Graphical user interface, text

Description automatically generated with medium confidence

Logo

Description automatically generated

**Abhilash Dikshit**

Course: ALY 6010

Instructor: Mohsen Soltanifar

Module 4 - R Practice

Date: 2022/11/27

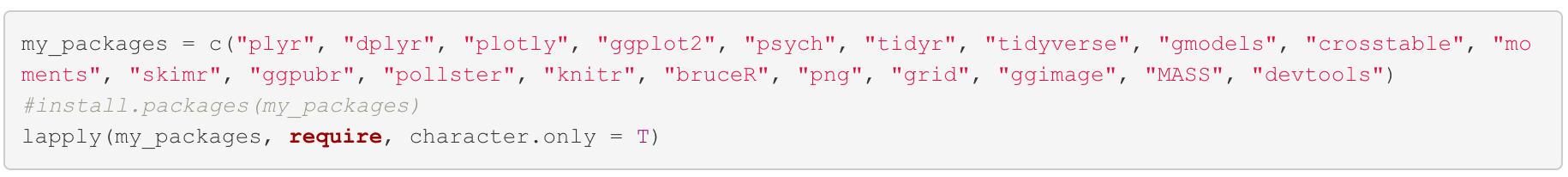
**Introduction**

Hypothesis testing is the art of testing if variation between two sample distributions can just be explained through random chance or not. If we have to conclude that two distributions vary in a meaningful way, we must take enough precaution to see that the differences are not just through random chance. A statistical hypothesis is an assumption made by the researcher about the data of the population collected for any experiment. It is not mandatory for this assumption to be true every time.Hypothesis testing, in a way, is a formal process of validating the hypothesis made by the researcher.

In this report, we will be conducting hypothesis testing on pre-loaded R data sets in the MASS package. For two-sample t-test, we use “cats” data set to test the variance of body weights in male and female cats based on gender. Later, we evaluated if the meditation improves the sleep quality or not before and after the workshop

**Installation of R package: MASS**

Initially we installed the package "MASS” to download the “cats” dataset for hypothesis testing.



**Part 1: Two-sample t-test with unequal variance (Cat Dataset):**

|  |  |
| --- | --- |
| Table  Description automatically generated | In the dataset “cats”, we renamed the column 2 and 3, and deleted ‘NA’ values, if any. In the data frame, we have 144 observations and 3 columns i.e., Sex, Body weight and Heart weight. The minimum Body weight is 2.0 and maximum is 3.8. |

**Identify if male and female cats have the same bodyweight or not?**

Our null and alternate hypothesis are:

H0 = Body weights of both male and female cats are same

H1 = Body weights of both male and female cats are different

Firstly, we separated the vectors for female and male cats based on bodyweights.



|  |  |
| --- | --- |
| Text, letter  Description automatically generated | After that, we performed t.test () with significance level of 0.05. |
|  | In a Two sample t-test, for the significance level 0.05, the P-value of 0.000000000000008831 is lesser than significance level. Hence, we reject the null hypothesis H0 due to insufficient data to identify if the body weights of both male and female cats are same. |

Boxplot representation of cat’s body weight and heart weight:

|  |  |
| --- | --- |
| Chart  Description automatically generated | From the following plot, we can interpret that by increasing the body weight of cats, their heart weight increase. Hence, body weight and heart weight are directly related to each other. |

We will check if the trend is the same for both the genders as the hypothesis test shows that their body weight is not equal to each other.

Relation between body weight and heart weight:

Chart, scatter chart, bubble chart

Description automatically generated

Plot A and B shows the relation between body weight and heart weight in Male and Female cats.

There is a slight difference in the relation between the male and female cats based on two gender as male cats have dense plot rather than female cats.

**Part 2: Two-sample t-test for evaluating whether meditation has an effect on sleep quality based on workshop:**

**The researchers claimed that meditation improves sleeping quality. Is it true?**

Firstly, we took the average sleeping quality scores in the week before the workshop and in the week following the workshop.

Graphical user interface, text, application, email

Description automatically generated

|  |  |
| --- | --- |
| Based on the summary of dataset and the mean values, the average sleeping score before and after meditation is 6.25 and 6.87 respectively which represents the increase in their sleeping quality. | Text  Description automatically generated |

Our null and alternate hypothesis are:

H0= meditation improve sleeping quality before and after the workshop

H1= meditation does not improve sleeping quality before and after the workshop

|  |  |
| --- | --- |
| Text  Description automatically generated | Text  Description automatically generated |

In a Two sample t-test, for the significance level 0.05, the P-value of 0.41 is greater than significant level. Hence, we accept null hypothesis H0 and we can conclude that sleep quality is same before and after meditation based on the workshop.

|  |  |
| --- | --- |
| In a Paired sample t-test, for the significance level 0.05, the P-value of 0.08322 is greater than significant level. Hence, we accept the null hypothesis H0 and we can conclude that sleep quality is same before and after meditation based on the workshop  In a Paired sample t-test, for the significance level 0.1, the P-value of 0.08322 is lesser than significant level. Hence, we reject the null hypothesis H0 and we can conclude that meditation does not improve sleeping quality before and after the workshop. | Text  Description automatically generated |

We can also identify by changing the significant level, the percent of confidence interval and its range changed. However, the p value didn’t change, and it did have an impact on null hypothesis HO.

Chart, box and whisker chart

Description automatically generated

**Justify the testing procedure which we used and why we use this test rather than other tests?**

For the hypothesis testing, we chose two sample t-test for both the dataset “cats” and “sleeping quality” along with paired t-test for our Part 2 analysis to check if meditation improves sleeping quality before and after the workshop.

In Part 1, we used two sample t-test to determine if male and female cats are equal in body weight. As our samples are statistically independent, we used two sample t-test instead of paired t-test.

In Part 2, initially we used two sample t-test and later we identified as the data is in the form of matched pairs, we used paired t-test. The **paired samples t-test** is used to compare the means between two related groups of samples. In this case, we have two values (i.e., pair of values) for the same samples. Hence, using **R software**, we wanted to know if meditation improves the sleeping quality before and after the workshop.

**Reference**:

1. Bevans ([2022, November 11](http://127.0.0.1:55106/rmd_output/3/#ref-R-Career));Datanovia ([2019, December 26](http://127.0.0.1:55106/rmd_output/3/#ref-R-Action));Brian Ripley ([October 12, 2022](http://127.0.0.1:55106/rmd_output/3/#ref-R-Cran));*RPubs - How Do i Get p-Values and Critical Values from r?* ([2017, March 1](http://127.0.0.1:55106/rmd_output/3/#ref-R-Material1));Investopedia ([2022, August 31](http://127.0.0.1:55106/rmd_output/3/#ref-R-Material2));*S.3.2 Hypothesis Testing (p-Value Approach) | STAT ONLINE. (N.d.). PennState: Statistics Online Courses.* ([n.d.](http://127.0.0.1:55106/rmd_output/3/#ref-R-Material3));*8. Calculating p Values — r Tutorial.* ([n.d.](http://127.0.0.1:55106/rmd_output/3/#ref-R-Material4))
2. *8. Calculating p Values — r Tutorial.* n.d. <https://www.cyclismo.org/tutorial/R/pValues.html>.
3. Bevans, R. 2022, November 11. *Hypothesis Testing | a Step-by-Step Guide with Easy Examples*. <https://www.scribbr.com/statistics/hypothesis-testing/>.
4. Brian Ripley, Douglas M. Bates, Bill Venables. October 12, 2022. *Support Functions and Datasets for Venables and Ripley’s MASS*. <https://cran.r-project.org/web/packages/MASS/MASS.pdf>.
5. Datanovia. 2019, December 26. *How to Do a t-Test in r: Calculation and Reporting*. <https://www.datanovia.com/en/lessons/how-to-do-a-t-test-in-r-calculation-and-reporting/>.
6. Investopedia. 2022, August 31. *What Is a Confidence Interval and How Do You Calculate It?*<https://www.investopedia.com/terms/c/confidenceinterval.asp>.
7. *RPubs - How Do i Get p-Values and Critical Values from r?* 2017, March 1. <https://rpubs.com/mdlama/spring2017-lab6supp1>.
8. *S.3.2 Hypothesis Testing (p-Value Approach) | STAT ONLINE. (N.d.). PennState: Statistics Online Courses.* n.d. <https://online.stat.psu.edu/statprogram/reviews/statistical-concepts/hypothesis-testing/p-value-approach>.