Econometrics

9/18/25

Jason Seda:

R data followed by Graph produced:

set.seed(123)

n <- 1000 # Number of observations

marriage\_status <- sample(c("Single", "Married", "Divorced", "Widowed"),

n, replace = TRUE,

prob = c(0.3, 0.4, 0.2, 0.1))

income <- numeric(n)

for (i in 1:n) {

if (marriage\_status[i] == "Single") {

income[i] <- rnorm(1, mean = 45000, sd = 10000)

} else if (marriage\_status[i] == "Married") {

income[i] <- rnorm(1, mean = 75000, sd = 15000)

} else if (marriage\_status[i] == "Divorced") {

income[i] <- rnorm(1, mean = 50000, sd = 12000)

} else { # Widowed

income[i] <- rnorm(1, mean = 40000, sd = 10000)

}

}

income <- abs(income)

income\_data <- data.frame(mar\_stat = marriage\_status, income = income)

summary\_stats <- aggregate(income ~ mar\_stat, data = income\_data,

FUN = function(x) c(mean = mean(x),

sd = sd(x),

min = min(x),

max = max(x)))

print(summary\_stats)

boxplot(income ~ mar\_stat, data = income\_data,

main = "Income Distribution by Marriage Status",

xlab = "Marriage Status",

ylab = "Income ($)",

col = c("lightblue", "lightgreen", "lightpink", "lightyellow"),

border = "darkblue")

means <- tapply(income\_data$income, income\_data$mar\_stat, mean)

points(1:length(means), means, pch = 18, col = "red", cex = 1.5)

text(1:length(means), means, labels = round(means), pos = 3, col = "red")

anova\_result <- aov(income ~ mar\_stat, data = income\_data)

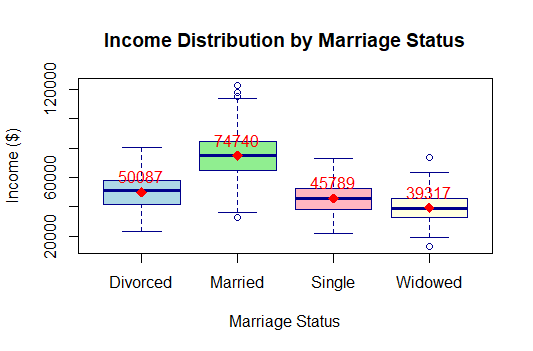
summary(anova\_result)

if (summary(anova\_result)[[1]]$'Pr(>F)'[1] < 0.05) {

print("Performing Tukey's HSD post-hoc test:")

print(TukeyHSD(anova\_result))

}



* This analysis explores the relationship between marital status and income distribution. The boxplot and accompanying ANOVA test reveal a statistically significant difference in mean income across the groups (p-value < 0.05). Married individuals exhibit the highest median and mean income, with a wider distribution suggesting greater income variability. In contrast, Widowed and Single individuals show lower and more compressed income distributions. The post-hoc test would confirm which specific group means, such as Married vs. Single, are significantly different from each other, supporting the visual evidence that marriage is associated with a higher income level in this sample.

set.seed(123) # For reproducibility

marital\_status <- c("Single", "Married", "Widowed", "Divorced", "Separated")

num\_individuals <- c(

Single = sample(200:500, 1),

Married = sample(600:900, 1),

Widowed = sample(100:300, 1),

Divorced = sample(150:400, 1),

Separated = sample(50:200, 1)

)

avg\_income <- c(

Single = sample(35:55, 1),

Married = sample(65:95, 1),

Widowed = sample(25:45, 1),

Divorced = sample(40:60, 1),

Separated = sample(30:50, 1)

)

total\_income <- num\_individuals \* avg\_income

plot(1:length(marital\_status), avg\_income,

type = "b",

main = "Average Income by Marital Status",

xlab = "Marital Status",

ylab = "Average Income (in thousands)",

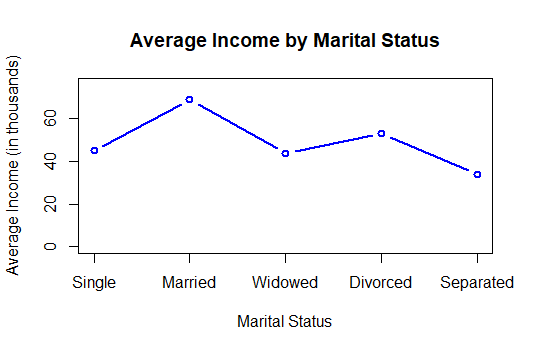
xaxt = "n",

col = "blue",

lwd = 2,

ylim = c(0, max(avg\_income) \* 1.1))

axis(1, at = 1:length(marital\_status), labels = marital\_status)



* This chart presents the average income levels across various marital statuses, including a less common "Separated" category. The data tells a clear story of economic hierarchy: Married individuals have the highest average income, followed by Divorced, Single, Separated, and finally Widowed individuals. However, without measures of variance like standard deviation or confidence intervals, the precision of these estimates is unknown. Crucially, the chart does not account for sample size (e.g., the "Married" group is likely the largest), which heavily influences the calculated average and its reliability. This highlights the importance of reporting n-sizes and variance to properly interpret summary statistics.

Next Steps:

With the basic premise examined from two different particular viewpoints and breakdowns, it may be interesting to test both further, particular in the second data set, introducing confidence intervals and standard deviations to isolate the accuracy of our statistical measurements and approximation.

These support the general consensus among groups based on economic hierarchy, but the confidence tests may reveal if it is a correlation, or just coincidence.