**Module 4**

**Week 4: R Practice**

**Logo

Description automatically generated**

**ALY6010**

**Instructor: Prof. Amin Karimpour**

**Date: 27 Nov 2021**

**Submitted by: -**

**Jeseeka Shah**

**Nuid – 002134289**

**Introduction**

The dataset is provided which has inbuild dataset in the mass library known as cats which has 144 rows and 3 features for performing the hypothesis testing. Along, with this there is another dataset given which has the measure of sleep quality which is impacted by medication or not needs to be tested.

**Exploratory data analytics**

**Part A**

The bar graph (Fig 1) shows the body weight of the female and the male cats. It is clear from the graph that have a lighter body weight compared to male cats. Also, the maximum number of female cats have weight that is near 2.3 and maximum number of male cats have it near 3. Fig 2 shows the height of female and male cats. Female cats have height range between 6 and 11. For male cats the maximum height frequently seen is between 10 and 15 as per the data given.

Chart, bar chart, histogram

Description automatically generated Chart, histogram

Description automatically generatedFig 1: Count of weight of female & male cats Fig 2: the different heights of female and male cats

|  |  |  |  |
| --- | --- | --- | --- |
| Mean: Female Weight | Mean: Male Weight | Mean: Female height | Mean: Male height |
| **2.36** | **2.9** | **9.2** | **11,32** |

Figure 3 shows the relation between height and weight in cats. The graph shows that there is a linear relationship between the height and weight of the cats.

Chart, scatter chart

Description automatically generated

Figure 3: linear relation between height and weight

**Performing Hypothesis testing**

Here, looking at the cat’s weight it’s clear that it’s not dependent on each other based on the gender. Hence, they are independent in nature.

**Question 1:**

**Do male and female cats have same weights?**

The weights of the male and female are not same. The boxplot (figure 4) shows the median weight of female and male cats respectively. And the red dot shows the mean weight in both boxplots which are different.

Chart, box and whisker chart

Description automatically generated

Figure 4: Boxplot of weight based on gender

Using test to check if both the weights are same or not. As there are two sample that are to be compared. Here, two sample test will be considered. The number of female cats is 47 (n1=47) and male cats are 97 (n2=97). Both the values are normally distributed and are larger than 30.

Chart

Description automatically generated

Figure 5: distribution of weight of cats

**Null Hypothesis**: Ho: Male and Female cats have same body weight



**Alternative Hypothesis**: H1: Male and Female cats don’t have same body weight



Text

Description automatically generated

**Answer 1:**

The two-sample test shows that p value is very small than significant level (known as alpha) which is 0.05 (p<0.05). Hence, the null hypothesis can be **rejected** and can concluded that female and male cats’ weights are different from one another.

**Part B: Checking if meditation has effect on quality of sleep**

Here, the data set is dependent on each other or also can be called as paired dataset. The dataset has the data which is recorded before and after meditation to check its effect on sleep or not.

Given sample: -

**before\_meditation**<- c (4.6, 7.8, 9.1, 5.6, 6.9, 8.5, 5.3, 7.1, 3.2, 4.4)

**after\_meditation** <- c (6.6, 7.7, 9.0, 6.2, 7.8, 8.3, 5.9, 6.5, 5.8, 4.9)

**Meditationeffectsleep**<-cbind (before\_meditation, after\_meditation)

**Meditationeffectsleep**<-as.data.frame (Meditationeffectsleep)

After making a data frame, as the data set is dependent on one another. I choose to perform paired t-test to perform the hypothesis testing. Prior to performing the testing, I have plotted the boxplot to see the effect on the sleep quality with respective to the meditation in figure 6. The before meditation has median less than that of after meditation. And it can be clearly seen though the boxplot showing the given data that sleep quality is better after meditation than that of before meditation.

Chart, box and whisker chart

Description automatically generated

Figure 6: Meditation effect on sleep quality

**Question 2:**

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

**Null Hypothesis**: Ho: Meditation has no effect on the sleep quality (sleep quality before medication> sleep quality after medication) **Alternative Hypothesis:** H1: Meditation has effect on the sleep quality (sleep quality before medication <= sleep quality after medication)

**Case 1**: Where confidence interval is **0.95** where **alpha =0.05**

Taking **paired t-test** as the data is dependent by setting paired=T (which is true) in t. test ().

Graphical user interface, text, application, email

Description automatically generated

**Case 2:** Where confidence interval is **0.90** where alpha=0.1

In general, the confidence interval is considered as 0.95. By setting confidence interval as 0.90, alpha will be set to 0.1 as the relationship between **confidence level + alpha =1.**

Graphical user interface, text, email

Description automatically generated

**Answer 2:**

|  |  |  |
| --- | --- | --- |
|  | **P value** | **Result** |
| **Case 1: alpha =0.05 & c=0.95** | **P=0.04161, p<alpha** | **Rejecting the null hypothesis**. Which means, sleep quality before meditation is not good compared to after. Thus, proved that researcher’s claim that meditation improves the sleep quality is **TRUE.** |
| **Case 2: alpha=0.1 & c=0.90** | **P=0.0832, p< alpha** | **Reject the null hypothesis.**  Similarly, even here the researcher’s claims come true that sleep quality has improved after meditation. |

**Summary**

* The task given here has two dataset one of cat’s weight based on their gender and second dataset has the before and after sleep quantity recorded based on meditation performed.
* Firstly, I carried out all the exploratory data analysis to understand the data that had to be process via hypothesis testing. Cats weight dataset was independent of its values and meditation dataset had dependent values in the dataset.
* This inference helped me decide which test must be performed by deciding the null and alternative hypothesis for respective problem statement.
* Null hypothesis in case of cat dataset was mean difference between female cats’ weight and males cat weight is zero. And vice-versa for alternative hypothesis for other.
* After performing, t.test() on cat dataset the inference followed that p values is less than 0.05. Hence, the null hypothesis gets rejected and concluded that female and male cats weights are different from one another.
* Next, I carried out analysis of meditation effect on the sleep quality based on the dataset records of sleep before and after meditation. As this data is dependent on one another , I have chosen to perform paired t-test.
* Results pointed out that sleep quality meditation before is not as good as the sleep quality after meditation. The p value is less than both 0.05 and 0.01 confidence intervals which indicated that null hypothesis can be rejected (before meditation sleep quality >after meditation sleep quality)

**Reference**

1)Axis labels not showing up. (2016, March 8). Stack Overflow.

<https://stackoverflow.com/questions/35876938/axis-labels-not-showing-up>

2)Data Cleanup: Remove NA rows in R. (2020, December 17). ProgrammingR.

<https://www.programmingr.com/examples/remove-na-rows-in-r/>

3) Wetherill, C. (n.d.). *How to Perform T-tests in R | DataScience+*. Datascience. <https://datascienceplus.com/t-tests/>

4) *The Open Educator - 9. Two Sample T-Test Unequal Variance*. (n.d.). Open Educator. <https://www.theopeneducator.com/doe/hypothesis-Testing-Inferential-Statistics-Analysis-of-Variance-ANOVA/Two-Sample-T-Test-Unequal-Variance>