Brown Fat Report

Cicily Balachandar

860-877-4088

cicily.balachandar@uconn.edu

Abstract

This report will examine brown fat presence and brown fat volume in relation to certain variables of interest. In particular, the season when the patient had the exam and whether the person had diabetes will be looked at.

After briefly describing the data and examining related graphs and tables, statistical inferences will be conducted. Both parametric and nonparametric methods will be used, where patients who had the exam in the winter will be compared to patients who had the exam in the summer. The means (parametric) and medians (nonparametric) of brown fat volume in those having brown fat will be compared. After checking for equal variances, using the F test for equality of variance (parametric) and equal dispersions, using the Ansari test (nonparametric), the corresponding t test and Satterthwaite test can be performed on the data.

Then, the association between diabetes and brown fat is also examined, by testing for significance of the odds ratio, where the odds ratio is the odds of having brown fat and not having diabetes to the odds of having brown fat and having diabetes.

The location parameters of the winter and summer groups were found to be significantly different, and the odds ratio was also found to be significantly different than 1, with a significance level of 5%.

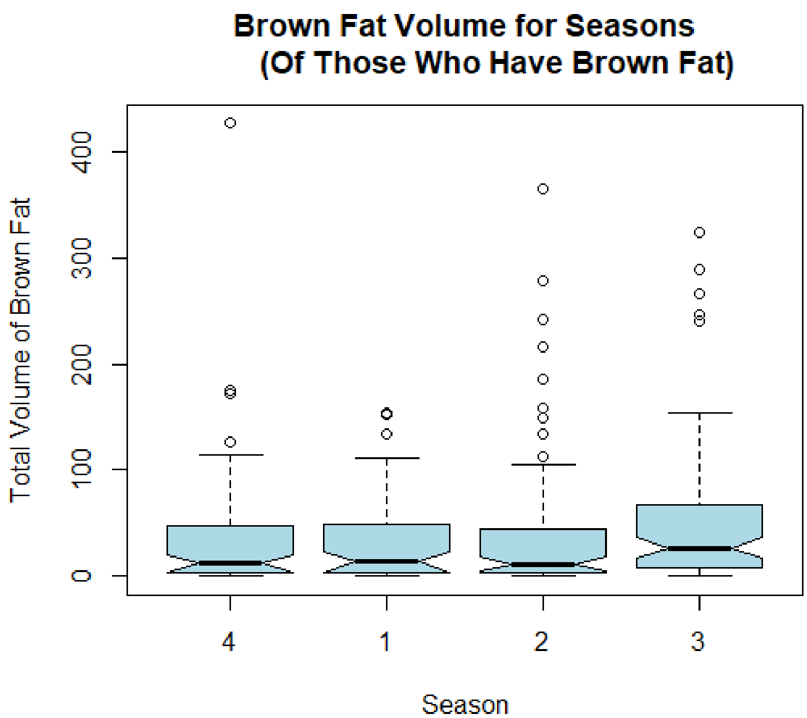
Section 1. Introduction and Description of the Data

The research questions include whether the location parameter for total volume of brown fat is greater in the winter season compared to the location parameter for total volume of brown fat in the summer season, among those who have brown fat, and also, if there is an association between brown fat and diabetes.

The Brown Fat Dataset is comprised of 4473 patients sorted based on 24 variables, both quantitative and qualitative. Quantitative variables include: *sex of the patient, month when the exam occurred, whether they have diabetes, season, whether they have brown fat, cancer status,* and *cancer type*. The qualitative variables are: *age in years when the exam occurred, day of the year the exam occurred, external temperature on the day of exam, average temperature of the last 2 days, average temperature of the last 3 days, average temperature of the last 7 days, average temperature of the last month, sunshine duration on exam day in minutes, weight (kg), size (cm), BMI (kg/m^2), blood sugar level (mmol/L), total volume of brown fat (mL), cervical volume (mL), paravertebral volume (mL), mediastinal volume (mL),* and *perirenal volume (mL)*.

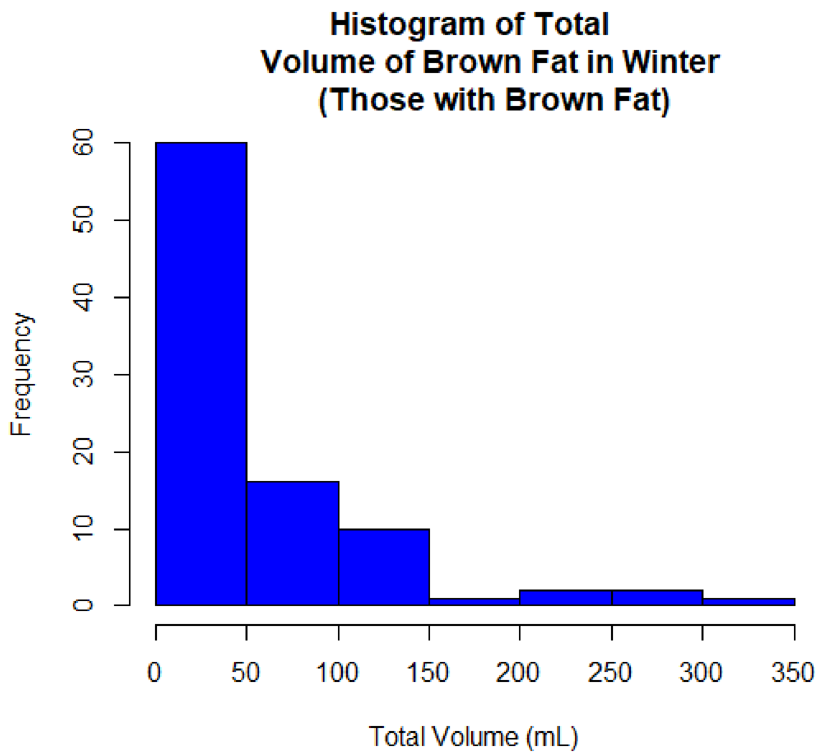
Section 2. Exploratory Data Analysis

Figure 1. Side-by-side notched boxplot of brown fat volume (of those who had brown fat) by season.



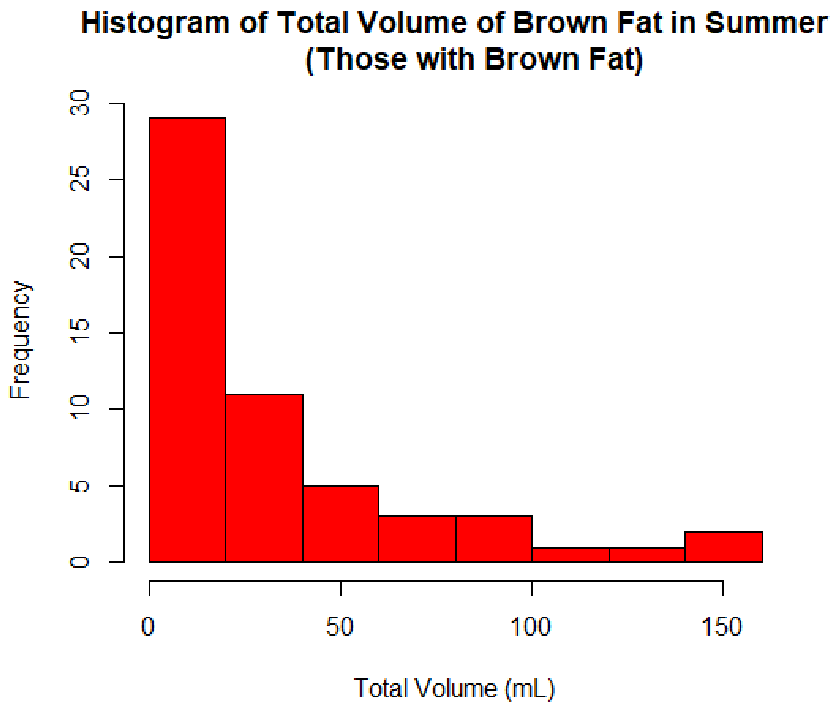
In the above **Figure 1**, the following seasons are shown for the total volume of brown fat in those who have brown fat, where 1 corresponds to spring, 2 to summer, 3 to fall, and 4 to winter. Since all of the notches are overlapping, no conclusions can be made between the different seasons. Instead looking at the individual seasons, the most notable feature is that each distribution is positively skewed.

Figure 2. Histogram of volume of brown fat of those with brown fat, who had the exam in winter.



In **Figure 2**, the distribution of the data is skewed right, and thus not normally distributed. It can also be seen that the distribution of volume of brown fat in summer for those with brown fat is also skewed right in **Figure 3** below.

Figure 3. Histogram of volume of brown fat of those with brown fat, who had the exam in summer.



Then, the Box Cox Transformation was applied to the data for each group. From **Figure 4** and **Figure 5** below, the transformed data for the winter group is approximately normal, but it is not clear whether the transformed data for the summer group is approximately normal. Looking at the Shapiro-Wilk normality test, the resulting p-values of 0.7547 and 0.1057, for the winter and summer groups, respectively, confirm that the transformed data for each of the groups is approximately normal at a 5% significance level.

Figure 4. Box Cox Transformation of data from winter group for brown fat volume in those with brown fat.

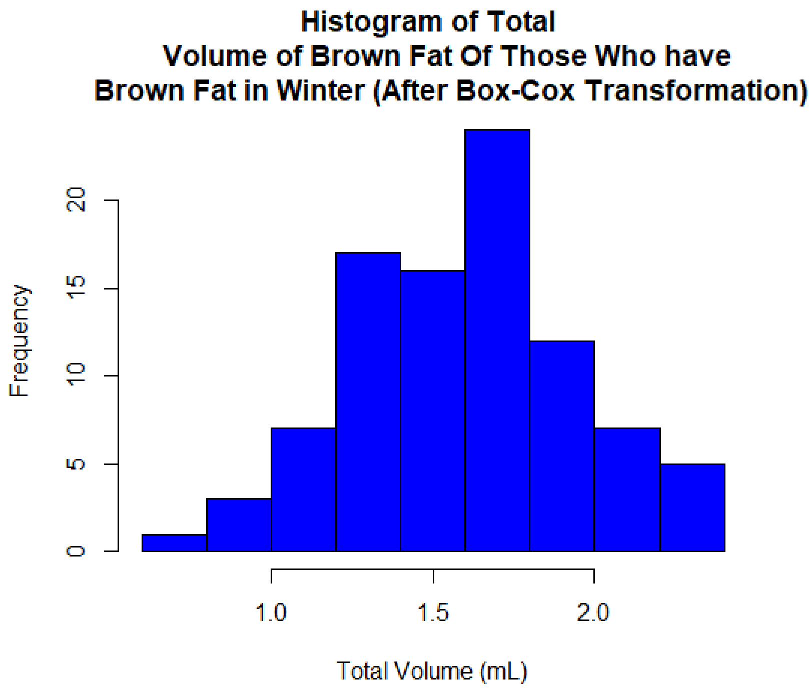
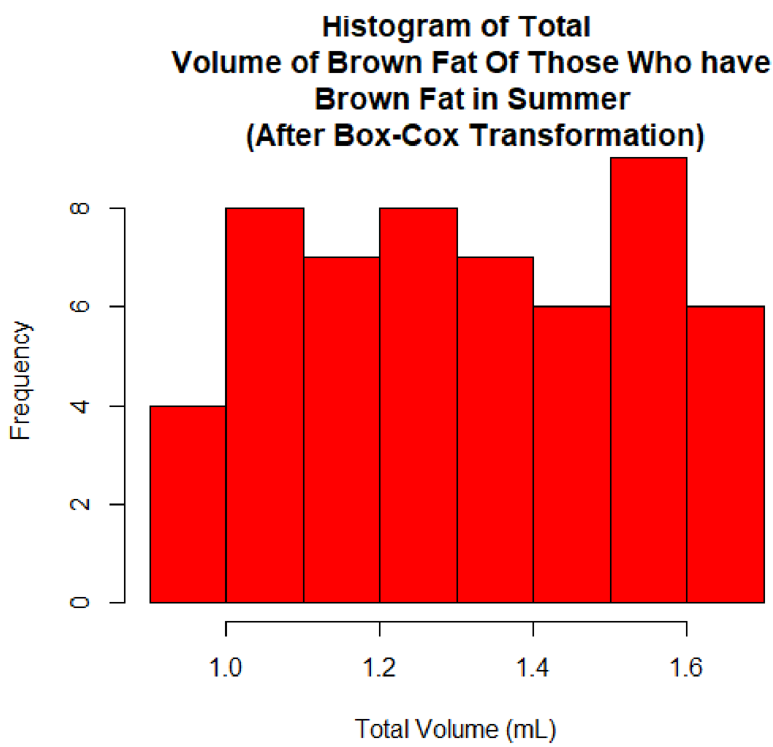


Figure 5. Box Cox Transformation of data from summer group for brown fat volume in those with brown fat.



The following two tables consolidate the data into a 2x2 table format, in order to test the association of diabetes and the presence of brown fat.

Table 1. Observed cell counts of diabetes and brown fat

	No Diabetes	Diabetes
Brown Fat	303	6
No Brown Fat	3695	469

Table 2. Expected cell counts of diabetes and brown fat

	No Diabetes	Diabetes
Brown Fat	276.1865	32.81355
No Brown Fat	3721.8135	442.18645

For **Table 1**, the observed counts of each combination of presence or absence of diabetes and presence or absence of brown fat are shown.

In **Table 2**, all expected cell counts are greater than 5 and multinomial sampling is assumed, so the Pearson chi-squared test of association can be performed.

Section 3. Statistical Inference

For the comparing the means of the winter and summer groups, the hypothesis test for two independent samples and normal populations will be used. Under the assumption that each patient is independent from each other, it follows that the winter and summer groups are independent.

Before testing the equality of the population means of brown fat volume, the test for equal variances is first conducted. A resulting p-value of 0.001501 is found, and at the 5% significance level, the null hypothesis of equal variances is rejected, since the p-value is less than 0.05. Thus, there is significant evidence to suggest that the population variances differ when the significance level is 0.05.

Since the population variances are unknown and unequal, the Satterthwaite test can be performed on the data. The null hypothesis is that population means of brown fat volume are equal between the two groups and the alternative hypothesis is that the population mean brown fat volume for the winter group is higher than the population mean brown fat volume for the summer group. After conducting this test, a p-value of 2.998e-09 is found, which is less than a 0.05 significance level. Thus, the null hypothesis is rejected and with a 5% significance level, there is sufficient evidence to conclude that the mean fat volume for the winter group is higher than the mean fat volume for the summer group, among those with brown fat.

The population medians can also be compared, with a null hypothesis of the population medians being equal and an alternative hypothesis of the winter group having a higher median of brown fat volume compared to the summer group.

First, the Ansari-Bradley procedure for dispersion of two samples is performed and the resulting p-value of 0.5706 is found. The p-value is greater than 0.05. Thus, the null hypothesis of equal dispersion is failed to reject at a significance level 0.05 and inference for the medians can continue under the assumption of equal dispersions.

Next, the Wilcoxon Rank Sum test can be performed, where the null hypothesis is equal population medians and the alternative hypothesis is the winter group having a higher population median in brown fat volume compared to the population median of the summer group. The resulting p-value for this test is 0.01669, so the null hypothesis is rejected at the 5% significance level. Thus, there is enough evidence to conclude that the population median for the winter group is higher than the population median for the summer group.

Moving to the inference of the association between diabetes and brown fat volume, the Pearson chi-squared test will be used to test the association. The null hypothesis is that the true odds ratio is equal to 1 and the alternative hypothesis is that the true odds ratio is not equal to one. The resulting p-value is 4.757e-07, which is less than 0.05. The null hypothesis is rejected at a 5% significance level and there is sufficient evidence to conclude that the odds ratio is not equal to one. In addition, the obtained odds ratio estimate is 6.409878. In this context, the odds of someone having brown fat and not having diabetes is 6.409878 times the odds of someone having brown fat and diabetes.

Section 4. Discussion of Results and Concluding Remarks

In conclusion, when using a 5% significance level, the population mean of brown fat volume, in those who took the exam in winter and with those having brown fat, was significantly higher than the population mean of brown fat volume, in those who took the exam in summer and with those having brown fat. Also, the same conclusion was reached with the population medians, where the population median of the winter group in brown fat volume was higher than the population median of the summer group.

In a different direction, there was found to be an association between diabetes and brown fat, also with a 5% significance level. Moreover, the odds of someone having brown fat and not having diabetes is 6.409878 times the odds of someone having brown fat and diabetes, suggesting an association in the presence of brown fat when diabetes is absent.

To take these conclusions further, regression would be a good technique to analyze these relationships. In particular, it would be useful to use a continuous regression model with the average temperature in the last month and the total volume of brown fat. Also, a categorical regression model can be used to model the relationship between diabetes and brown fat. In addition, other factors can then be included into the model, like BMI and blood sugar level.

Section 5. Appendix

```
brownfat <- read_excel("P:/BIST 5505/brownfat_projectdata.xlsx")


# Data Manipulation
diabetes <- brownfat$Diabetes
bf <- brownfat$BrownFat
brown <- brownfat$Total_vol[which(brownfat$BrownFat == "1")]

YY <- bf[which(bf == "1" & diabetes == "1")]
NN <- bf[which(bf == "0" & diabetes == "0")]
YN <- bf[which(bf == "1" & diabetes == "0")]
NY <- bf[which(bf == "0" & diabetes == "1")]

length(YY)
length(NN)
length(YN)
length(NY)

Ydiab <- diabetes[which(diabetes == "1")]
Ndiab <- diabetes[which(diabetes == "0")]

length(Ydiab)
```

```
length(Ndiab)

Ybf <- bf[which(bf == "1")]
Nbf <- bf[which(bf == "0")]

length(Ybf)
length(Nbf)

tot_vol <- brownfat$Total_vol
winter_fat <- tot_vol[which(brownfat$Season == "4")]
summer_fat <- tot_vol[which(brownfat$Season == "2")]

tot_adj <- tot_vol[which(tot_vol > 0)]
wint_adj <- winter_fat[which(winter_fat > 0)]
summ_adj <- summer_fat[which(summer_fat > 0)]

# Exploratory Data Analysis

hist(tot_vol, main="Histogram of Total Volume of Brown Fat",
    xlab="Total Volume (mL)", col="blue")
shapiro.test(tot_vol)

hist(winter_fat, main="Histogram of Total Volume of Brown Fat in Winter",
    xlab="Total Volume (mL)", col="purple")
shapiro.test(winter_fat)

hist(summer_fat, main="Histogram of Total Volume of Brown Fat in Summer",
    xlab="Total Volume (mL)", col="red")
```

```
shapiro.test(summer_fat)

qqnorm(winter_fat)
qqnorm(summer_fat)

hist(wint_adj, main="Histogram of Total
    Volume of Brown Fat in Winter
    (Those with Brown Fat)",
    xlab="Total Volume (mL)", col="blue")
shapiro.test(wint_adj)

hist(summ_adj, main="Histogram of Total Volume of Brown Fat in Summer
    (Those with Brown Fat)",
    xlab="Total Volume (mL)", col="red")
shapiro.test(summ_adj)

boxplot(brown ~ brownfat$Season[which(brownfat$BrownFat == "1")],
        names = unique(brownfat$Season[which(brownfat$BrownFat == "1")]),
        xlab="Season", ylab="Total Volume of Brown Fat",
        col="lightblue", notch=T, main="Brown Fat Volume for Seasons
        (Of Those Who Have Brown Fat)")

library(car)

Y = wint_adj
bc=powerTransform(Y) # default method is Box-Cox
YBC=Y^bc$lambda
hist(YBC, main="Histogram of Total
```

```
    Volume of Brown Fat Of Those Who have
    Brown Fat in Winter (After Box-Cox Transformation)",
    xlab="Total Volume (mL)", col="blue")
shapiro.test(YBC)
qqnorm(YBC)

X = summ_adj
bc=powerTransform(X) # default method is Box-Cox
XBC=X^bc$lambda
hist(XBC, main="Histogram of Total
    Volume of Brown Fat Of Those Who have
    Brown Fat in Summer
    (After Box-Cox Transformation)",
    xlab="Total Volume (mL)", col="red")
shapiro.test(XBC)
qqnorm(XBC)

# Statistical Inference

var.test(YBC, XBC, alternative="two.sided", conf.level=0.95)
t.test(YBC,XBC, alternative = "greater", var.equal=FALSE)



ansari.test(wint_adj, summ_adj, alternative= "two.sided")
wilcox.test(wint_adj, summ_adj, alternative = "greater")

fat = matrix(c(303, 3695, 6, 469), nrow = 2,
             dimnames = list(Diabetes = c("No", "Yes"),
```

```
                             BrownFat = c("Yes", "No")))

E11 <- 309*3998/4473
E21 <- 4164*3998/4473
E12 <- 309*475/4473
E22 <- 4164*475/4473

expected = matrix(c(E11, E21, E12, E22), nrow = 2,
                  dimnames = list(Diabetes = c("No", "Yes"),
                                  BrownFat = c("Yes", "No")))
chisq.test(fat)
ORhat <- (303*469)/(6*3695)
```