Bangabandhu Sheikh Mujibar
Rahman
Aviation and Aerospace University
(BSMRAAU)

ASSIGNMENT

Name of Assignment: Statistical Analysis of United Airlines Aircraft Operating Statistics

Date of Submission: 26/11/2024

Name: Md. Mahdi Kamal Alif

ID: 22024019

Department: **Aeronautical Engineering (AVIONICS)**

Course Name: Probability and Statistics

Course Code: MAT 4509

Submitted To: Assoc Prof Dr. M Siddikur Rahman

Table of Contents

Introduction:

The following report does an in-depth statistical analysis of fleet performance related to salaries, maintenance costs, and aircraft utilization using frequency distributions and graphical representations of the operational data for the period ranging from 1995 to 2004 across fleet categories such as Small Narrowbodies, Large Narrowbodies, Widebodies, and Total Fleet. This will be helpful in extracting information on operational strategy concerning United Airlines that will help optimize resource utilization through better decision-making.

Question 01: Frequency Distribution of "Salaries and Wages"

Table I.I: Small Narrowbodies

Class Intervals	Frequency
291.555,343.555	5
343.555,395.555	5
395.555,447.555	5
447.555,499.555	1
499.555,551.555	1

Table I.II: Large Narrowbodies

Class Intervals	Frequency
349.52,407.52	6
465.52,523.52	4
512.86-572.86	2
523.52,581.52	3
581.52,639.52	3

Table I.III: Widebodies

Class Intervals	Frequency
581.0153,685.0153	8
685.0153,789.0153	2
789.0153,893.0153	2
893.0153,997.0153	5
997.0153,1101.015	1

Table I.IV: Total Fleet

Class Intervals	Frequency
385.51,455.51	5
455.51,525.51	6
525.51,595.51	1
595.51,665.51	5

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("D:/vscode/R/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 19 non-NA values from a specified row
get_row <- function(row, data) {
  na.omit(as.numeric(data[row, -1]))[1:19]
}

# 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	402.401	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	340.491, 388.037, 441.287	397.111, 439.566, 543.969	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 Q3

```
library(readxl)
library(fdth)

# Load data
data <- read_excel("D:/vscode/R/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 19 non-NA values from a specified row
get_row <- function(row, data) {
  na.omit(as.numeric(data[row, -1]))[1:19]
}

# 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("D:/vscode/R/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])
}
```

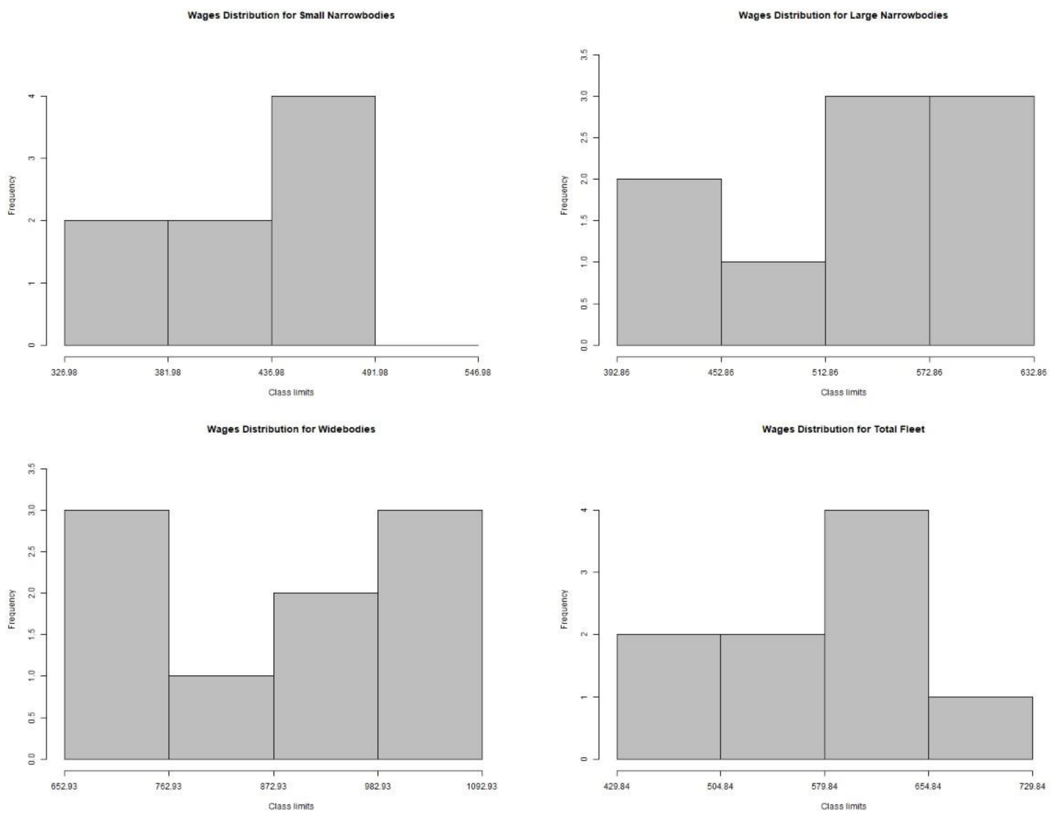

Figure 1: Frequency distribution histograms

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 Q4

```
library(readxl)

#load Data
data <- read_excel("D:/vscode/R/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)


get_row <- function(row, data) {
  na.omit(as.numeric(data[row, -1]))[1:19]
}

get_row_sum <- function(row, data) {
  sum(na.omit(as.numeric(data[row, -1]))[1:19])
}

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 <- "D:/vscode/R/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)


if (!dir.exists(save_path)) {
  dir.create(save_path, recursive = TRUE)
}

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

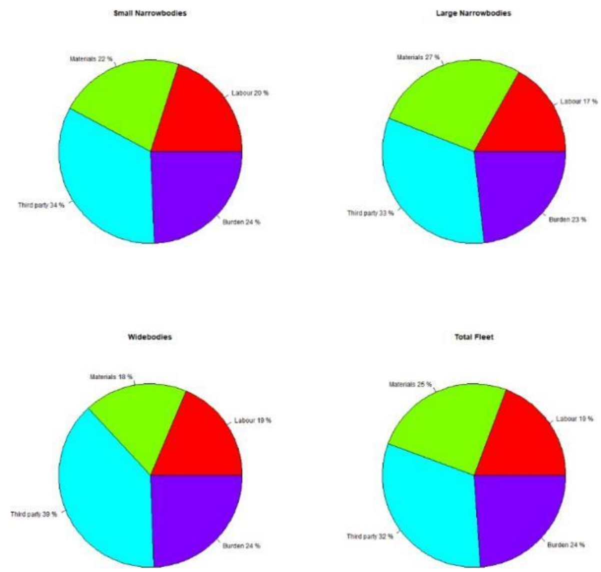

Figure 2: Pie Charts for the "Maintenance" variable

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.

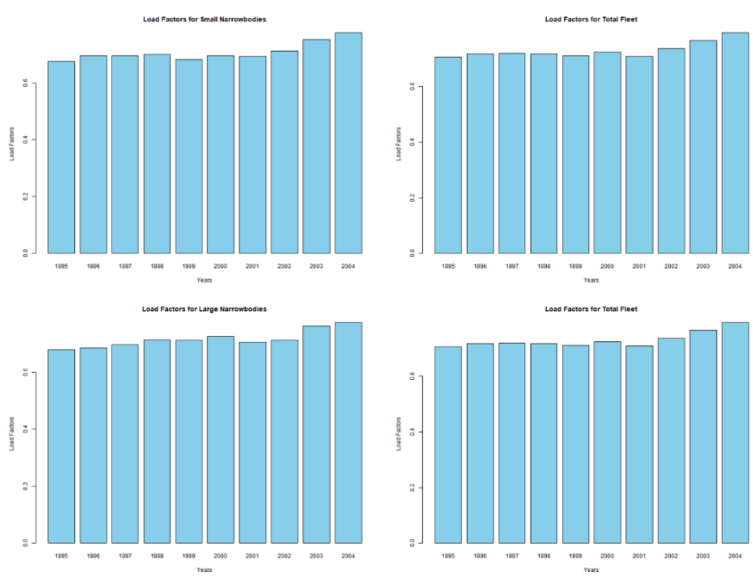

Figure 3: Barplot for "Load factor" variable

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:

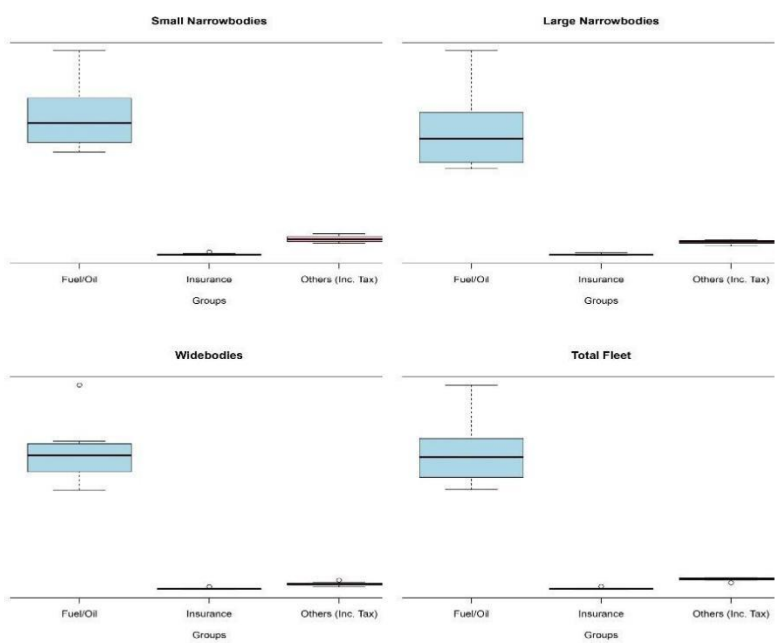

Figure 4: Box plots for "Purchased goods"

Aircraft Ownership:

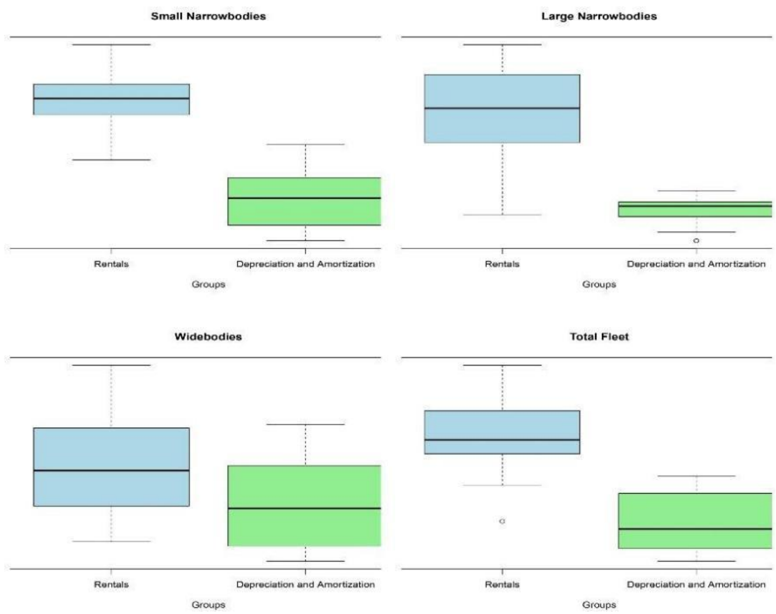

Figure 5: Box plots for "Aircraft ownership

Daily Utilization per Hour:

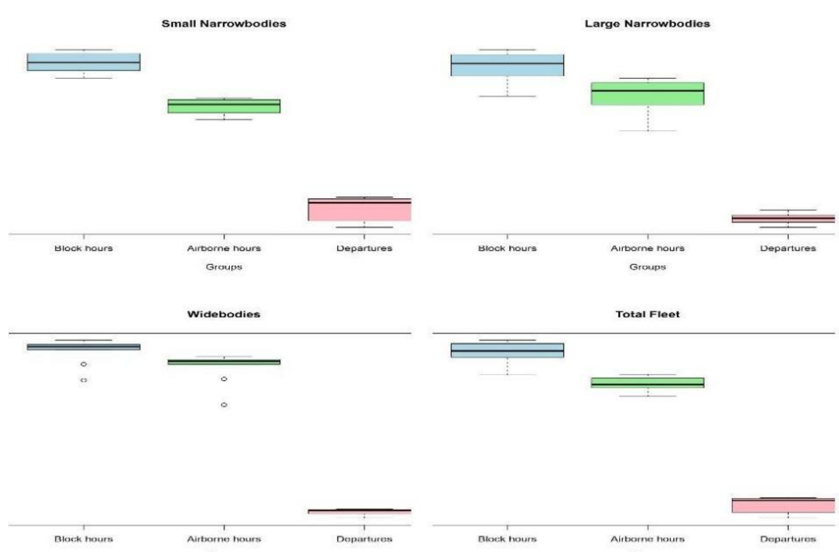

Figure 6: Box plots for "Daily Utilization per Aircraft"

Question 06: Plot summary

United Airlines' data offers a detailed comparison of various fleet types across operational and financial metrics. Analyzing salary distributions reveals that Small Narrowbodies exhibit a bimodal pattern, with most entries concentrated in the middle range, while the highest class remains unrepresented. Large Narrowbodies display a relatively uniform distribution, with higher classes having greater representation. Widebodies, however, show a highly skewed distribution, where salaries cluster at the extremes with minimal presence in the middle. The Total Fleet presents a steady distribution, peaking in the mid-to-upper salary range. Regarding maintenance costs, Small and Large Narrowbodies are evenly balanced, with the largest share allocated to third-party services and similar proportions for labor and burden costs. Widebodies incur higher labor expenses while maintaining reliance on third-party services. Overall, the Total Fleet reflects United Airlines' strategic approach to maintenance, tailored to specific aircraft requirements.

Examining load factors between 1995 and 2004, Small Narrowbodies show a positive trend, starting lower in 1995 and improving significantly by 2004. Large Narrowbodies experience a gradual increase, peaking higher than Small Narrowbodies by the end of the period. Widebodies maintain consistently high load factors throughout the decade, while the Total Fleet exhibits a clear upward trend, indicating steady improvement over time. Fuel and oil costs emerge as the most significant and variable expense among the fleet types, with Widebodies displaying the greatest variability. Large Narrowbodies show moderate variability, while Small Narrowbodies have more compact cost distributions. Insurance and other costs, including taxes, are lower and feature compact distributions with fewer outliers.

In comparing Rentals to Depreciation and Amortization, Small Narrowbodies have higher median rental costs with moderate variability. Large Narrowbodies follow a similar trend, though rental costs exhibit a wider spread and an outlier in depreciation. Widebodies demonstrate the greatest diversity in both rental and depreciation costs, favoring rentals. The Total Fleet shows a higher median rental cost, with a broader spread compared to depreciation, including a single rental outlier.

Utilization metrics—Block Hours, Airborne Hours, and Departures—highlight varying patterns across fleet types. Small Narrowbodies lead in departures, with average block and airborne hours indicating steady use. Large Narrowbodies have higher block and airborne hours but fewer departures, reflecting a balanced usage pattern. Widebodies, on the other hand, achieve the highest block and airborne hours but the fewest departures, indicating the most variable usage pattern. The Total Fleet strikes a balance, with reasonable block and airborne hours alongside evenly distributed departure numbers.

In summary, smaller aircraft tend to operate with higher flight frequencies but shorter durations, while larger aircraft like Widebodies operate fewer flights with extended durations. These insights align with operational strategies tailored to the specific roles and route structures of each aircraft type.