t-test of Cats & Sleep

R Practice 4: Two-sample t-test

ALY6010

Prepared by: Heejae Roh
Presented to: Professor Behzad Ahmadi
Date: Dec 4th, 2022

INTRODUCTION

Two-sample t-test is the process of recognizing differences and changes and seeing if there are any differences. It is very important in the process of identifying change and wanting to make process better, in other words it could be 'optimization'. Two-sample t-test is comparing two parameters such as two means, two proportions, or two variances (Bluman, 2017), so we can compare means, proportions, or variances also.

In this module, we will learn how to do a t-test on average of different samples. To help the process understood, I wanted to use visualization which can give us understanding and insight of data.

PART 1

Question: Do male and female cat samples have the same bodyweight ("Bwt")? (using α = 0.05)

Step 1 Hypothesis.

- H0: μ1 = μ2 (claim) and H1: μ1 ≠ μ2
- H0: μ1 - μ2 = 0 (claim) and H1: μ1 - μ2 ≠ 0
- H0: there is no significant difference in the average bwt of cat by sex. (claim)
- H1: there is a significant difference in the average bwt of cat by sex.

Data analysis & Visualization.

```
 Descriptive statistics by group
Sex: F
     vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
Sex*    1 47 1.00 0.00    1.0    1.00 0.00 1.0   1   0.0  NaN      NaN 0.00
Bwt     2 47 2.36 0.27    2.3    2.33 0.30 2.0   3   1.0 0.87    -0.21 0.04
Hwt     3 47 9.20 1.36    9.1    9.20 1.48 6.3  13   6.7 0.10    -0.17 0.20
---------------------------------------------------------------------------------
Sex: M
     vars  n  mean   sd median trimmed  mad min  max range skew kurtosis   se
Sex*    1 97  2.00 0.00    2.0    2.00 0.00 2.0  2.0   0.0  NaN      NaN 0.00
Bwt     2 97  2.90 0.47    2.9    2.89 0.59 2.0  3.9   1.9 0.13    -0.80 0.05
Hwt     3 97 11.32 2.54   11.4   11.21 2.67 6.5 20.5  14.0 0.55     0.58 0.26
```

Figure 1 descriptive analysis of Cats data

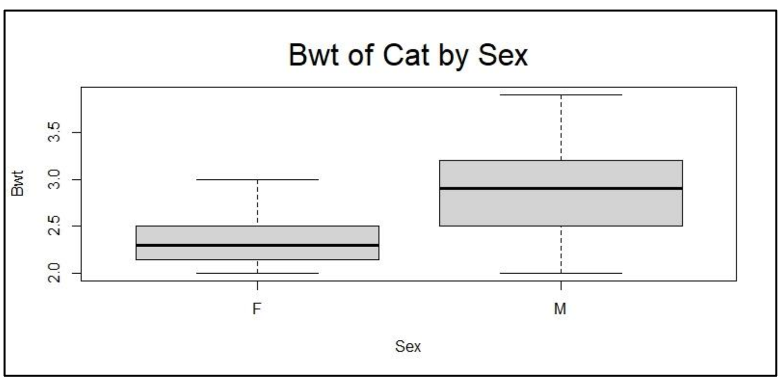

Figure 2 (boxplot) bwt of cats by sex

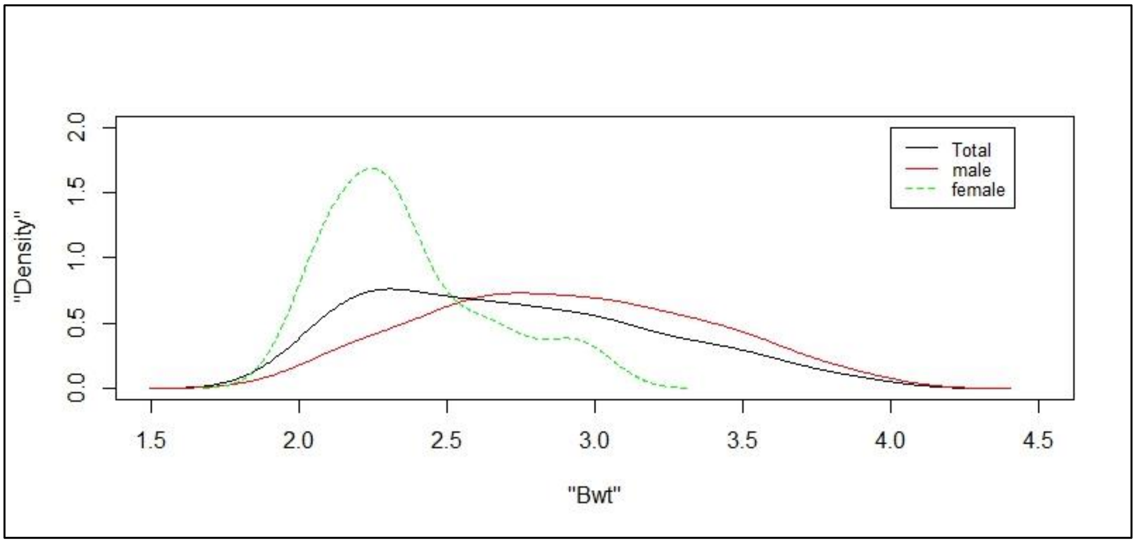

Figure 3 (density plot) bwt of cats by sex

Finding:

1. Mean of female(F) bwt is 2.36 and median is 2.3. sd of F bwt is 0.27
2. Mean of male(M) bwt is 2.9 and median is 2.89. sd of M bwt is 0.47
3. Median of M is higher than F, F has smaller range comparing with M. F data is more centralized around 2.3.
4. Density of all is Right-Skewed, F is mostly Right-Skewed and has low sd.

Step 2 Find the critical value. (paired=FALSE)

- α = 0.05, two-tailed t-value is 0.475, 0.864

Step 3 Compute the test value.

- t-value is 7.3307 with t-test().

Step 4 Make the decision.

- Reject the null hypothesis and accept claim. Since 0.864 < 7.3307

Step 5 Summarize the results.

- There is a significant difference between female and male bwt of cat.

Interpretation:

1. The answer is male and female cats sample have different bodyweight.
2. I use Rule of Thumb which is ‘if the ratio of the larger variance to the smaller variance is less than 4 then we can assume the variances are approximately equal (Zach, 2021)’, so I set ‘var.equal = TRUE’. because the sd ratio of male and female cats sample is 0.47/0.27=1.74 < 4.
3. (Data) There is enough evidence to support that female and male cat samples bodyweight is not same.
4. (Data) Mean of male cat samples’ bodyweight is greater than mean of female cats

sample’s bodyweight.
5. If we used visualization, we could expect what would be the result and cross-checking result with visualization.
6. If the result was not the thing we expected, we can explain what the reason is.
7. We can use α, p-value, and CI to confirm the hypothesis is right or not.

PART 2

Question: The researchers claimed that meditation improves sleeping quality. Is it true? (using α = 0.05)

Step 1 Hypothesis.

- H0: μ1 = μ2 and H1: μ1 < μ2 (claim)
- H0: μ1- μ2 = 0 and H1: μ1 - μ2 < 0 (claim)
- H0: there is no significant difference in sleeping quality before and after meditaion
- H1: sleeping quality after doing meditation is greater than before (claim)

Data Analysis & Visualization:

```
 Descriptive statistics by group
Workshop: after
          vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
sleep        1 10 6.87 1.28   6.55    6.85 1.41 4.9   9   4.1 0.18    -1.38 0.4
workshop*    2 10 1.00 0.00   1.00    1.00 0.00 1.0   1   0.0  NaN      NaN 0.0
---------------------------------------------------------------------------------
Workshop: before
          vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
sleep        1 10 6.25 1.93   6.25    6.28 2.37 3.2 9.1   5.9 -0.02    -1.52 0.61
workshop*    2 10 1.00 0.00   1.00    1.00 0.00 1.0 1.0   0.0   NaN      NaN 0.00
```

Figure 4 descriptive analysis of sleep quality before & after meditation

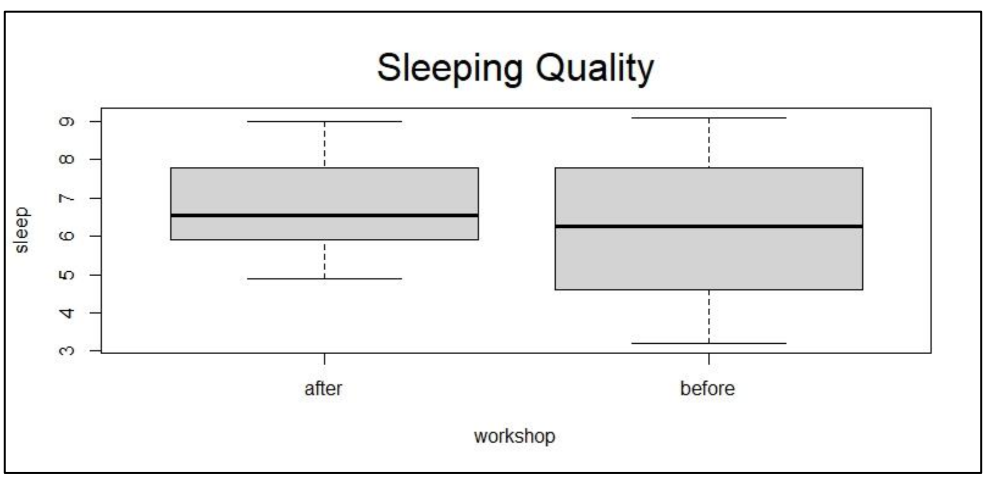

Figure 5 (boxplot) sleep quality of before and after meditation

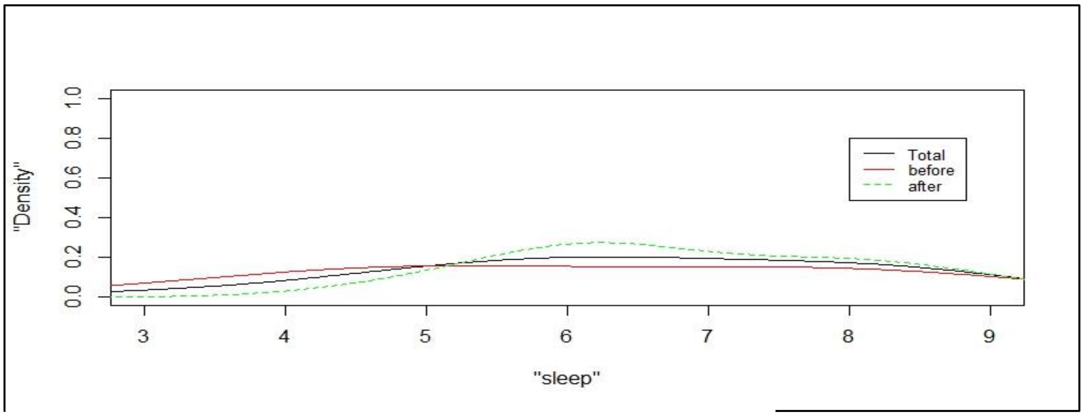

Figure 6 (density plot) sleep quality of before and after meditation

Finding:

1. Mean of sleep quality after workshop (SQ of A) is 6.87 and median is 6.55. sd of after meditation sleep quality is 1.28.
2. Mean of sleep quality before workshop (SQ of B) is 6.25 and median is 6.28. sd of before meditation sleep quality is 1.98.
3. Median of SQ of A is higher than SQ of B, SQ of A has smaller range comparing with SQ of B. SQ of A data is more centralized around 6.55.
4. SQ of A has low sd and higher mean & median.

Step 2 Find the critical value. (var.equal = TRUE, since Rule of Thumb)

- α = 0.05/2 in two-tailed, and p-value = 0.08

Step 3 Compute the test value.

- α/2(0.025) < 0.08

Step 4 Make the decision.

- Not reject the null hypothesis and reject claim. Since α/2(0.025) < 0.08

Step 5 Summarize the results.

- There is a no significant difference in sleep quality between before and after meditation.

Interpretation:

1. The answer is ‘not true’ There is no significant evidence that sleep quality between after and before workshop (meditation) is different.
2. In descriptive analysis and visualization, the result looks getting better after workshop, but in the t-test we can confirm that there is not significant difference.
3. In real world, maybe we can do more test with changing steps.
4. If the significance value changes from 0.05 to 0.1, the critical value is 0.1/2= 0.05, so the result is same.

QUESTION by MYSELF

1. This should be a paired T-test?
   First cats' bwt example uses unpaired t-test, because male and female bwt is independent.
   Second sleep quality example uses paired t-test, because before and after sleep quality is dependent.
2. Does your conclusion change if the level of significance changes from 0.05 to 0.1?
   This case can be used in second sleep quality example, the critical value is 0.1/2= 0.05, so the result is same. If the critical value is 0.2, then the result will be change. In other words, if the critical value is 0.2, the null hypothesis will be rejected.
3. Justify the testing procedure you choose, why use the test of your choice rather than other tests?
   Two-sample t-test is comparing two parameters such as two means, two proportions, or two variances (Bluman, 2017). In the cats bwt example I have two sample which is independent and want to see whether the mean of bwt is different or not. Therefore, I use unpaired two sample t-test.
   In the sleep quality example I have two sample which is dependent and want to see whether the mean of sleep quality is different or not. Therefore, I used paired two sample t-test.

CONCLUSION

In this assignment, two samples were examined. Cats' bodyweight sample had a significant difference between the two means and sleep quality samples had no difference. During this process, I was even able to think about if I were a researcher in the real world.

I also saw the power of visualization once again. No matter how hard we try to explain with numbers, writing, and words, the truth is hard to make other concentrated. But what if you show them visualization things like graph? It becomes a skill to intuitively persuade the other person.

Going back to the t-test part, we first need to check if the conditions match the t-test. After that, change the real-world question into hypothesis form and check the claim again. Next step is that we find the critical value and compare it to the test statistic and draw conclusions. I feel like I have the power to solve a certain problem through a t-test.

REFERENCE

Bluman, Allan. (2017). Elementary statistics: a step by step approach 10th edition. McGraw-Hill.

STHDA. (n.d.). Unpaired two-samples t-test in R. Retrieved from http://www.sthda.com/english/wiki/unpaired-two-samples-t-test-in-r

idrrio. (2022, September 29). Describe.by: basic summary statistics by group. Retrieved from https://rdrr.io/cran/psych/man/describe.by.html

RDocumentation. (n.d.). T.test: student's t-test. Retrieved from https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/t.test

Ivan Calandra. (2010, November 10). [R] plot & xlim/ylim & range of axis. Retrieved from https://stat.ethz.ch/pipermail/r-help/2010-November/259254.html

STHDA. (n.d.). Add legends to plots in R software : the easiest way. Retrieved from http://www.sthda.com/english/wiki/add-legends-to-plots-in-r-software-the-easiest-way

ZACH. (2021, April 11). How to Determine equal or uequal variance in t-tests. STATOLOGY. Retrieved from https://www.statology.org/determine-equal-or-unequal-variance/

STHDA. (n.d.). Paired samples t-test in R. Retrieved from http://www.sthda.com/english/wiki/paired-samples-t-test-in-r

R-Codes

```
install.packages("MASS")
install.packages("psych")
install.packages("stats")
library(easypackages)
libraries("MASS", "psych", "ggpubr", "gplots", "graphics")

str(cats)
headtail(cats)

cats

# descriptive statistics and do subset()
describeBy(cats,list(Sex=cats$Sex))
male <- subset(cats, subset=(cats$Sex=="M"))
female <- subset(cats, subset=(cats$Sex=="F"))
male

#Boxplot
plot(Bwt ~ Sex, data = cats)
mtext("Bwt of Cat by Sex", side=3, line=1, cex=2)

#checking density plot
plot(density(male$Bwt))
plot(density(female$Bwt))

#overlay plot
table(female$Bwt)
plot(x="Bwt",
    y="Density",
    xlim=range(1.5:4.5),
    ylim=range(0:2))
lines(density(cats$Bwt), col = "black")
lines(density(male$Bwt), col = "red")
lines(density(female$Bwt), col = "green", lty=2)
legend(4, 2, legend=c("Total", "male", "female"),
      col=c("black", "red", "green"), lty=c(1,1,2), cex=0.8)

#t.test, two tail
t.test(male$Bwt, female$Bwt, alternative = "two.sided", var.equal = TRUE)

#can use this code also
t=t.test(Bwt ~ Sex, data = cats, var.equal=FALSE)

#importing dataset
```

```r
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)

#making data.frame from imported dataset
df.b <- data.frame(before,"before")
df.b
df.w <- data.frame(after,"after")
df.w

#change colnames before visualization
colnames(df.b)[which(names(df.b) == "X.before.")] <- "workshop"
colnames(df.w)[which(names(df.w) == "X.after.")] <- "workshop"
colnames(df.b)[which(names(df.b) == "before")] <- "sleep"
colnames(df.w)[which(names(df.w) == "after")] <- "sleep"

#merge the columns
df <- rbind(df.b, df.w)
df

#descriptive statistics and making factor
describeBy(df,list(Workshop=df$workshop))
df$workshop <- as.factor(df$workshop)
df
str(df)

#Boxplot
plot(sleep ~ workshop, data = df)
mtext("Sleeping Quality", side=3, line=1, cex=2)

before <- subset(df, subset=(df$workshop=="before"))
after <- subset(df, subset=(df$workshop=="after"))

#overlay plot
table(df$sleep)
plot(x="sleep",
    y="Density",
    xlim=range(3:9),
    ylim=range(0:1))
lines(density(df$sleep), col = "black")
lines(density(before$sleep), col = "red")
lines(density(after$sleep), col = "green", lty=2)
legend(8, 0.8, legend=c("Total", "before", "after"),
      col=c("black", "red", "green"), lty=c(1,1,2), cex=0.8)

#t-test
t.test(df.b$sleep, df.w$sleep, paired=TRUE, alternative = "two.sided", var.equal = TRUE)
```