VSN Sai Krishna Mohan Kocherlakota

Title: Module 4 R Practice

Submission Date: 04th December 2023

INTRODUCTION:

In the world of statistical analysis, understanding tests like t-tests is crucial. They help compare means between two groups, like male and female cat bodyweights or the impact of a workshop on sleep quality. This code demonstrates conducting and visualizing both independent (Part 1) and paired (Part 2) t-tests in R, providing insights into differences between groups or before-and-after scenarios. Utilizing libraries like ggplot2 and t.test functions, it assesses variations, aiding in informed decision-making based on observed data. Understanding these statistical methods is invaluable in various fields, empowering sound analysis and decision-making.

TASKS:

1. Part-1

```
> #Required Libraries
> library(ggplot2)
> library(ggthemes)
> library(ggeasy)
> library(pacman)
> library(lubridate)
> library(janitor)
> library(tidyverse)
> library(knitr)
> library(gridExtra)
> library(MASS)
> #Part-1
> data(cats)
> male <- subset(cats, subset = (cats$Sex == "M"))$Bwt
> female <- subset(cats, subset = (cats$Sex == "F"))$Bwt
>
> t_test_result <- t.test(male, female, var.equal = FALSE)
> t_test_result
```

This code above conducts a two-sample t-test to compare the bodyweights of male and female cats using the `cats` dataset. It separates the dataset into male and female observations and then performs a t-test to check if there's significant difference in bodyweights between the two sexes.

The `t_test_result` variable holds the result of this statistical test. The `boxplot` function creates a visual representation of the bodyweight distribution for males and females using different colors for each group. This visualization helps in understanding the distribution and potential differences between male and female cat bodyweights.
The output you've provided is the result of a Welch Two Sample t-test comparing male and female cat bodyweights.
Let's break down what each part means:

- t-value: 8.7095 - This is the calculated t-statistic. It represents how many standard deviations the difference between the two-sample means is, relative to the variation in the data.

- Degrees of Freedom (df): 136.84 - Approximate degrees of freedom used in the t-distribution for this test. It's calculated based on the sample sizes and variances of the two groups.

- Confidence Interval (CI): The 95% confidence interval for the difference in means is 0.418 to 0.663. This means that we can be 95% confident that the true difference in means between male and female cat bodyweights falls within this range.

Overall, this test strongly suggests a significant difference in bodyweights between male and female cats, with males, on average, having higher bodyweights than females.

```
> # Visualizing two-sample t-test (Part 1)
> boxplot(list(Male = male, Female = female), col = c("blue", "pink"), main = "Male vs Female Cat Bodyweights")
```

The code generates a boxplot visualization comparing the bodyweights of male and female cats. Here's a breakdown of what each part does:

- `boxplot`: This function creates a boxplot. The `list(Male = male, Female = female)` part specifies the data for the boxplot, with 'Male' and 'Female' as labels for the two boxplots. `col = c("blue", "pink")` assigns colors to each boxplot, using blue for male and pink for female categories. `main = "Male vs Female Cat Bodyweights"` sets the main title of the plot.

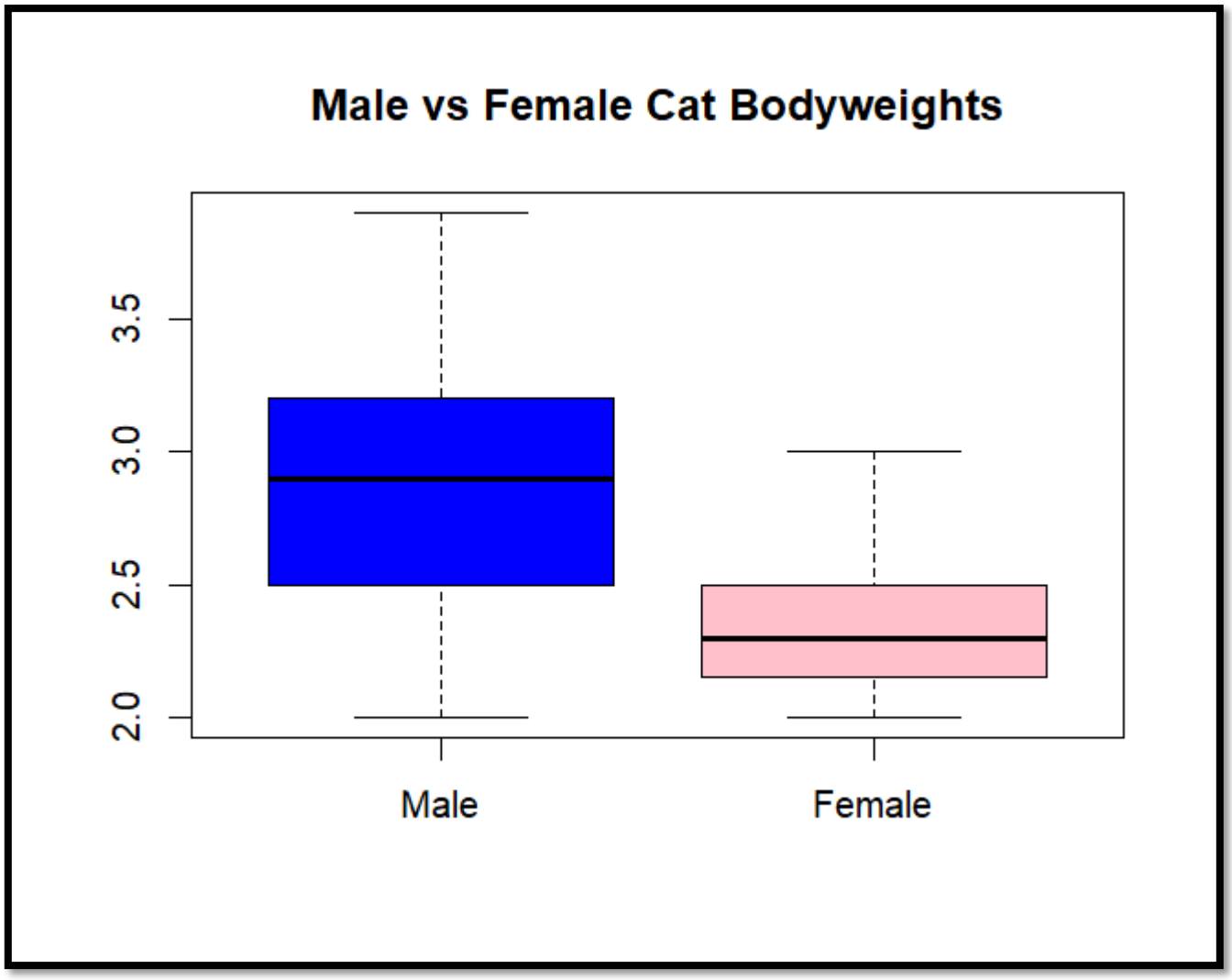

Figure 1: - Box plot of Male and Female Cat Weights

2.Part-2

```
> #Part-2
> before_workshop <- c(4.6, 7.8, 9.1, 5.6, 6.9, 8.5, 5.3, 7.1, 3.2, 4.4)
> after_workshop <- c(6.6, 7.7, 9.0, 6.2, 7.8, 8.3, 5.9, 6.5, 5.8, 4.9)
>
> paired_t_test <- t.test(before_workshop, after_workshop, paired = TRUE)
> paired_t_test

        Paired t-test

data:  before_workshop and after_workshop
t = -1.9481, df = 9, p-value = 0.08322
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -1.33995222  0.09995222
sample estimates:
mean difference
          -0.62
```

This code above performs a paired t-test to compare the sleep quality before and after a workshop.
Here's what the output means:

- Paired t-test: This test compares the means of two related groups, in this case, the before and after workshop sleep qualit y ratings for the same individuals.
- t-value: -1.9481 - This is the t-statistic for the paired t-test. It represents how many standard deviations the mean differen ce is away from zero.
- Degrees of Freedom (df): 9 - The degrees of freedom used in the t-distribution for this test, calculated based on the numb er of pairs.
- Confidence Interval (CI): The 95% confidence interval for the difference in means ranges from -1.34 to 0.10. This interval i ncludes zero, reinforcing the suggestion that there might not be a significant change in sleep quality.

```
> # Visualizing paired t-test (Part 2)
> plot(before_workshop, after_workshop, xlab = "Sleep Quality Before Workshop", ylab =
"Sleep Quality After Workshop",
+      main = "Sleep Quality Before vs After Workshop", pch = 19)
> abline(0, 1, col = "red") # Adds a diagonal line for reference (equality line)
```

This code generates a scatter plot to visualize the paired data of sleep quality ratings before and after a workshop. Here's a breakdown of the code:
- `plot`: This function creates a scatter plot. `before_workshop` values are plotted on the x-axis, `after_workshop` values on the y-axis. The `xlab` and `ylab` parameters set the labels for the x and y-axes, respectively. `main` sets the title of the plot, and `pch = 19` modifies the plotting character to solid circles.
- `abline(0, 1, col = "red")`: This function adds a reference line to the plot. The line `y = x` represents perfect equality between sleep quality ratings before and after the workshop. It's marked in red for visibility.

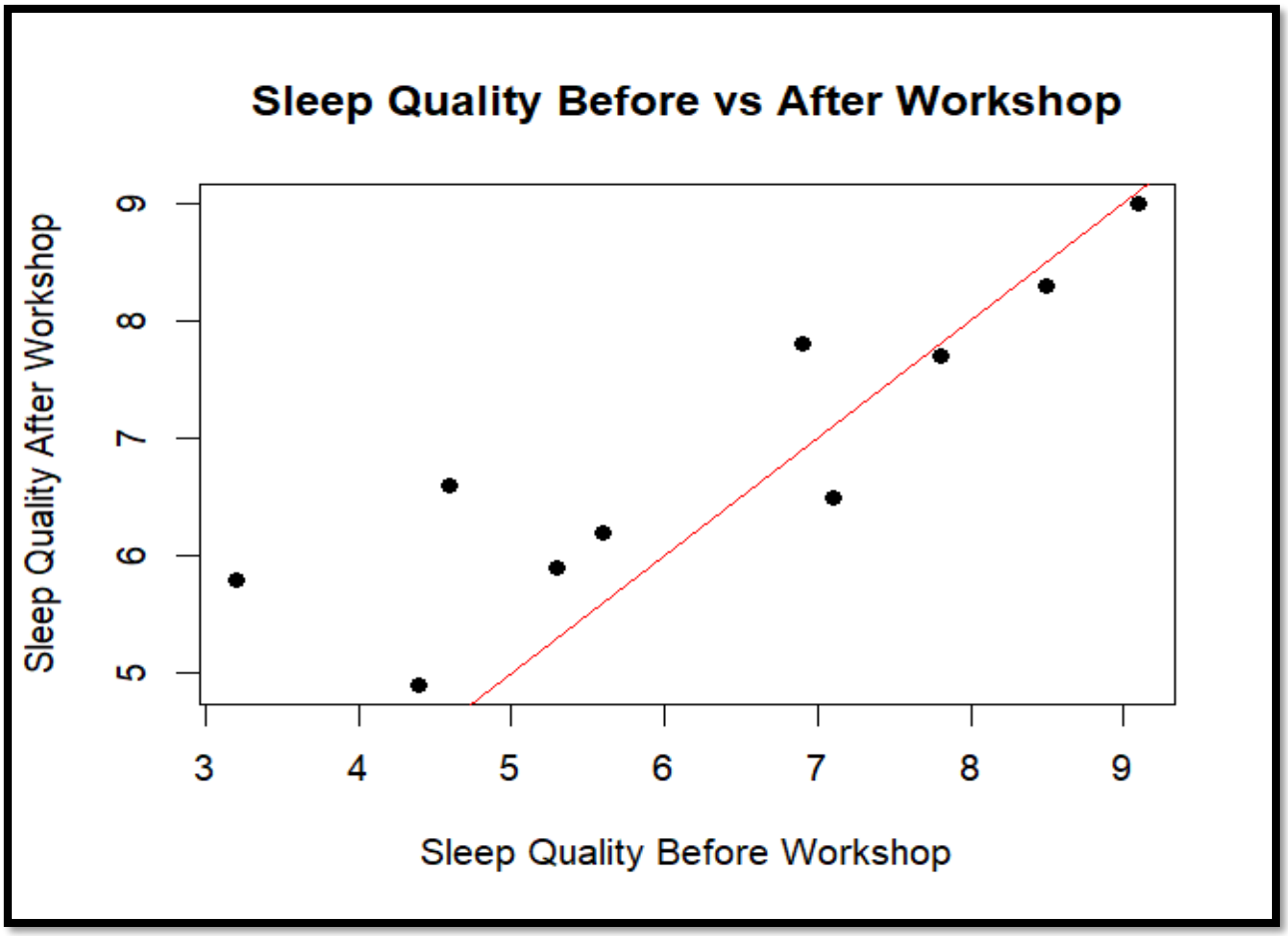

Figure 2: Sleep Quality

CONCLUSION

The analysis conducted consists of two parts: a comparison of male and female cat bodyweights and an evaluation of sleep quality before and after a workshop.

For the cat bodyweight comparison, a Welch Two Sample t-test was performed, showing a significant difference between male and female cats ($t = 8.7095$, $p < 0.001$). The average bodyweight for male cats (2.9) was notably higher than for female cats (2.36), supported by a 95% confidence interval (0.42 to 0.66).

Regarding sleep quality, a paired t-test revealed a mean decrease of 0.62 in sleep quality after the workshop, but the p-value (0.083) suggests this change might not be statistically significant. Visualizing the paired data showed varying results; while some points suggested decreased sleep quality after the workshop, others showed improvement. The diagonal reference line indicated no change.

In conclusion, the analysis highlights a significant difference in bodyweights between male and female cats. However, regarding sleep quality before and after the workshop, the evidence for a substantial change is inconclusive. The workshop's impact on sleep quality appears varied among participants, with no clear overall trend observed in the data. Further investigation or a larger sample size might be needed to draw definitive conclusions about the workshop's effect on sleep quality.