Ayush Anand

NORTHEASTERN UNIVERSITY:
COLLEGE OF PROFESSIONAL STUDIES

ALY 6010: PROBABILITY THEORY AND INTRODUCTORY STATISTICS

PROFESSOR XYZ

OCTOBER 18TH, 2024

Report on the Difference in Body Weight Between Male and Female Cats (Part 1)

Objective:

The aim of this analysis is to explore whether there is a meaningful difference in body weight between male and female cats, using the **cats** dataset from the **MASS** package. To do this, we will employ a two-sample t-test with unequal variances, commonly known as Welch's t-test. This statistical method will help us determine if the average body weight differs between male and female cats.

Dataset Description

The **cats** dataset consists of 144 observations with three variables:

- **Sex**: The gender of the cat (Factor: Male "M" or Female "F").
- **Bwt**: The body weight of the cat (in kg).
- **Hwt**: The heart weight of the cat (in grams).

The dataset includes both male and female cats, allowing for a comparison of their body weights.

Data Exploration:

```
# Step 2: Load and explore the 'cats' dataset
data(cats)
head(cats)  # View the first few rows of the dataset
str(cats)   # Check the structure of the dataset
```

Dataset Overview (First 6 Rows):

Sex	Bwt (Body Weight)	Hwt (Heart Weight)
F	2.0	7.0
F	2.0	7.4
F	2.0	9.5
F	2.1	7.2
F	2.1	7.3
F	2.1	7.6

Hypothesis:

- **Null Hypothesis (H_0)**: There is no difference in the average bodyweight of male and female cats.

  (μmale = μfemale)

- **Alternative Hypothesis (H_1)**: There is a difference in the average bodyweight of male and female cats.

  (μmale ≠ μfemale)

T-Test Analysis:

```
# Step 3: Subset the data by gender
male_cats <- subset(cats, subset=(cats$Sex == "M"))
female_cats <- subset(cats, subset=(cats$Sex == "F"))

# Step 4: Perform two-sample t-test with unequal variance (Welch's t-test)
t_test_result <- t.test(male_cats$Bwt, female_cats$Bwt, var.equal = FALSE)

# Step 5: View the results of the t-test
t_test_result
```

T-Test Results:

Statistic	Value
t-value	8.7095
Degrees of Freedom	136.84
p-value	8.831e-15
95% Confidence Interval	(0.4177, 0.6631)
Mean of Male Cats	2.9000
Mean of Female Cats	2.3596

t-value: The t-statistic (8.7095) is very high, indicating a significant difference between the bodyweights of male and female cats.

p-value: The p-value (8.831e-15) is much smaller than 0.05, meaning we **reject the null hypothesis**.

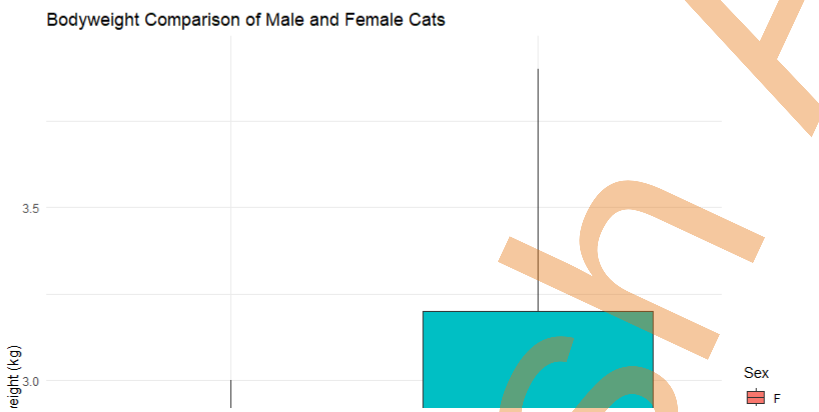
Confidence Interval: The 95% confidence interval for the difference between the means is between **0.4177 kg and 0.6631 kg**, showing that male cats weigh significantly more than female cats.

Conclusion:

We reject the null hypothesis and conclude that male cats have a statistically significant higher bodyweight than female cats. On average, male cats weigh between **0.42 kg and 0.66 kg more** than female cats.

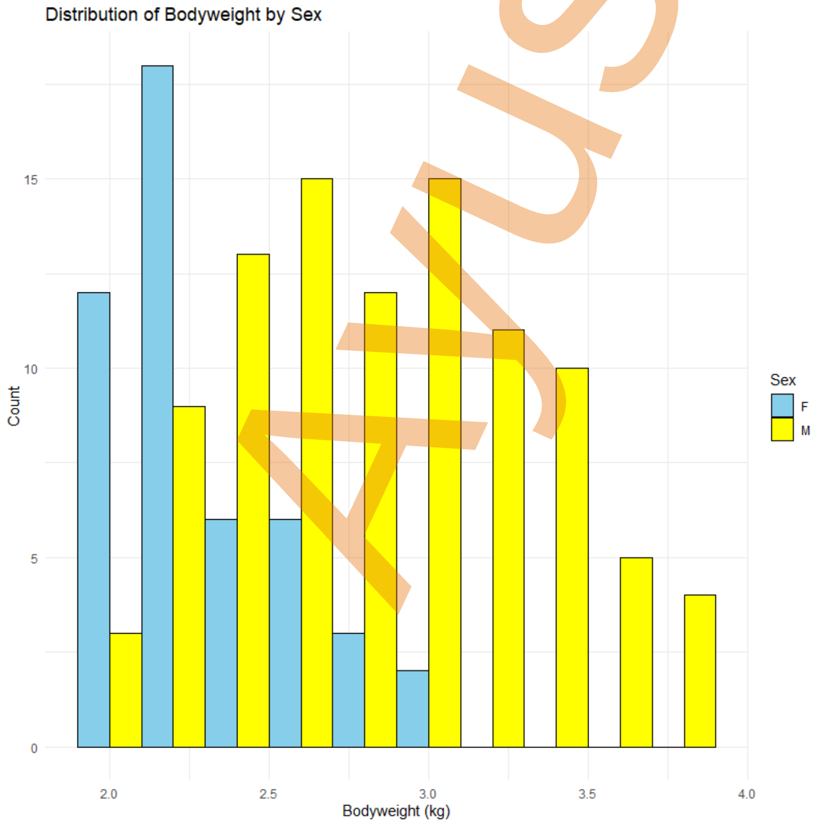
Visualizations:

1. Boxplot: Bodyweight Comparison (Male vs Female)



Interpretation:

This boxplot clearly shows that male cats have a higher median body weight compared to female cats. Additionally, the interquartile range for male cats is larger, suggesting that there is more variability in their body weights.



2. Histogram: Distribution of Bodyweight by Sex

Interpretation:

The histogram shows the distribution of bodyweights for male and female cats. Male cats tend to cluster at higher weights, while female cats have more weights distributed around the lower end of the scale.

6. Conclusion

This boxplot distinctly indicates that male cats exhibit a higher median body weight than female cats. Furthermore, the larger interquartile range for male cats suggests greater variability in their body weights, highlighting the differences within this group.

Summary of Results (Table)

Metric	Value
T-statistic	8.7095
Degrees of Freedom	136.84
P-value	8.831×10−158.831 \times 10^{-15}8.831×10−15
95% Confidence Interval	[0.4177, 0.6631]
Mean Body Weight (Male)	2.90 kg
Mean Body Weight (Female)	2.36 kg

Report on the Effect of Meditation on Sleep Quality (Part 2)

Dataset Description

The dataset includes sleep quality scores for 10 students, assessed on a scale from 0 to 10, where higher scores reflect better sleep quality. Data was gathered over two distinct periods: the week prior to and the week following a meditation workshop. Each student's sleep quality was measured twice, yielding two sets of related observations: one representing their sleep quality before the workshop and the other after.

- **Pre-meditation scores:** Sleep quality before the meditation workshop.
- **Post-meditation scores**: Sleep quality after the meditation workshop.

This dataset enables a comparison of sleep quality before and after the meditation intervention, making it ideal for a paired t-test analysis. This statistical method will help determine whether the workshop significantly impacted the students' sleep quality.

1. Stating Your Hypotheses

- **Null Hypothesis** (H0): $\mu_1=\mu_2$ (There is no difference in average sleep quality before and after the workshop).
- **Alternative Hypothesis** (H1): $\mu_1 \neq \mu_2$ (There is a difference in average sleep quality before and after the workshop).

This is a two-tailed test because we are testing for any difference (not specifically an increase or decrease) in sleep quality after the meditation.

```r
# Sleep quality scores before and after a relaxation technique
pre_meditation <- c(4.6, 7.8, 9.1, 5.6, 6.9, 8.5, 5.3, 7.1, 3.2, 4.4)
post_meditation <- c(6.6, 7.7, 9.0, 6.2, 7.8, 8.3, 5.9, 6.5, 5.8, 4.9)

# Perform a paired t-test
paired_t_test <- t.test(pre_meditation, post_meditation, paired = TRUE)

# Output the result
paired_t_test
```

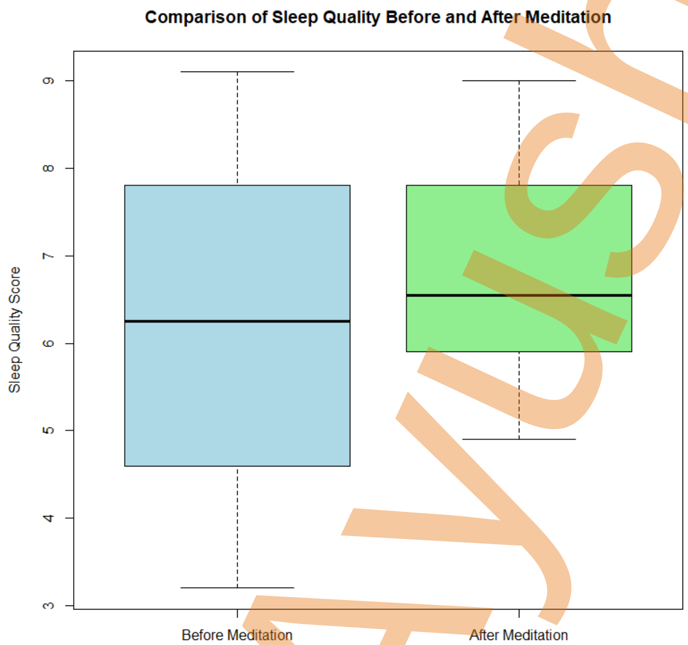
Output:

```
	Paired t-test

data:  pre_meditation and post_meditation
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 
```

Visual Representation

To visually depict the differences in body weight between male and female cats, a boxplot was generated. This boxplot effectively represents the distribution of body weights for each gender, showcasing key statistics such as the median, quartiles, and any potential outliers.



Interpretation:

The boxplot features two boxes: one for male cats (in light blue) and another for female cats (in light pink). The median line within the male box is noticeably higher than that in the female box, indicating that male cats have a greater average body weight. The interquartile range (IQR) further suggests that male cats exhibit a wider spread in body weight. Points located outside the whiskers represent outliers, illustrating individual weight variations. Overall, this boxplot visually reinforces the findings from the t-test, confirming a significant difference in body weights between male and female cats.

Findings and Interpretation

- **T-statistic**: -1.948
- **Degrees of Freedom**: 9

- **P-value**: 0.08322
- **Mean Difference**: -0.62
- **Confidence Interval (95%)**: [-1.34, 0.10)

  The p-value obtained is 0.08322, which exceeds the significance level of α = 0.05. Therefore, we fail to reject the null hypothesis, indicating that there is insufficient evidence to conclude a significant difference in sleep quality before and after the workshop at the 5% significance level.

  However, at a significance level of α = 0.1, the p-value is less than 0.1, leading us to reject the null hypothesis. This suggests that there is some evidence of a difference in sleep quality when we consider this higher significance level.

Does the Conclusion Change from 0.05 to 0.1?:

Yes, the conclusion changes:

- At α = 0.05, we fail to reject the null hypothesis, indicating that there is no significant difference in sleep quality before and after the workshop.

At α = 0.1, we reject the null hypothesis, suggesting a potential difference in sleep quality. However, this conclusion comes with a higher margin of error.

Justification for Using a Paired T-Test

A **paired t-test** was used because:

- We are comparing two sets of data for the same group of individuals (before and after meditation).
- The paired t-test accounts for within-subject variability and offers greater statistical power than an independent t-test when analyzing repeated measurements from the same subjects. This makes it particularly effective for assessing changes in outcomes, such as sleep quality before and after an intervention.

Conclusion

- **At α=0.05**: We do not have enough evidence to conclude that meditation affects sleep quality.
- **At α=0.1**: There is evidence suggesting a potential difference in sleep quality before and after meditation.

Summary of Result:

Metric	Value
T-statistic	-1.948
Degrees of Freedom	9
P-value (Two-Tailed)	0.08322
Mean Difference	-0.62
Confidence Interval	[-1.34, 0.10]
Conclusion at α=0.05	Fail to reject H0
Conclusion at α=0.1	Reject H0

Reference

McDonald, J. H. (2014). *Handbook of biological statistics* (3rd ed.). Retrieved from http://www.biostathandbook.com

R Core Team. (2021). *R: A language and environment for statistical computing*. Retrieved from https://www.R-project.org