows to analyze
wages_rows <- c(8, 47, 86, 125) - 2
fleets <- c("Small Narrowbodies", "Large Narrowbodies", "Widebodies", "Total Fleet")

# Function to get the first 10 non-NA values from a specified row
get_row <- function(row, data) {
  na.omit(as.numeric(data[row, -1]))[1:10]
}

# Function to calculate mode
get_mode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
```

```
# Function to format numbers to 3 decimal places
format_number <- function(x) {
  format(round(x, 3), nsmall = 3)
}

# Calculate statistics for specified rows
results_list <- lapply(wages_rows, function(i) {
  wages <- get_row(i, data)
  list(
    mean = format_number(mean(wages, na.rm = TRUE)),
    mode = format_number(get_mode(wages)),
    median = format_number(median(wages, na.rm = TRUE)),
    range = format_number(max(wages, na.rm = TRUE) - min(wages, na.rm = TRUE)),
    sd = format_number(sd(wages, na.rm = TRUE)),
    variance = format_number(var(wages, na.rm = TRUE)),
    quartiles = paste(format_number(quantile(wages, probs = c(0.25, 0.5, 0.75), na.rm = TRUE)),
collapse = ", "),
    deciles = paste(format_number(quantile(wages, probs = c(0.9, 1), na.rm = TRUE)), collapse = ", ")
  )
})

# Convert results into a matrix format
results_matrix <- do.call(cbind, lapply(results_list, function(x) {
  c(x$mean, x$mode, x$median, x$range, x$sd, x$variance, x$quartiles, x$deciles)
}))

# Assign row names and column names
rownames(results_matrix) <- c("mean", "mode", "median", "range", "sd", "variance", "quartiles",
"deciles")
colnames(results_matrix) <- fleets

# Convert the matrix to data frame for better printing
results_df <- as.data.frame(results_matrix)

# Print the results in the desired format
print(results_df, row.names = TRUE)
```

The output of the given code gives the following output for the data of 4 catagories.

Table III: Central tendency for the data

	Small Narrowbodies	Large Narrowbodies	Widebodies	Total Fleet
Mean	442.020	538.084	913.591	595.351
Mode	388.037	495.274	758.697	510.407
Median	437.950	542.676	964.563	611.027
Range	241.462	261.546	478.201	327.958
Standard deviation	76.219	81.071	155.987	102.360
Variance	5809.266	6572.426	24331.900	10477.637
Quartiles	393.517, 437.950, 473.707	500.650, 542.676, 584.232	783.633, 964.563, 993.432	521.051, 611.027, 643.327
Deciles	554.266, 568.444	631.807, 654.404	1086.465, 1131.132	723.612, 757.800

Question 03: Histograms for "Grouped Salaries"

There can be four histograms for the frequency distribution tables from the question 1.

Table IV: Code for Q2

```
library(readxl)
library(fdth)

# Load data
data <- read_excel("B:/BSMRAAU 22024010 5th/MAT_4509/Programming/United Airlines Aircraft Operating Statistics- Cost Per Block Hour (Unadjusted).xls", range="b2:w158")
data <- as.data.frame(data)

fleets <- c("Small Narrowbodies", "Large Narrowbodies", "Widebodies", "Total Fleet")
wages_rows <- c(8, 47, 86, 125) - 2

# Function to get the first 10 non-NA values from a specified row
get_row <- function(row, data) {
  na.omit(as.numeric(data[row, -1]))[1:10]
}

# Function to create a frequency distribution table and save the plot
create_and_save_table <- function(data, fleet_name) {
  interval <- as.integer((max(data) - min(data)) / (log2(length(data)) + 1))
  table <- fdt(data, start = min(data), end = max(data), h = interval)

  # Save plot as PNG
  png_filename <- paste0("B:/BSMRAAU 22024010 5th/MAT_4509/Programming/Charts/", fleet_name, "_wages_distribution.png")
  png(filename = png_filename, width = 800, height = 600)
```

```
  plot(table, main = paste("Wages Distribution for", fleet_name))
  dev.off()  # Close the PNG device
}

# Main code to create tables and save plots
for (i in seq_along(wages_rows)) {
  pilots_wages <- get_row(wages_rows[i], data)
  create_and_save_table(pilots_wages, fleets[i])
}
```

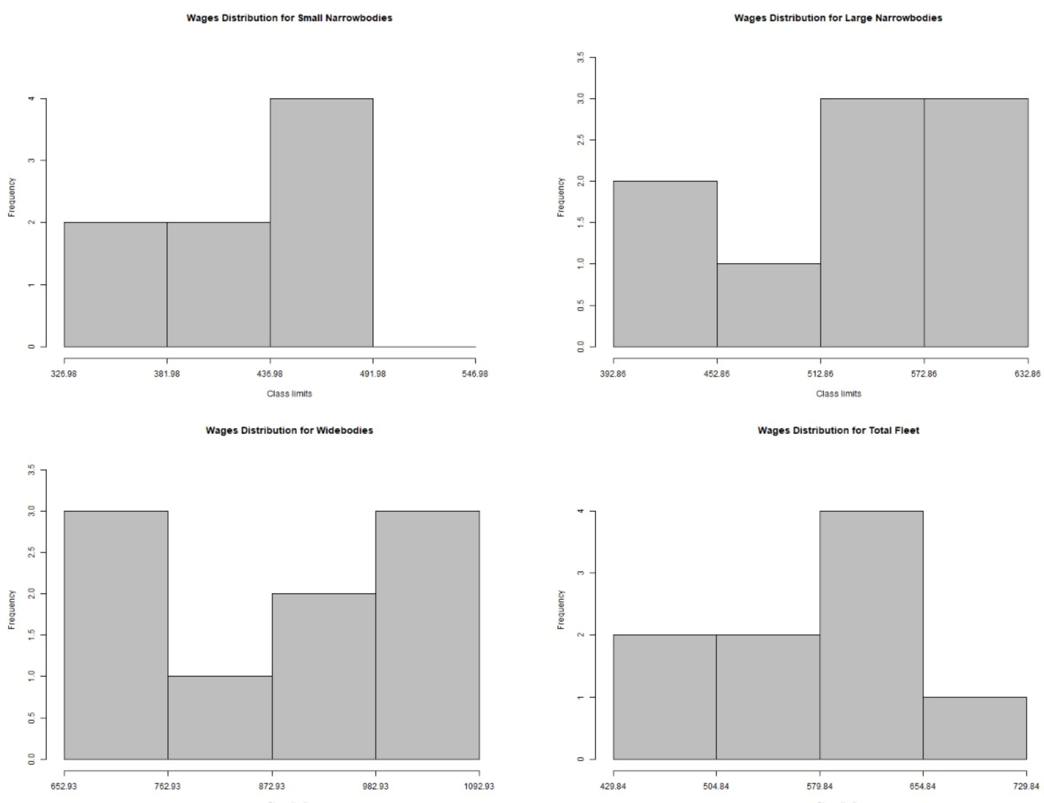

Figure 1: Frequency distribution histograms

Question 04: Pie Chart and Bar Diagram for “Maintenance” and “Load factor”

The following code plots pie charts for the “Maintenance” variables and bar plots for “Load factors”.

Table V: Code for Q2

```
library(readxl)

######################################

########## Data Loading ############

######################################

data <- read_excel("B:/BSMRAAU 22024010 5th/MAT_4509/Programming/United Airlines Aircraft Operating Statistics- Cost Per Block Hour (Unadjusted).xls", range="b2:w158")

data <- as.data.frame(data)

fleets <- c("Small Narrowbodies", "Large Narrowbodies", "Widebodies", "Total Fleet")

years_row <- 3 - 2

load_factor_rows <- c(36, 75, 114, 153) - 2

maintanance_rows <- matrix(c(19,20,21,23,58,59,60,62,97,98,99,101,136,137,138,140) - 2, nrow = 4, byrow = FALSE)

######################################

########## Functions ###############

######################################


get_row <- function(row, data) {

  na.omit(as.numeric(data[row, -1]))[1:10]

}


get_row_sum <- function(row, data) {

  sum(na.omit(as.numeric(data[row, -1]))[1:10])

}


get_totals <- function(row_matrix, data) {

  sapply(1:4, function(i) sapply(1:4, function(j) get_row_sum(row_matrix[i, j], data)))
```

```
}


plot_and_save_pies <- function(labour, materials, third_party, burden, labels, save_path) {
 for (i in 1:4) {
  values <- c(labour[i], materials[i], third_party[i], burden[i])
  percentages <- round(values / sum(values) * 100)
  pie_labels <- paste(c("Labour", "Materials", "Third party", "Burden"), percentages, "%")
  file_name <- paste0(save_path, "/", labels[i], "_maintenance_pie.png")
  # Open a PNG device and save the plot
  png(filename = file_name, width = 600, height = 600)
  pie(values, labels = pie_labels, main = labels[i], col = rainbow(length(values)))
  dev.off()  # Close the device to save the file
 }
}


create_barplot <- function(years_row, load_factor_rows, fleets, save_path) {
 years <- get_row(years_row, data)
 for (i in seq_along(load_factor_rows)) {
  load_factors <- get_row(load_factor_rows[i], data)
  fleet_name <- fleets[i]
  file_name <- paste0(save_path, "/barplot_", fleet_name, ".png")
  png(file_name, width = 800, height = 600)
  barplot(load_factors, names.arg = years, col = "skyblue",
      main = paste("Load Factors for", fleet_name),
      xlab = "Years", ylab = "Load Factors")
  dev.off()
 }
 message("Bar plots saved successfully in ", save_path)
```

```
}


# Define save path

save_path <- "B:/BSMRAAU 22024010 5th/MAT_4509/Programming/Charts"

# Create directory if it doesn't exist

if (!dir.exists(save_path)) {

  dir.create(save_path, recursive = TRUE)

}

# Calculate totals for each category

totals <- get_totals(maintanance_rows, data)

total_labour <- totals[1, ]

total_materials <- totals[2, ]

total_third_party <- totals[3, ]

total_burden <- totals[4, ]

# Generate and save pie charts for each fleet

plot_and_save_pies(total_labour, total_materials, total_third_party, total_burden, fleets,
save_path)


save_path <- "B:/BSMRAAU 22024010 5th/MAT_4509/Programming/Charts"


if (!dir.exists(save_path)) {

  dir.create(save_path, recursive = TRUE)

}


create_barplot(years_row, load_factor_rows, fleets, save_path)
```

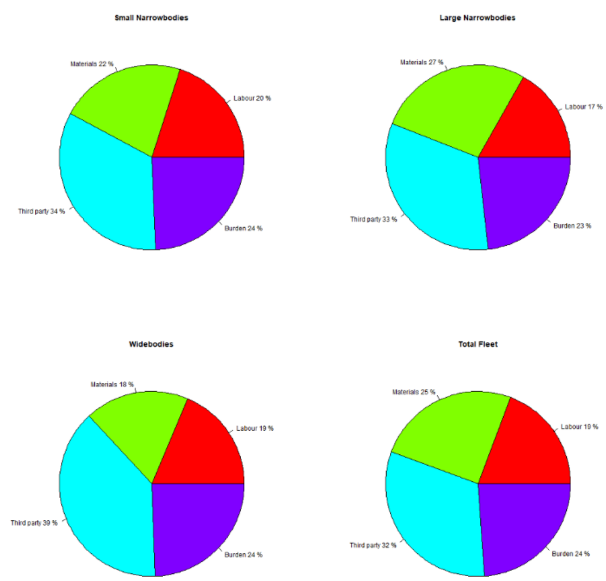

Figure 2: Pie Charts for the "Maintenance" variable

Here the “Maintenance” variable had the following "Labour", "Materials", "Third party", "Burden" sections. The pie chart was plotted by talking the average of 10 years for each type of fleets. We get 4 pie charts out of this.

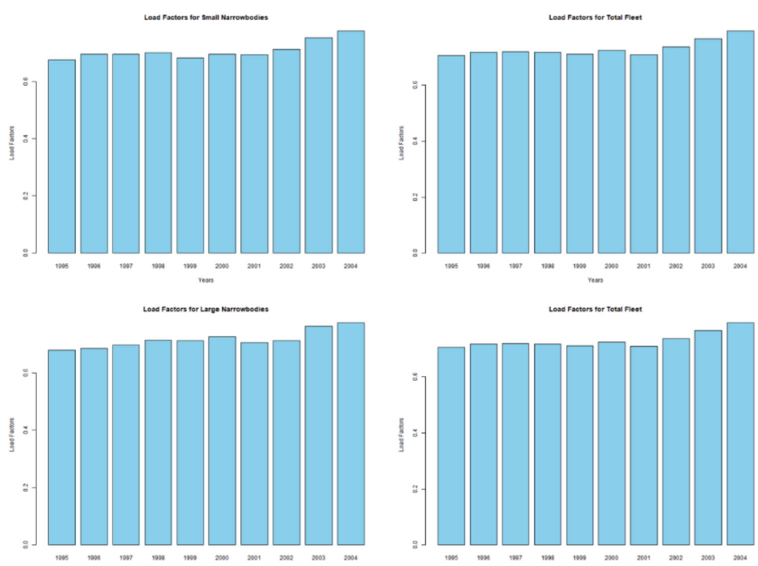

Figure 3: Barplot for "Load factor" variable

At the 2nd part of the code the bar plot for the “Load factor” variable was plotted, for each type of fleets the load factor was plotted for different years, in this case for 10 years

Question 05: Box Plot for “Purchased Goods”, “Aircraft Ownerships” and “Daily Utilization per Aircraft”

The following code generates box plot for the following variables,

- Purchased Goods:
    - Fuel/Oil
    - Insurance
    - Other (inc. Tax)
- Aircraft Ownership:
    - Rentals
    - Depreciation and Amortization
- Daily Utilization per Aircraft:
    - Block Hours
    - Airborne Hours
    - Departures

Each of these sections contains multiple variables. The box plots were plotted for those variables for each section and each type of aircrafts.

Purchased goods:

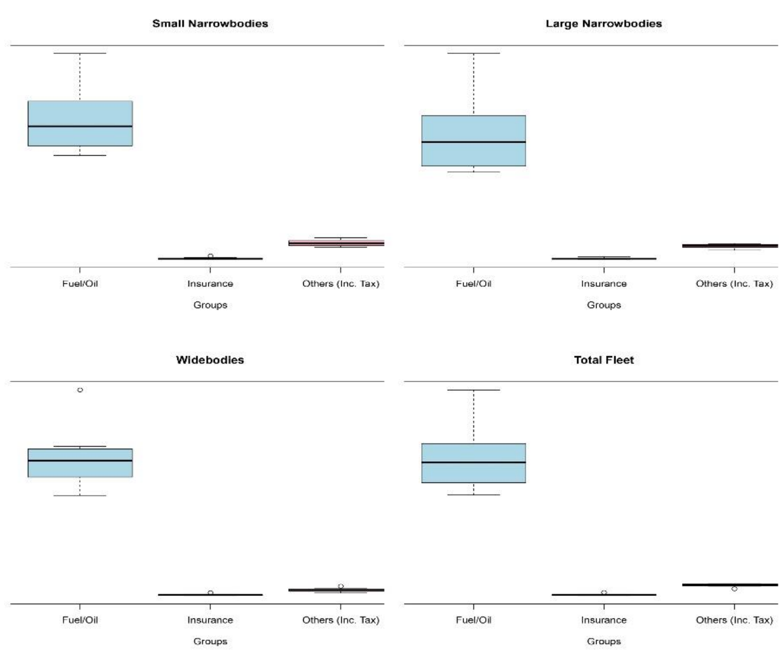

Figure 4: Box plots for "Purchased goods"

Aircraft Ownership:

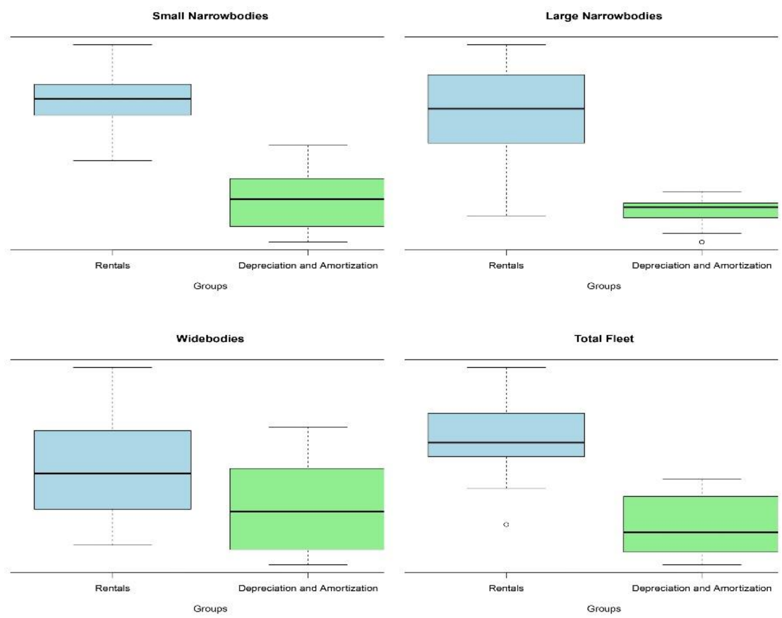

Figure 5: Box plots for "Aircraft ownership

Daily Utilization per Hour:

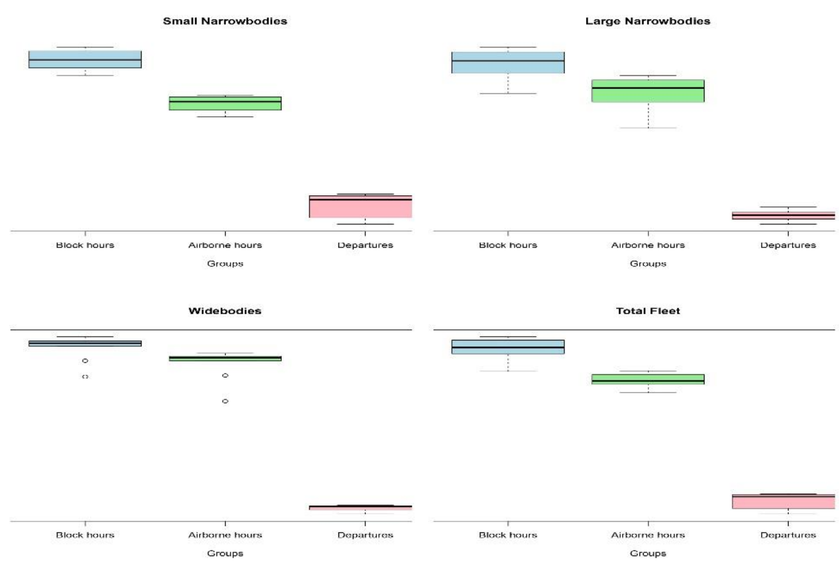

Figure 6: Box plots for "Daily Utilization per Aircraft"

Question 06: Plot summary

United Airlines' data provides an in-depth comparison of several fleet types on a variety of operational and financial metrics.

From the distribution of salaries, one can find out that Small Narrowbody is bimodal, though the most subjects are in the middle class, while its highest class does not get any representatives; Large Narrowbody is more or less uniformly distributed, with higher classes receiving more representatives. Widebodies are highly skewed, with most of the salaries at the extreme ends and very few in the middle. The Total Fleet follows a very consistent distribution, peaking in the middle to upper range of salaries. Looking at the shares of maintenance costs, Small and Large Narrowbodies are well-balanced, as the largest share is for third-party services, while labor and burden costs are similar in size. Widebodies have higher labor costs but are still very reliant on third-party services. The total fleet gives a good overview of its maintenance cost structure, which reflects United's overall maintenance strategy while adjusting for specific aircraft needs.

Analyzing the load factors-1995-2004, Small Narrowbodies reflect an upward trend, showing higher values towards the end, while in the beginning, in 1995, it had relatively lower values. Large Narrowbodies reflect a gradual increase that peaks higher than Small Narrowbodies, with substantial improvement by 2004. Widebodies are very stable and seem to maintain high load factors across the decade, while the Total Fleet is one of steady improvement, whereby there is a clear increasing trend upwards from 1995 to 2004.

Fuel/oil costs are the highest and most variable for the fleet types; Widebodies have the largest variability. This is reflected in Large Narrowbodies showing moderate variability, while the distributions of Small Narrowbody are more compact. Insurance and Other Costs - including tax - are considerably lower, with more compact distributions and fewer outliers.

Moving to Rentals vs. Depreciation & Amortization, median rental costs are higher than depreciation for Small Narrowbodies, with moderate variability. For Large Narrowbodies, similar patterns are seen, but rental costs have a greater spread, while there is a noticeable outlier in depreciation. Widebodies are the most diverse in both rental and depreciation costs, generally rental-heavy. The Total Fleet median rental cost is higher; the spread is wider than for depreciation, with one outlier in rentals.

Utilization metrics here are Block Hours, Airborne Hours, and Departures. In this case, Small Narrowbodies show the greatest number of departures, while block and airborne hours are more middling, which would indicate consistent use. Large Narrowbodies indicate higher block hours and airborne hours but fewer departures, showing a pattern that is somewhat balanced. Widebodies display the highest block and airborne hours but with the lowest number of departures; thus, this reflects the most variable utilization pattern. The Total Fleet reflects a more balanced pattern with reasonable block and airborne hours and rather balanced departure numbers.

These findings point out that smaller planes have a higher frequency of flights but shorter hours of flying, while larger ones, such as Widebodies, can boast of more flight hours and fewer departures. This reflects operational strategies appropriate for each aircraft type's intended use and route structure.