Module 4 R Practice

Name: Aishwarya Doni

College of Professional Studies, Northeastern University

ALY6010: Probability Theory and Introductory Statistics
Professor: Kenneth Parker

November 23, 2024

Introduction

This report examines statistical tests applied to “cats” datasets using R and to draw insights.

- Part 1 investigates differences in body weights between male and female cats using the cats dataset from the MASS package.
- Part 2 evaluates the impact of meditation on sleep quality using a paired t-test.

Working with Datasets in MASS

1. **Viewing Available Datasets**
   The MASS package contains built-in “cats” datasets that is utilized for statistical analysis. To view the available dataset, the library(MASS) code is used after installing it.
2. **Exploring the “cats” dataset**
   This dataset contains information about 144 cats, including their gender, body weight, and heart weight.
   To explore this dataset, the following steps are used:
   - **Load and view the dataset**:

```
> data(cats)
> head(cats)
  Sex Bwt Hwt
1   F 2.0 7.0
2   F 2.0 7.4
3   F 2.0 9.5
4   F 2.1 7.2
5   F 2.1 7.3
6   F 2.1 7.6
```

     The image displays the first few rows of the cats dataset from the MASS package in R, retrieved using head(cats). The first six rows showcase the data for female cats (Sex=F) with their corresponding body and heart weights. For example, the first cat has a body weight of 2.0kg and a heart weight of 7.0g.
   - **Preview and summarize the dataset**:

```
> str(cats)
'data.frame':   144 obs. of  3 variables:
 $ Sex: Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 ...
 $ Bwt: num  2 2 2 2.1 2.1 2.1 2.1 2.1 2.1 2.1 ...
 $ Hwt: num  7 7.4 9.5 7.2 7.3 7.6 8.1 8.2 8.3 8.5 ...
> summary(cats)
 Sex         Bwt             Hwt
 F:47   Min.   :2.000   Min.   : 6.30
 M:97   1st Qu.:2.300   1st Qu.: 8.95
        Median :2.700   Median :10.10
        Mean   :2.724   Mean   :10.63
        3rd Qu.:3.025   3rd Qu.:12.12
        Max.   :3.900   Max.   :20.50
>
```

Descriptive Statistics from the Summary (for 144 cats, Females: 47 and Males: 97)

a) **Minimum value:**
   Body weight 2.0kg, heart weight 6.30g- these are the smallest observed values, and this cat has the lowest body weight in the dataset. This cat might be ill or just weak.

b) **Maximum value:**
   Body weight 3.9kg, Heart weight 20.50g- these are the highest observed values, and this cat has the highest body weight in the dataset.

c) **1st Quartile:**
   25% of cats fall below the value of body weight and heart weight 2.3kg and 8.95g respectively.

d) **Median:**
   Cat with body weight 2.7kg and heart weight 10.10g has the central values for the dataset.

e) **Mean:**
   Mean values for body weight and heart weight are 2.724kg, 10.63g respectively, which is slightly high than the median values. This implies that there are very few cats which have little high body and heart weight (slightly skewed by outliers).

f) **3rd Quartile:**
   75% of cats fall below the value of body weight and heart weight 3.025kg and 12.12g respectively

g) **Body weight:** Symmetric distribution, most weights fall between 2.3 and 3.025kg.
h) **Heart weight:** Slightly skewed positively, most values range from 8.95 to 12.12g.

Analysis of Part1: Cats dataset

Hypotheses

- **Null Hypotheses (H0):** Male and female cats have the same mean body weight (μmale=μfemale).
- **Alternative Hypotheses (Ha):** Male and female cats have different mean body weights (μmale≠μfemale).
- **Result:**

```
> 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)
> print(t_test_result)

        Welch Two Sample t-test

data:  male and female
t = 8.7095, df = 136.84, p-value = 8.831e-15
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.4177242 0.6631268
sample estimates:
mean of x mean of y
 2.900000  2.359574
```

- **Interpretation:**
  a) The t-statistic is 8.7095, and the p-value is extremely small (8.831e-15), well below the significance level of 0.05.
  b) 95% confidence interval for the difference in means is [0.4177, 0.6631], indicating a significant positive difference.
  c) The mean body weight of male cats is 2.90 kg, while the mean body weight of female cats is 2.36 kg.
- **Conclusion:**
  H0 is rejected and male and female cats have significantly different body weights. Male cats are, on average, heavier than female cats.

Analysis of Part 2: Meditation and Sleep Quality

Hypotheses

- **Null Hypotheses (H0):** Meditation has no effect on sleep quality (μdifference=0).
- **Alternative Hypotheses (Ha):** Meditation improves sleep quality (μdifference>0).
- **Result:**

```
> before <- c(4.6, 7.8, 9.1, 5.6, 6.9, 8.5, 5.3, 7.1, 3.2, 4.4)
> after <- 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(after, before, paired = TRUE, alternative = "greater")
> print(paired_t_test)

        Paired t-test

data:  after and before
t = 1.9481, df = 9, p-value = 0.04161
alternative hypothesis: true mean difference is greater than 0
95 percent confidence interval:
 0.03659503        Inf
sample estimates:
mean difference
           0.62
```

- **Interpretation:**
    d) The t-statistic is 1.9481, and the p-value is 0.04161, which is less than significance level of 0.05.
    e) The mean difference in sleep quality scores (after - before) is 0.62, with a 95% confidence interval starting from 0.0366 to infinity.
- **Conclusion:**
  H0 is rejected and that meditation significantly improves sleep quality.
- **Justification for using Paired t-test:**
  This test is appropriate as the same individuals were measured before and after the intervention. This test supports for the dependency in observations. Using an independent t-test would ignore this dependency and reduce the accuracy of the results.

Comparison of Statistical Tests and Significance Levels:

1. **Type of Test Justification:**
   - **Part 1** (Cats Dataset): The Welch Two-Sample t-test was used because the goal was to compare the mean body weights of two independent groups (male and female cats). A paired t-test would not be appropriate here as the two groups are unrelated.
   - **Part 2** (Meditation and Sleep Quality): A paired t-test was used as the same individuals were measured before and after meditation, making the observations dependent. An independent t-test would ignore this dependency and reduce accuracy.
2. **Impact of Significance Level:**
   - **Part 1:** Changing the significance level from 0.05 to 0.1 does not alter the conclusion, as the p-value (8.831e-15) is far below both thresholds, confirming a significant difference in body weights.
   - **Part 2:** The p-value (0.04161) is below 0.05 but would still be significant at the 0.1 level, meaning the conclusion remains unchanged (meditation improves sleep quality).
3. **Justification for the Testing Procedures:**
   - The Welch Two-Sample t-test was chosen for Part 1 due to the independence of groups and potential unequal variances.
   - The paired t-test was selected for Part 2 as it accounts for the dependency in observations, which enhances result accuracy compared to other tests like the independent t-test.

Conclusion

- Male cats are significantly heavier than female cats, with a mean body weight of 2.90 kg compared to 2.36 kg. The null hypothesis (equal weights) is rejected as the p-value is extremely small (8.831e-15).
- Meditation significantly improves sleep quality, with a mean difference of 0.62 and a p-value of 0.04161, indicating statistical significance.

- The use of a paired t-test confirms that the observed differences in sleep quality are due to the same individuals being tested before and after meditation.
- Overall, the statistical tests successfully identified meaningful differences and effects in both datasets.

References:

1. Bluman, A. G. (2017). **Elementary statistics: A step by step approach** (10th ed.). McGraw-Hill Education.
2. Triola, M. F. (2018). **Essentials of statistics** (6th ed.). Pearson.
3. R Core Team. (2024). **R: A language and environment for statistical computing.** R Foundation for Statistical Computing. https://www.R-project.org/