**R PRACTICE REPORT**

**PROBABILITY AND STATISTICS ASSIGNMENT**

**WEEK 4**

**MODULE 4**

By : **HARSHIT GAUR**

MASTER OF PROFESSIONAL STUDIES IN ANALYTICS

ALY 6010 : PROBABILITY THEORY AND INTRODUCTORY STATISTICS

DECEMBER 2, 2021

To : **PROF. AMIN KARIMPOUR**

# **ABSTRACT**

**Data analytics** is a discipline focused on extracting insights from data. It comprises of the processes, tools and techniques of data analysis and management, including the collection, organization, and storage of data. The chief aim of data analytics is to apply statistical analysis and technologies on data in order to find trends and solve problems.

**Hypothesis testing** is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis. Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data.

# **INTRODUCTION**

In hypothesis testing, an analyst tests a statistical sample, with the goal of providing evidence on the plausibility of the null hypothesis. Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analysed. All analysts use a random population sample to test two different hypotheses: the null hypothesis and the alternative hypothesis. The null hypothesis is usually a hypothesis of equality between population parameters; e.g., a null hypothesis may state that the population mean return is equal to zero. The alternative hypothesis is effectively the opposite of a null hypothesis (e.g., the population mean return is not equal to zero). Thus, they are mutually exclusive, and only one can be true. However, one of the two hypotheses will always be true.

We used the library of 'MASS' for data set of 'cats' and performed hypothesis testing on another data set provided in the assignment problem statement.   
  
 All Hypothesis testing are performed using 4 steps :

1. The first step is for the analyst to state the two hypotheses so that only one can be right.
2. The next step is to formulate an analysis plan, which outlines how the data will be evaluated.
3. The third step is to carry out the plan and physically analyse the sample data.
4. The fourth and final step is to analyse the results and either reject the null hypothesis, or state that the null hypothesis is plausible, given the data.

We're going to use 2 types of hypothesis testing in this assignment.

* Two-Sample t-Test
* Paired t-Test

The features available in the 'cats' data set are -

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Feature** | **Dictionary** |
| 1. | Sex | Gender of the cats (M,F) |
| 2. | Bwt | Body weight of the cats (in kilograms) |
| 3. | Hwt | Heart weight of the cats (in grams) |

*Table 1: Features of the 'cats' data set with their dictionary.*

From the structure of the data set, the features, their types, and values can be determined.

|  |
| --- |
| *Figure 1: Structure of the data set.* |

Some of the data points from the summary of the data set are present below.

|  |
| --- |
| *Figure 2: Summary of the data set.* |

# 

# **EXPLORATORY DATA ANALYSIS**

The descriptive statistics of the features of the data set can be summarised to calculate the statistics -

|  |
| --- |
| *Figure 3: Summary of the 'cat' data set.* |

We used the **'DESCRIBE()**' function from the package **'PSYCH'** to find out the descriptive statistics of features of the data set. The following observations can be made using the statistics found in summary of the data set -

1. The total number of observations in the 'cats' data set is 144.
2. The **mean** of the attribute *Bwt (Body weight in Kilograms)* is around **2.72** with a **standard deviation** of **0.49** and **quartiles value (Lower Quartile - 2.30, Higher Quartile - 3.025).**   
   From the observations, we can calculate that the data points has the **maximum value (3.90)** and the **minimum value (2.0)** in this feature.
3. The **mean** of the attribute *Hwt (Heart weight in grams)* is around **10.63** with a **standard deviation** of **2.43** and **quartiles value (Lower Quartile - 3.18, Higher Quartile - 12.13).**   
   From the observations, we can calculate that the data points has the **maximum value (20.5)** and the **minimum value (6.3)** in this feature.
4. The attribute *Sex\** is a categorical feature with values as M/F.

# Part 1 - Two Sample t-Test with unequal variance

For comparison of the body weights of male and female cats, the following steps should be performed.

1. The data set needs to be filtered based on the *Sex\** feature and created as two vectors.
2. Check both the vectors (groups) for normality using ways like
   1. Density Plots
   2. Q-Q Plots
   3. Shapiro Wilks Test
3. Perform t-Test to compare both the groups.

# FILTER THE DATA SET BASED ON GENDER (SEX\*)

The data set is filtered based on gender (Sex\*) and 2 new vectors are created using the feature *Bwt (Bodyweight).*

|  |
| --- |
| *Figure 4: Filter the data set based on Gender (M & F) and vectorize on the Bwt feature.* |
| *Figure 5: Code for Summary of the r-binded male & female data set.* |
| *Figure 6: Summary of the r-binded male & female data set.* |

The descriptive statistics of the body-weight data set of male & female cats tell us that the mean and variance of both the groups are different. Therefore, we would need Two-Sample t-Test to perform hypothesis testing in this case.

# NORMALITY (Density Plots)

The normality of both the data sets was checked to understand the type of distribution -

|  |
| --- |
| *Chart, line chart  Description automatically generated*  *Figure 7: Density Plot of male & female cats data set.* |

The density plots of both the weight groups look approximately normal and requires no other transformation.

# NORMALITY (shapiro-wilks test & q-q plot)

The normality of both the male & female data set was checked using Shapiro-Wilks & Q-Q Plots to understand the type of distribution -

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | **S. No.** | **Data Set** | **Shapiro Wilks Test (P-value)** | **W-value** | | 1. | Body-weight group of Male Cats | 0.118962 | 0.97883 | | 2. | Body-weight group of Female Cats | 0.000375422 | 0.89096 |   *Table 2: Shapiro-Wilks Test of the data set.* |
| *Figure 8: Q-Q Plots of both the data sets.* |

The Shapiro-Wilks Test, and the Quantiles-Quantiles plots of the features gave out the below results about the data set -

1. **Shapiro-Wilks Test** was used to verify the normality of a feature of the data set. The test gives out a W-value and a p-value as its result. Higher the W-value and greater p-value will usually signify that the distribution tends to be normal and the NULL hypothesis can be rejected (the data is not normal). But, this is not a universal fact to abide by and should be taken with a pinch of grain.
2. The ***body-weight data set of male cats*** has a p-value (**0.118962**) which is greater than 5% and a W-value (**0.97883**) which is large. Therefore, this data set tends to be normally distributed.
3. The ***body-weight data set of female cats*** has a p-value (**0.000375422**) which is less than 5% and a W-value (**0.97883**) which is large. Since, the p-value conditions is not satisfied and the fact that the data set has less number of observations, we cannot confirm the normality of this female data set using Shapiro-Wilks test.
4. The **Quantiles-Quantiles plot** are used to check if the sample data came from the same theoretical distribution, like Normal or Gaussian distribution. If the theoretical and sample quantile fall in a straight line, the attribute (feature or data set) tends to follow the theoretical distribution (ex, Normal distribution).
5. The ***body-weight data set of male cats*** tends to follow the straight line between theoretical and sample quantiles. It is considered normally distributed.
6. But, the ***body-weight data set of female cats*** do not follow a straight line between theoretical and sample quantiles very closely. Therefore, it is a possibility that it can be considered either to be normally distributed or not.

**Two Sample T-Tests**

Since, both the data sets come from the same population and that we do not know the standard deviation, we need to perform Two-Sided / Two-Sample t-Test as statistical hypothesis testing to compare both the data sets or groups.

# Bodyweight of Male Cats vs Bodyweight of Female Cats (TWO-SIDED)

|  |
| --- |
| *Figure 9: Two Sample t-Test of Male vs Female cats for Bodyweight feature (Two-Sided)* |
| *Figure 10: Result of Two Sample t-Test of Male vs Female cats for Bodyweight feature* |
| *Figure 11: P-Value of Two Sample t-Test of Male vs Female cats for Bodyweight feature* |

**NULL HYPOTHESIS, H0** : True Difference in Means of Bodyweight of both Cats Populations (Male & Female) is equal to 0.

**ALTERNATE HYPOTHESIS, H1** : True Difference in Means of Bodyweight of both Cats Populations (Male & Female) is not equal to 0.

**ALTERNATIVE** : Two. Sided

**DEGREE OF FREEDOM** : 137

**P-VALUE** : 0.000000000000008831034

**Observation** : Since, the p-value of our Two Sample t-test is 0.000000000000008831034, which is **much less than alpha = 0.05, we can reject the Null Hypothesis of the test.** The **T.TEST()** function in R takes the confidence interval as 95% (0.95) by default.

This means that we do have sufficient evidence to say that **True Difference in Means of Bodyweight of Both Population** of the *Male Cats & Female Cats* is not equal to 0.

# Bodyweight of Male Cats vs Bodyweight of Female Cats (GREATER)

|  |
| --- |
| *Figure 12: Two Sample t-Test of Male vs Female cats for Bodyweight feature (Greater)* |
| *Figure 13: Result of Two Sample t-Test of Male vs Female cats for Bodyweight feature* |
| *Figure 14: P-Value of Two Sample t-Test of Male vs Female cats for Bodyweight feature* |

**NULL HYPOTHESIS, H0** : True Difference in Means of Bodyweight of both Cats Populations (Male & Female) is equal to 0.

**ALTERNATE HYPOTHESIS, H1** : True Difference in Means of Bodyweight of both Cats Populations (Male & Female) is greater than 0.

**ALTERNATIVE** : Greater

**DEGREE OF FREEDOM** : 137

**P-VALUE** : 0.000000000000004415517

**Observation** : Since, the p-value of our Two Sample t-test is 0.000000000000004415517, which is **much less than alpha = 0.05, we can reject the Null Hypothesis of the test.** The **T.TEST()** function in R takes the confidence interval as 95% (0.95) by default.

This means that we do have sufficient evidence to say that **True Difference in Means of Bodyweight of Male Cats Population and Female Cats Population** is greater than **0.**

# Part 2 - Effect of Meditation on sleep quality

For comparison of the body weights of male and female cats, the following steps should be performed. For conducting the experiment of assessing the effect of meditation on sleep quality, 10 students were recruited for a meditation workshop. These 10 students agreed to wear sleeping evaluators to measure their sleeping quality (which is on a scale of 0-10, the higher the better).

The pre- and post- meditation workshop sleep scores were obtained for these 10 students and a statistical test was applied on them to comprehend whether meditation improved the sleep quality or not.

|  |
| --- |
| *Figure 15: Vector for Pre-Meditation Average Sleeping Score.* |
| *Figure 16: Vector for Post-Meditation Average Sleeping Score.* |
| *Figure 17: Data frame for Pre- and Post-Meditation Average Sleeping Score.* |

The vectors or data frame were created for pre- and post- meditation data and appended with each other in a data frame with proper naming conventions.

**Descriptive Statistics of Cumulated Sleeping Scores of Pre and Post Meditation Workshop**

|  |
| --- |
| *Figure 18: Group by and Summarizing descriptive statistics for Pre and Post--Meditation Average Sleeping Score.* |
| *Figure 19: Descriptive Statistics for Pre- and Post-Meditation Average Sleeping Score.* |

# VISUALISATION OF THE DISTRIBUTION USING BOXPLOT

Box plot can help in understanding the distribution of the data set and comprehend the difference between the pre and post meditation average sleep scores (quality).

|  |
| --- |
| *Figure 20: Box Plot of both the pre and post meditation data sets.* |

The figure depicts that the pre-meditation workshop sample data has a larger variance with lower mean and median values than the post-meditation workshop sample. The point within the boxplot represents the mean value of and the middle horizontal line signifies the median value of the data sets.

# NORMALITY (Density Plots)

The normality of the difference in means of average sleeping scores of both the data sets (pre and post meditation) was checked to understand the type of distribution -

|  |
| --- |
| *Figure 21: Density Plot of difference in means of both data set.* |

The density plots of the difference in means of both the pre and post meditation groups look approximately normal and requires no other transformation.

# NORMALITY (shapiro-wilks test & q-q plot)

The normality of difference in means of average sleeping scores of both the data sets (pre and post meditation was checked using Shapiro-Wilks & Q-Q Plots to understand the type of distribution -

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | **S. No.** | **Data Set** | **Shapiro Wilks Test (P-value)** | **W-value** | | 1. | Difference in Means of sleeping score | 0.2177084 | 0.89975 |   *Table 3: Shapiro-Wilks Test of the data set.* |
| *Figure 22: Q-Q Plot of difference in means of average sleeping score of both the data sets.* |

The Shapiro-Wilks Test, and the Quantiles-Quantiles plots of the features gave out the below results about the data set -

1. The **Shapiro-Wilks Test** for ***difference in means of average sleeping score of both the pre and post meditation data set*** has a p-value (**0.2177084**) which is greater than 5% and a W-value (**0.89975**) which is large. Therefore, this data set tends to be normally distributed.
2. The **Quantiles-Quantiles plot** for ***difference in means of average sleeping score of both the pre and post meditation data set*** tends to follow the straight line between theoretical and sample quantiles. It can be considered normally distributed.

**Paired Sample T-Tests**

Since, both the data sets come from the same population and that it contains the paired pre- and post- values of an experiment, we need to perform a Paired t-Test as statistical hypothesis testing to compare both the data sets or groups.

# Pre Meditation Sleep Scores vs Post Meditation Sleep Scores

|  |
| --- |
| *Figure 23: Paired t-Test of Pre vs Post Meditation Average Sleeping Scores.* |
| *Figure 24: Result of Paired t-Test of Pre vs Post Meditation Average Sleeping Scores.* |
| *Figure 25: P-Value of Paired t-Test of Pre vs Post Meditation Average Sleeping Scores.* |

**NULL HYPOTHESIS, H0** : True Difference in Means of Pre-Meditation and Post-Meditation average sleeping scores is equal to 0.

**ALTERNATE HYPOTHESIS, H1** : True Difference in Means of Pre-Meditation and Post-Meditation average sleeping scores is greater than 0.

**ALTERNATIVE** : Paired (Greater)

**DEGREE OF FREEDOM** : 9

**P-VALUE** : 0.04161026

**Observation** : Since, the p-value of our Two Sample t-test is 0.04161026, which is **less than alpha = 0.05, we can reject the Null Hypothesis of the test.** The **T.TEST()** function in R takes the confidence interval as 95% (0.95) by default. The *paired = TRUE* signifies that the t-Test is a Paired t-Test.

This means that we do have sufficient evidence to say that **True Difference in Means of Pre-Meditation sleeping score and Post-Meditation average sleeping score** is greater than 0.

**NOTE :** Since, the p-value of t-Test at the alpha-value of 0.05 (95%) is 0.04161026, we have rejected the Null Hypothesis of the test. ***If we take alpha-value of 0.1 (90%), the p-value would be far less than this alpha-value (0.1) and the Null hypothesis would get rejected in this case as well.***

# **CONCLUSION**

The dataset of *'cats'* from the library *'MASS'* has provided various insights about the body weights and heart weights of the cats based on their gender; and the second case study of Pre-Meditation and Post-Meditation sleeping quality of 10 people also studied and analysed. We performed exploratory data analysis, calculated various statistics, plotted a few visualisation graphs in order to understand the analysis properly, and performed normality tests and various hypothesis tests. The below points can be inferred from the analysis :

* We used density plots to determine the normality of the bodyweight distribution in male and female cats population visually and found out that we can anticipate male data set to be normally distributed very closely and the female cats tending to be normally distributed.
* The Shapiro-Wilks Tests and Q-Q plots provided better information about the normality of the features. **Both the data sets in 1st Part and data set belonging to 2nd Part** can be considered to be normally distributed when used in original format as well.
* The application of Two Sample t-test on the male and female cats data set based on *Bwt (Bodyweight)* helped in rejecting the null hypothesis and provided the better understanding of the true populations.
  + **True Difference in Means of Bodyweight of Male Population and Female Population** is equal to 0 with **alternate being not equal to 0**, p-value :0.000000000000008831034
  + **True Difference in Means of Bodyweight of Male Population and Female Population** is equal to 0 with **alternate being greater than 0**, p-value :0.000000000000004415517
* We plotted a boxplot to visually and understood that the pre-meditation workshop sample data has a larger variance with lower mean and median values than the post-meditation workshop sample.
* The **Shapiro-Wilks Test** for ***difference in means of average sleeping score of both the pre and post meditation data set*** has a p-value (**0.2177084**) which is greater than 5% and a W-value (**0.89975**) which is large; and the Q-Q Plot also tends to follow a straight line between theoretical and sample quantiles. Therefore, this data set tends to be normally distributed.
* We applied Paired-Sample t-Test on the meditation data set to observe that the null hypothesis can be rejected and that the alternate hypothesis of **True Difference in Means of Pre-Meditation sleeping score and Post-Meditation average sleeping score** being greater than 0 can be considered with p-value :0.04161026
* Since, the p-value of Paired t-Test at the alpha-value of 0.05 (95%) is 0.04161026, we have rejected the Null Hypothesis of the test. ***If we take alpha-value of 0.1 (90%), the p-value would be far less than this alpha-value (0.1) and the Null hypothesis would get rejected in this case as well.***

# **BIBLIOGRAPHY**

1. *R Companion: Paired t–test*. (2021). Paired T-Test | RCompanion. https://rcompanion.org/rcompanion/d\_09.html
2. *How Hypothesis Testing Works*. (2021, October 28). Investopedia. https://www.investopedia.com/terms/h/hypothesistesting.asp
3. *Normality Test in R - Easy Guides - Wiki - STHDA*. (2021). STHDA. http://www.sthda.com/english/wiki/normality-test-in-r
4. *A Grammar of Data Manipulation*. (2021). Dplyr. https://dplyr.tidyverse.org/
5. *Home - RDocumentation*. (2021). Functions in R - Documentation. https://www.rdocumentation.org/
6. *Chapter 5 Probability and Statistics in R | Applied Statistics with R*. (2021, July 23). Probability and Statistics in R. https://daviddalpiaz.github.io/appliedstats/probability-and-statistics-in-r.html
7. *assign row name while rbind a row in a data frame*. (2017, February 1). Stack Overflow. https://stackoverflow.com/questions/41982549/assign-row-name-while-rbind-a-row-in-a-data-frame

# **APPENDIX**

#---------------------- Week\_4\_Module\_4\_R-Script ----------------------#

print("Author : Harshit Gaur")

print("Week 4 Assignment - Module 4 R Pratice")

# Importing the packages.

listOfPackages <- c(

"MASS",

"dplyr", "tidyr", "plyr", "tidyverse", "RColorBrewer", "plotrix", "scales", "ggplot2",

"data.table", "reshape", "gridExtra", "vtable", "moments", "ggpubr", "psych", "GGally"

)

for (package in listOfPackages) {

if (package %in% rownames(installed.packages()) == FALSE)

{ install.packages(package) }

# Importing the package.

library(package, character.only = TRUE)

}

# Use & Display the 'cats' data set.

data(cats, package = "MASS")

View(cats)

# Print the structure of 'cats' data set

str(cats)

# Print the summary of 'cats' data set

summary(cats)

st(cats)

# Describe the summary of the data set by providing Descriptive Statistics.

View(describe(cats, skew = FALSE, quant = c(0.25, 0.75), IQR = TRUE))

#------------------- 1st Question -------------------#

# Filter the data set based on 'Sex (M/F)' and Make it a vector using 'Body weight (Bwt)' feature

male\_cats\_Bwt <- cats %>% filter(Sex == "M") %>% pull(Bwt)

female\_cats\_Bwt <- cats %>% filter(Sex == "F") %>% pull(Bwt)

# Describe the summary of both the Gender based vectors by providing Descriptive Statistics.

View(rbind

("Male" = describe(male\_cats\_Bwt, skew = FALSE, quant = c(0.25, 0.75), IQR = TRUE),

"Female" = describe(female\_cats\_Bwt, skew = FALSE, quant = c(0.25, 0.75), IQR = TRUE)))

#------------------- Exploratory Data Analysis -------------------#

# Check Normality using Density Graphs of all the uni-variates.

normality\_male\_bwt <- ggdensity(male\_cats\_Bwt, main = "Density plot of Male Cats' Bodyweight", xlab = "Male Cats' Bodyweight", fill = "#ffa514") +

stat\_overlay\_normal\_density(color = 'red', linetype = 'dashed')

normality\_female\_bwt <- ggdensity(female\_cats\_Bwt, main = "Density plot of Female Cats' Bodyweight", xlab = "Female Cats' Bodyweight", fill = "#edf759") +

stat\_overlay\_normal\_density(color = 'red', linetype = 'dashed')

grid.arrange(normality\_male\_bwt,normality\_female\_bwt)

# Check Normality using Shapiro-Wilks Test

format(shapiro.test(male\_cats\_Bwt)$p.value, scientific = FALSE)

format(shapiro.test(female\_cats\_Bwt)$p.value, scientific = FALSE)

#-------- Check Normality using Q-Q Plot of all the numeric features. --------#

# Function to plot graph

qq\_plot <- function(numeric\_feature, mainTitle) {

qqnorm(numeric\_feature, pch = 5, frame = TRUE, main = mainTitle)

qqline(numeric\_feature, col = "#ffa514", lwd = 2)

}

# Changing Plot Matrix Size to 1x2.

par(mfrow = c(2,1))

# Check Normality using Q-Q Plot of 'male\_cats\_Bwt' Features.

qq\_plot(male\_cats\_Bwt, "Male Cats' Bodyweight Q-Q Plot")

# Check Normality using Q-Q Plot of 'female\_cats\_Bwt' Features.

qq\_plot(female\_cats\_Bwt, "Female Cats' Bodyweight Q-Q Plot")

# Resetting Plot Matrix Size to 1x1.

par(mfrow = c(1,1))

# Check Skewness of the features

skewness(male\_cats\_Bwt)

skewness(female\_cats\_Bwt)

# Check Kurtosis of the features

kurtosis(male\_cats\_Bwt)

kurtosis(female\_cats\_Bwt)

# ----------- Two Sample Testing (t-Test) ----------- #

# Two-Sample t-test for 'Bodyweight of Male Cats vs Bodyweight of Female Cats'

ttest <- t.test(male\_cats\_Bwt, female\_cats\_Bwt, alternative = "two.sided")

format(ttest$p.value, scientific = FALSE)

# ----------- One Sample Testing (t-Test) ----------- #

# One-Sample t-test for 'Bodyweight of Male Cats vs Female Cats'

ttest <- t.test(male\_cats\_Bwt, female\_cats\_Bwt, alternative = "greater")

format(ttest$p.value, scientific = FALSE)

#------------------- 2nd Question -------------------#

# Average Sleeping Quality Scores BEFORE workshop (Pre-Meditation)

avg\_scores\_pre\_med <- c(4.6, 7.8, 9.1, 5.6, 6.9, 8.5, 5.3, 7.1, 3.2, 4.4)

# Average Sleeping Quality Scores AFTER workshop (Post-Meditation)

avg\_scores\_post\_med <- c(6.6, 7.7, 9.0, 6.2, 7.8, 8.3, 5.9, 6.5, 5.8, 4.9)

# Grouping of the Meditation data.

sleeping\_data <- data.frame(meditation\_status = rep(c("pre", "post"), each = 10),

sleep\_score = c(avg\_scores\_pre\_med, avg\_scores\_post\_med))

group\_by(sleeping\_data, meditation\_status) %>%

dplyr::summarise(

count = n(),

mean = mean(sleep\_score, na.rm = TRUE),

sd = sd(sleep\_score, na.rm = TRUE))

# Plot an Box Plot for 'Average Sleeping Score Pre and Post Meditation' using 'ggplot'

# ----------- Plot 1: Average Sleeping Score Pre vs Post Meditation ----------- #

par(mar = c(2,4,2,4))

ggplot(data = sleeping\_data, aes(x = meditation\_status, y = sleep\_score, color = meditation\_status)) +

geom\_boxplot(outlier.size = 3.5) +

stat\_summary(fun=mean, geom="point", shape=18, size=3, color="steelblue") +

labs(title = 'Box Plot of Average Sleeping Score Pre vs Post Meditation', x = 'Meditation Status', y = 'Average Sleeping Score') +

theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1, size = 10))

# Check Normality using Density Graphs.

# Computing the difference in between the sleeping average score of Pre- and Post- Meditation.

avg\_sleep\_score\_diff <- with(sleeping\_data, sleep\_score[meditation\_status == "pre"] - sleep\_score[meditation\_status == "post"])

par(mar = c(4,4,4,4))

ggdensity(avg\_sleep\_score\_diff, main = "Density Plot of Difference in Average Sleeping Score", xlab = "Difference in Average Sleeping Score of Pre & Post Meditation", fill = "#40c7f7") +

stat\_overlay\_normal\_density(color = 'red', linetype = 'dashed')

# Check Normality using Shapiro-Wilks Test

format(shapiro.test(avg\_sleep\_score\_diff)$p.value, scientific = FALSE)

par(mar = c(4,4,4,4))

# Check Normality using Q-Q Plot of 'male\_cats\_Bwt' Features.

qq\_plot(avg\_sleep\_score\_diff, "Q-Q Plot of Difference in Avg Sleeping Score of Pre- and Post- Meditation")

# ----------- Paired Testing (t-Test) ----------- #

# Paired t-test for 'Pre Meditation vs Post Meditation Sleep Score'

paired\_ttest <- t.test(data = sleeping\_data, sleep\_score ~ meditation\_status, paired = TRUE, alternative = "greater")

format(paired\_ttest$p.value, scientific = FALSE)