# Exercise1 Data Analysis

Angelos Papageorgiou

2025-03-22

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

#Question1

data <- read.csv("C:/Users/aggel/Desktop/NEA E A 1 (2025)/ObesityDataSet.csv") summary(data)

#Check the dimensions of the dataset dim(data)

#Check the types of variables str(data)

#Convert chr variables to factors $data$Gender $= as.factor(data$Gender) $data$CALC $= as.factor(data$CALC) $data$FAVC = as.factor(data$FAVC) $data$SCC = as.factor(data$SCC) $data$SMOKE = as.factor(data$SMOKE) $data$family_history_with_overweight = as.factor(data$family\_history\_with\_overweight) $data$CAEC = as.factor(data$CAEC) $data$MTRANS = as.factor(data$MTRANS) $data$NObeyesdad <- as.factor(data$NObeyesdad)

#Check the types after conversion str(data)

#Question2

#Create Age_f $data$Age_f = ifelse(data$Age <18, 1, 2) $data$Age_f = factor(data$Age\_f, levels = c(1,2), labels = c("Minor", "Adult"))

#Calculate quartiles for the Weight variable Q1 = quantile(data$Weight, 0.25) $Q2 = quantile(data$Weight, 0.5) Q3 = quantile(data$Weight, 0.75)

#Create Weight_f $data$Weight_f = cut(data$Weight, breaks = c(-Inf,Q1,Q2,Q3,Inf), labels = c("Flyweight", "Lightweight", "Middleweight", "Heavyweight"))

#Question3 #Create dataframe with quantitative variables numeric_data = data[, c("Age", "Height", "Weight", "FCVC", "NCP", "CH2O", "FAF", "TUE")]

#Install sjmisc library install.packages("sjmisc")

#Load sjmisc library library(sjmisc)

#Calculate descriptive measures descr(numeric_data)

#Question4

#Install ggplot2 install.packages("ggplot2") #Load ggplot2 library library(ggplot2)

#Create Histogram for Age ggplot(data, aes(x = Age)) + geom_histogram(aes(y = ..density..), binwidth = 2, fill = "skyblue", color = "black", alpha = 0.7) + geom_density(color = "red", linewidth = 1) + labs(title = "Age Histogram",x = "Age",y = "Probability Density Function") +

theme_minimal()

#Boxplot for Age ggplot(data, aes(y = Age)) + geom_boxplot(fill = "lightgreen", color = "black") + labs(title = "Age Boxplot",y = "Age") +

theme_minimal()

#QQ plot for Age ggplot(data, aes(sample = Age)) + geom_qq(color = "blue") + geom_qq_line(color = "red") +labs(title = "QQ Plot for Age",x = "Theoretical Quantiles",y = "Sample Quantiles") +

theme_minimal()

#Question5 #Install and load libraries install.packages("sjmisc") install.packages("sjPlot") library(sjmisc) library(sjPlot) library(ggplot2)

#Frequency table and relative frequencies frq(data$MTRANS)

#Bar plot for MTRANS ggplot(data, aes(x = MTRANS)) + geom_bar(fill = "skyblue", color = "black") + labs(title = "Bar Plot for MTRANS Variable", x = "Transportation Method", y = "Frequency") + theme_minimal()

#Calculate frequencies freq_table <- as.data.frame(table(data$MTRANS)) colnames(freq_table) <- c("MTRANS", "Frequency")

#Pie chart ggplot(freq_table, aes(x = " ", y = Frequency, fill = MTRANS)) + geom_bar(stat ="identity", width = 1) + coord_polar("y", start = 0) + geom_text(aes(label = paste0(Frequency)), position = position_stack(vjust = 0.5)) + labs(title = "Pie Chart for MTRANS Variable", fill = "Transportation Method") + theme_void()

#Question6 #Load ggplot2 library library(ggplot2)

#Scatterplot for Height and Weight ggplot(data, aes(x = Height, y = Weight)) + geom_point(color = "blue", size = 2) + labs(title = "Scatterplot: Relationship between Height and Weight", x = "Height (m)", y = "Weight (kg)") + theme_minimal()

#Scatterplot with Age_f ggplot(data, aes(x = Height, y = Weight, color = Age_f, shape = Age_f)) + geom_point(size = 3) + labs(title = "Scatterplot: Relationship between Height and Weight based on Age", x = "Height (m)", y = "Weight (kg)", color = "Age Category", shape = "Age Category") + theme_minimal() + scale_color_manual(values = c("Minor" = "red", "Adult" = "blue")) + scale_shape_manual(values = c("Minor" = 16, "Adult" = 17))

#Question7 #Load ggplot2 library library(ggplot2)

#Boxplot for Weight by Gender ggplot(data, aes(x = Gender, y = Weight, fill = Gender)) + geom_boxplot() + labs(title = "Boxplot of Weight by Gender", x = "Gender", y = "Weight (kg)") + theme_minimal() + scale_fill_manual(values = c("Female" = "pink", "Male" = "lightblue"))

#Question8 #Install and load libraries install.packages("gmodels") install.packages("ggplot2") library(gmodels) library(ggplot2)

#Contingency table for Age_f and Weight_f CrossTable(data$Age_f, data$Weight_f, prop.chisq = FALSE)

#Grouped bar plot ggplot(data, aes(x = Weight_f, fill = Age_f)) + geom_bar(position = "dodge") + labs(title = "Grouped Bar Plot: Age_f and Weight_f", x = "Weight Category", y = "Frequency", fill = "Age Category") + theme_minimal()

#Stacked bar plot ggplot(data, aes(x = Weight_f, fill = Age_f)) + geom_bar(position = "stack") + labs(title = "Stacked Bar Plot: Age_f and Weight_f", x = "Weight Category", y = "Frequency", fill = "Age Category") + theme_minimal()