1.  
# Load a built-in dataset, for example, mtcars

data(mtcars)

# Compute correlation between 'mpg' (miles per gallon) and 'wt' (weight)

correlation <- cor(mtcars$mpg, mtcars$wt)

print(paste("Correlation between mpg and wt:", round(correlation, 3)))

# Scatter plot to visualize the relationship

plot(mtcars$wt, mtcars$mpg,

main = "Scatter plot of MPG vs Weight",

xlab = "Weight (1000 lbs)",

ylab = "Miles Per Gallon (MPG)",

pch = 19, col = "blue")

# Add a regression line

abline(lm(mpg ~ wt, data = mtcars), col = "red", lwd = 2)  
  
  
2.  
# Load dataset

data(mtcars)

# Pearson correlation test

cor\_test <- cor.test(mtcars$mpg, mtcars$wt, method = "pearson")

print(cor\_test)

# Q-Q plots to check normality of the two variables

par(mfrow = c(1, 2)) # Plot side by side

qqnorm(mtcars$mpg, main = "Q-Q Plot of MPG")

qqline(mtcars$mpg, col = "red")

qqnorm(mtcars$wt, main = "Q-Q Plot of Weight")

qqline(mtcars$wt, col = "red")

# Reset plotting area

par(mfrow = c(1, 1))  
  
  
3.  
# Sample data (replace this with your actual data)

price\_quotes <- data.frame(

Mary = c(50, 60, 55, 65, 70),

Barry = c(55, 58, 57, 62, 68)

)

# Summary statistics

summary(price\_quotes)

# Additional descriptive stats for clarity

library(dplyr)

price\_quotes %>%

summarise(

Mary\_Mean = mean(Mary),

Mary\_SD = sd(Mary),

Mary\_SE = sd(Mary) / sqrt(n()),

Barry\_Mean = mean(Barry),

Barry\_SD = sd(Barry),

Barry\_SE = sd(Barry) / sqrt(n())

)  
  
  
4.  
# Sample data (replace this with your actual data)

treatment\_data <- data.frame(

Facility = rep(c("Facility1", "Facility2"), each = 4),

Reengineering = rep(c("Before", "After"), times = 4),

Behavioral\_Problems = c(15, 10, 18, 12, 20, 14, 22, 16),

Staff\_Turnover = c(8, 6, 9, 7, 10, 8, 11, 9)

)

# 1. Summary statistics of Treatment Facility (Behavioral Problems & Staff Turnover)

library(dplyr)

summary\_stats <- treatment\_data %>%

group\_by(Facility) %>%

summarise(

BP\_Mean = mean(Behavioral\_Problems),

BP\_SD = sd(Behavioral\_Problems),

ST\_Mean = mean(Staff\_Turnover),

ST\_SD = sd(Staff\_Turnover)

)

print(summary\_stats)

# 2. Effect of Reengineering on Behavioral Problems and Staff Turnover

# Compare means Before and After reengineering

effect\_stats <- treatment\_data %>%

group\_by(Reengineering) %>%

summarise(

BP\_Mean = mean(Behavioral\_Problems),

BP\_SD = sd(Behavioral\_Problems),

ST\_Mean = mean(Staff\_Turnover),

ST\_SD = sd(Staff\_Turnover)

)

print(effect\_stats)

# Optional: Paired t-tests (if data is paired)

# Assuming paired data by facility, test if reengineering changed Behavioral Problems

bp\_before <- treatment\_data$Behavioral\_Problems[treatment\_data$Reengineering == "Before"]

bp\_after <- treatment\_data$Behavioral\_Problems[treatment\_data$Reengineering == "After"]

bp\_ttest <- t.test(bp\_before, bp\_after, paired = TRUE)

print(bp\_ttest)

# Similarly for Staff Turnover

st\_before <- treatment\_data$Staff\_Turnover[treatment\_data$Reengineering == "Before"]

st\_after <- treatment\_data$Staff\_Turnover[treatment\_data$Reengineering == "After"]

st\_ttest <- t.test(st\_before, st\_after, paired = TRUE)

print(st\_ttest)  
  
  
5.  
# Sample data (replace with your real data)

baggage\_data <- data.frame(

Airline = rep(c("American Eagle", "Hawaiian", "United"), each = 10),

Complaints = c(5,4,3,6,7,3,4,5,6,5, 3,2,1,4,3,2,3,4,5,3, 7,6,5,8,9,7,8,9,10,8),

Month = rep(1:10, 3),

Destination = rep(c("NY", "LA", "SF", "NY", "LA", "SF", "NY", "LA", "SF", "NY"), 3),

Season = rep(c("Winter", "Spring", "Summer", "Fall", "Winter"), 6),

Travel\_Volume = sample(1000:5000, 30, replace = TRUE)

)

library(dplyr)

library(ggplot2)

# 1. Summary statistics

summary\_stats <- baggage\_data %>%

group\_by(Airline) %>%

summarise(

Mean\_Complaints = mean(Complaints),

SD\_Complaints = sd(Complaints),

Min\_Complaints = min(Complaints),

Max\_Complaints = max(Complaints)

)

print(summary\_stats)

# 2. Comparing airlines - Which airline has best/worst record?

# Lower mean complaints → better record

# Plot complaints over time by airline

ggplot(baggage\_data, aes(x = Month, y = Complaints, color = Airline)) +

geom\_line() +

geom\_point() +

labs(title = "Baggage Complaints Over Time by Airline",

x = "Month",

y = "Number of Complaints") +

theme\_minimal()

# Check trends over time with linear regression by airline

library(broom)

baggage\_trends <- baggage\_data %>%

group\_by(Airline) %>%

do(tidy(lm(Complaints ~ Month, data = .)))

print(baggage\_trends)

# Analyze other factors using a linear model

lm\_model <- lm(Complaints ~ Airline + Season + Destination + Travel\_Volume, data = baggage\_data)

summary(lm\_model)  
  
  
6.  
# Sample data (replace with your real data)

malpractice\_data <- data.frame(

ClaimPayment = c(50000, 120000, 70000, 90000, 60000, 110000, 130000, 80000, 75000, 115000),

Specialty = c("Anesthesiologist", "Dermatologist", "Orthopedic Surgeon", "Anesthesiologist",

"Orthopedic Surgeon", "Dermatologist", "Dermatologist", "Anesthesiologist",

"Orthopedic Surgeon", "Dermatologist"),

PatientAge = c(45, 50, 60, 35, 70, 40, 55, 65, 50, 60)

)

library(dplyr)

library(ggplot2)

# 1. Descriptive statistics for Claim Payment

summary(malpractice\_data$ClaimPayment)

# Boxplot to visualize Claim Payment distribution by Specialty

ggplot(malpractice\_data, aes(x = Specialty, y = ClaimPayment, fill = Specialty)) +

geom\_boxplot() +

labs(title = "Claim Payment Amount by Specialty", y = "Claim Payment ($)") +

theme\_minimal()

# Identify factors influencing Claim Payment (basic linear model)

lm\_payment <- lm(ClaimPayment ~ Specialty + PatientAge, data = malpractice\_data)

summary(lm\_payment)

# 2. Percentage of cases by specialty

percent\_specialty <- malpractice\_data %>%

group\_by(Specialty) %>%

summarise(Count = n()) %>%

mutate(Percentage = 100 \* Count / sum(Count))

print(percent\_specialty)

# Relationship between Patient Age and Claim Payment

correlation <- cor.test(malpractice\_data$PatientAge, malpractice\_data$ClaimPayment)

print(correlation)

# Scatter plot with regression line

ggplot(malpractice\_data, aes(x = PatientAge, y = ClaimPayment)) +

geom\_point(color = "blue") +

geom\_smooth(method = "lm", se = TRUE, color = "red") +

labs(title = "Relationship between Patient Age and Claim Payment",

x = "Patient Age",

y = "Claim Payment ($)") +

theme\_minimal()