# MODULE SECTION: ESTIMATION

## Demonstration: how to calculate standard error and confidence intervals in R

R code from Chatgpt

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

setwd("C:/Users/… ")

# Load necessary libraries

library(readxl) # To read Excel files

library(dplyr) # For data manipulation

library(tidyverse) # For summarizing and managing data

# Read the data from Excel

farm\_data <- read\_excel("farm\_data.xlsx", sheet = "Sheet1")

#file\_path <- "farm\_data.xlsx" # Replace with your actual file path

#data <- read\_excel(file\_path)

# 1. Calculate the overall mean, standard error, and confidence interval for FarmIncome

# Helper function to compute the standard error

standard\_error <- function(x) {

sd(x) / sqrt(length(x))

}

# Overall statistics

overall\_mean <- mean(farm\_data$FarmIncome)

overall\_se <- standard\_error(farm\_data$FarmIncome)

# Calculate the 95% confidence interval

alpha <- 0.05

z\_value <- qnorm(1 - alpha / 2) # Z-score for 95% confidence

overall\_ci <- c(

overall\_mean - z\_value \* overall\_se,

overall\_mean + z\_value \* overall\_se

)

# Print the overall results

cat("Overall Mean:", overall\_mean, "\n")

cat("Standard Error:", overall\_se, "\n")

cat("95% Confidence Interval:", overall\_ci, "\n")

# 2. Calculate mean, standard error, and confidence interval for each FarmSize group

group\_stats <- farm\_data %>%

group\_by(FarmSize) %>%

summarise(

Mean = mean(FarmIncome),

SE = standard\_error(FarmIncome),

CI\_Lower = Mean - z\_value \* SE,

CI\_Upper = Mean + z\_value \* SE

)

# Print group results

cat("\nStatistics by FarmSize:\n")

print(group\_stats)

# Load necessary libraries

library(readxl) # To read Excel files

library(ggplot2) # For plotting

# Create the plot

ggplot() +

# Violin plot for each farm size

geom\_violin(

data = farm\_data,

aes(x = FarmSize, y = FarmIncome),

fill = "skyblue",

alpha = 0.7,

color = "darkblue"

) +

# Add boxplot for each farm size inside violin plots

geom\_boxplot(

data = farm\_data,

aes(x = FarmSize, y = FarmIncome),

width = 0.2,

color = "black",

outlier.shape = 19

) +

# Violin plot for overall data

geom\_violin(

data = farm\_data,

aes(x = "Overall", y = FarmIncome),

fill = "lightgray",

alpha = 0.7,

color = "black"

) +

# Add boxplot for overall data inside the violin plot

geom\_boxplot(

data = farm\_data,

aes(x = "Overall", y = FarmIncome),

width = 0.2,

color = "black",

outlier.shape = 19

) +

# Customize the labels and theme

labs(

title = "Farm Income Distribution by Farm Size and Overall",

x = "Farm Size",

y = "Farm Income"

) +

scale\_x\_discrete(limits = c("small", "large", "Overall")) + # Ensure "Overall" appears

theme\_minimal(base\_size = 14) +

theme(

panel.background = element\_rect(fill = "white", color = NA), # Plot area background white

plot.background = element\_rect(fill = "white", color = NA), # Surrounding area white

axis.title = element\_text(color = "black"), # Axis labels in black

axis.text = element\_text(color = "black"), # Axis tick labels in black

plot.title = element\_text(color = "black", hjust = 0.5) # Title in black and centered

)

# Save the plot (optional)

ggsave("farm\_income\_violin\_boxplot\_white\_bg.png", width = 8, height = 6)

## Output of R codes

**Print the overall results**

> cat("Overall Mean:", overall\_mean, "\n")

Overall Mean: 1206.427

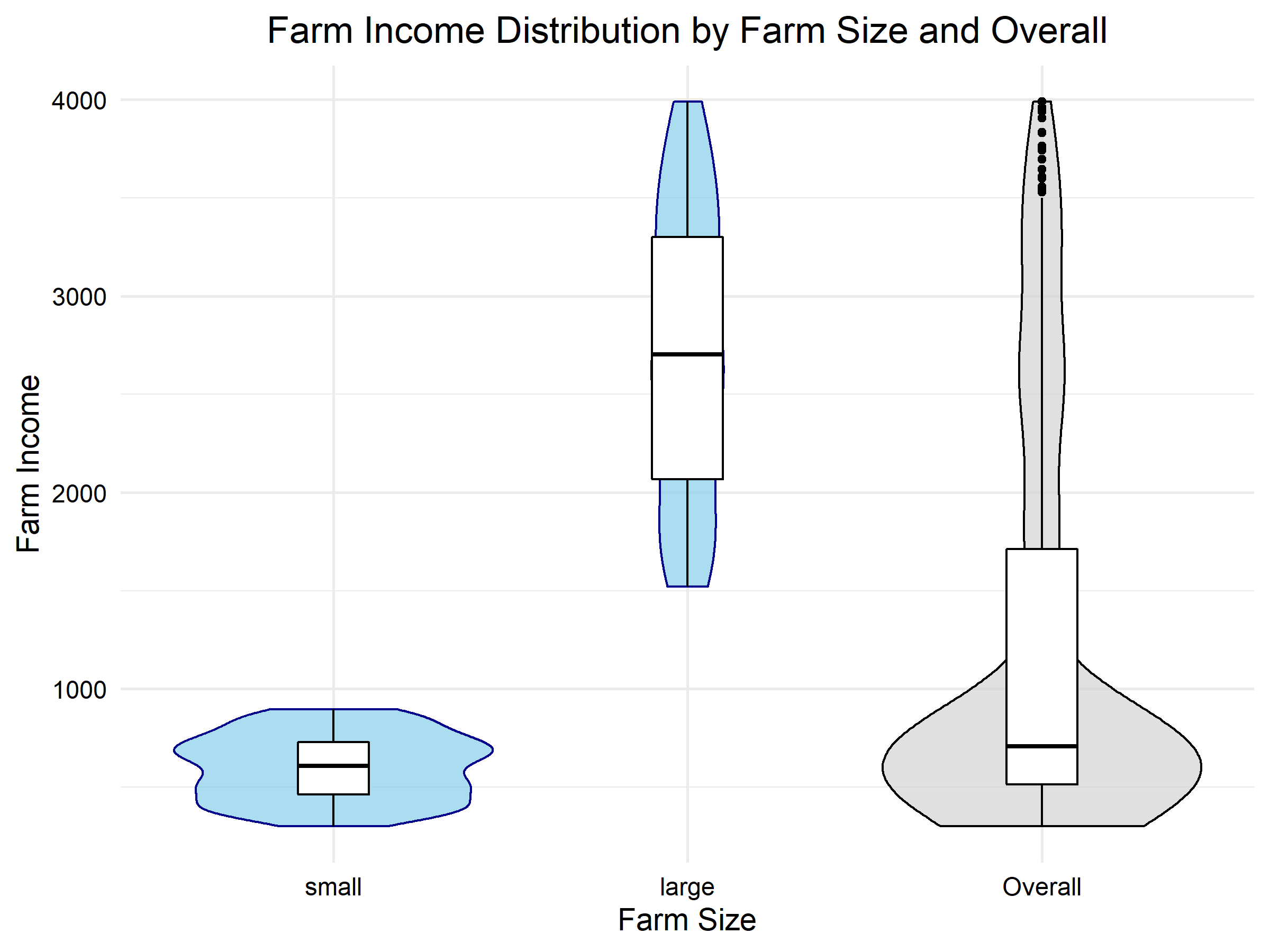
Standard Error: 60.09113

95% Confidence Interval: 1088.65 1324.203

**Statistics by FarmSize:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| FarmSize | Mean | SE | CI\_Lower | CI\_Upper |
| 1 large | 2713 | 77.9 | 2560 | 2866 |
| 2 small | 601 | 11.4 | 579 | 623 |

**Plot of Farm Income Distribution by Size and Overall**

****