IS4250 Health Care Analysis Report

An Exploratory Analysis of Liver Patients from USA and INDIA

Group 08: Huang Jiushi (A0105542R)
Ye Shujian (A0105702W)

1. Introduction

1.1. Background

According to the World Health Organization (WHO) report published in 2014, the number of Liver Disease Deaths in India has reached 216,865 or 2.44% of total deaths. The age adjusted Death Rate is 21.96 per 100,000 of population which made Liver Disease become one of the top 10 causes of death in India. In order to improve the liver medical diagnosis accuracy rate, Prof Tekkali, from Aditya Institute of Technology and Management in India, performed several researches based on different data sets. Significant differences in classification accuracy of various classifiers with different data sets have been indicated. In order to envisage the reason for the difference between different data sets to further improve the diagnosis accuracy, an exploratory analysis was carried out.

1.2. Project Objectives

In order to improve the diagnosis efficiency, India is currently developing the automatic medical diagnosis system to instead the manual diagnosis methods. With the purpose of improving the system accuracy, the results of this study are expected to help the development of the automatic medical diagnosis system and the needs for its localization settings based on the geographical region.

1.3. Data Set Introduction

The study contains 2 data sets, taken from the University of California at Irvine (UCI) Machine Learning Repository and a local hospital in north east, Andhra Pradesh, India respectively. The reason why they choose one data set from USA and one from India is because significant differences within these two data sets are expected to be found due to the regional disparity, which will be helpful for the localization settings of the system.

The first data set indicates UCI data set which contains 345 records with 6 attributes. 145 of them are labelled as liver patients and others are not. The attributes of the UCI data set are MCV(mean corpuscular volume) , Alkphos(alkaline phosphotase), SGPT(alamine aminotransferase), SGOT(aspartate aminotransferase), Gammagt(gamma-glutamyl transpeptidase) and Drinks(number of half-pint equivalents of alcoholic beverages drunk per day).

The second data set indicates India data set which has 583 liver patient records with 10 attributes including Gender, Age, Total_Bilirubin, Direct_Bilirubin, Total_Protiens, Albumin, A/G_Ratio(Albumin and Globulin Ratio), SGPT(alamine aminotransferase), SGOT(aspartate aminotransferase) and Alkphos(alkaline phosphotase). Similarly, 416 records of India data set are labelled as liver patients and others are considered as non-liver patients. The

common attributes from the two data sets are Alkphos, SGPT and SGOT, which will be taken for the purpose of comparison.
Specific tables of data sets are shown like in the below:

Attribute	Type
Mcv	Integer
Alkphos	Integer
SGPT	Integer
SGOT	Integer
Gammagt	Real Number
Drinks	Real Number

Table 1: UCI Liver data set specifications

Attribute	Type
Gender	Alphabetic
Age	Real Number
Total_bilirubin	Real Number
Direct_Bilirubin	Real Number
Total_Protiens	Real Number
Albumin	Real Number
A/G ratio	Real Number
SGPT	Integer
SGOT	Integer
Alkphos	Integer

Table 2: India Liver data set specifications

2. Methodology

Standard statistical methods of One-way Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA) are applied to evaluate the significance between two data sets for better classification.

2.1. One way Analysis of Variance (ANOVA)

ANOVA is a collection of statistical models used to analyze the differences in a single dependent variable among group means and their associated procedure. It will compare the means between the groups and determine whether any of those means are significantly different from each other. Specifically, it tests the null hypothesis:

$$H_0: \ \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$$

Where μ = group and k = number of groups. If the one-way ANOVA returns a significant result, the alternative hypothesis will be accepted, which means that
there are at least 2 group means that are significantly different from each other.

To test the null hypothesis, first the F-ratio need to be calculated. Suppose there are g groups:

- Y_{ij} = Observation from subject j in group i
- n_i = Number of subjects in group i
- $N = n_1 + n_2 + \ldots + n_g$ = Total sample size.
- $\bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ = Sample mean for group i.
- $\bar{y}_{..} = \frac{1}{N} \sum_{i=1}^{g} \sum_{j=1}^{n_i} Y_{ij}$ = Grand mean.

Sum of squares between groups = $\sum_{i=1}^{g} n_i (\overline{y}_{i.} - \overline{y}_{..})^2$

Sum of squares within groups = $\sum_{i=1}^{g} \sum_{j=1}^{n_i} (Y_{ij} - \overline{y}_{i.})^2$

Numerator degree of freedom = g - 1
Denominator degree of freedom = N - g

Mean square between group = Sum of squares between groups / Numerator degree of freedom
Mean square within groups = Sum of squares within groups / Denominator degree of freedom

F-ratio = Mean square between groups / Mean square within groups

After calculating the F-ratio, the value of the F-ratio will be compared with the critical value of the F-ratio. The critical value of F-ratio can be found in the F-distribution tables using numerator degree of freedom and denominator degree of freedom. If the F-ratio is larger than the critical value of the F-ratio, the null hypothesis can be rejected.

2.2. Multivariate Analysis of Variance (MANOVA)

MANOVA is a technique to compare multivariate sample means. As a multivariate procedure, it is used when there are two or more dependent variables, and is typically followed by significance tests involving individual dependent variables separately. It can help find the relationship among the dependent variables and independent variables respectively. The purpose of the MANOVA is to look for an effect of one or more independent variables on several dependent variables at the same time.

3. Results and Discussion

Three common attributes, SGOT, SGPT and Alkphos, exist between the US data set and India data set, and all of them are integer attributes. To show the significant differences between Indian liver patients and the US liver patients, these three common attributes will be the only measures.

Three experiments will be conducted on three different samples. For each experiment, 7 comparisons will be tested to show that the significant differences exist. The first three comparisons will examine whether significant differences on individual common attribute exist between Indian patients and the US patients, i.e. comparison on SGOT, comparison on SGPT, and comparison on Alkphos. The next three comparisons will examine whether significant differences exist on the combination of two common attributes between India patients and US patients, i.e. comparison on SGOT and SGPT, comparison on SGOT and Alkphos, and comparison on SGPT and Alkphos. The last comparison will examine whether significant differences on the combination of all the three common variables exist between Indian patients and the US patients, i.e. comparison on SGOT, SGPT, and Alkphos. The first three comparisons will be tested by ANOVA because only the difference of data sets on single variable will be examined. MANOVA will be used for the rest of the comparisons since the difference of data sets on multiple variables is expected to be examined.

Comparison Number	Dependent Variables	Experiment Methodology
1	SGOT	ANOVA
2	SGPT	ANOVA
3	Alkphos	ANOVA
4	Alkphos & SGPT	MANOVA
5	Alkphos & SGOT	MANOVA
6	SGPT & SGOT	MANOVA
7	Alkphos, SGOT & SGPT	MANOVA

Table 3: Experiment Comparisons Specifications

The comparison on Alkphos in the first experiment will be taken as an example to elaborate ANOVA and the comparison on the combination of Alkphos and SGPT as an example to elaborate MANOVA. The rest of the comparisons in the first experiments as well as experiment 2 and 3 will be briefly discussed.

3.1. Experiment 1

For the first experiment, the sample will be all the records in the two data sets, which means both liver patients and non-liver patients are included in the experiment.

First ANOVA should be used to compare the US patients and Indian patients on Alkphos. The calculation results show that the mean of Alkphos of the US patients and Indian patients is 69.87 and 290.58 respectively. Then the sum of squares of Alkphos between the two groups (Indian patients and US patients) can be calculated as 10557739.946. The numerator degree of freedom is 1 and it is equal to number of groups minus 1. Hence, we can get the mean square between two groups which is 10557739.946 and it is equal to sum of squares between two groups divided by the numerator degree of freedom.

For the sum of squares within two groups, it is equal to 34464783.484. The denominator degree of freedom is 926 and it is equal to the total number of records in two groups (928) minus number of groups. Hence we can get the mean square within two groups which is 37218.989 and it is equal to sum of squares within two groups divided by the denominator degree of freedom.

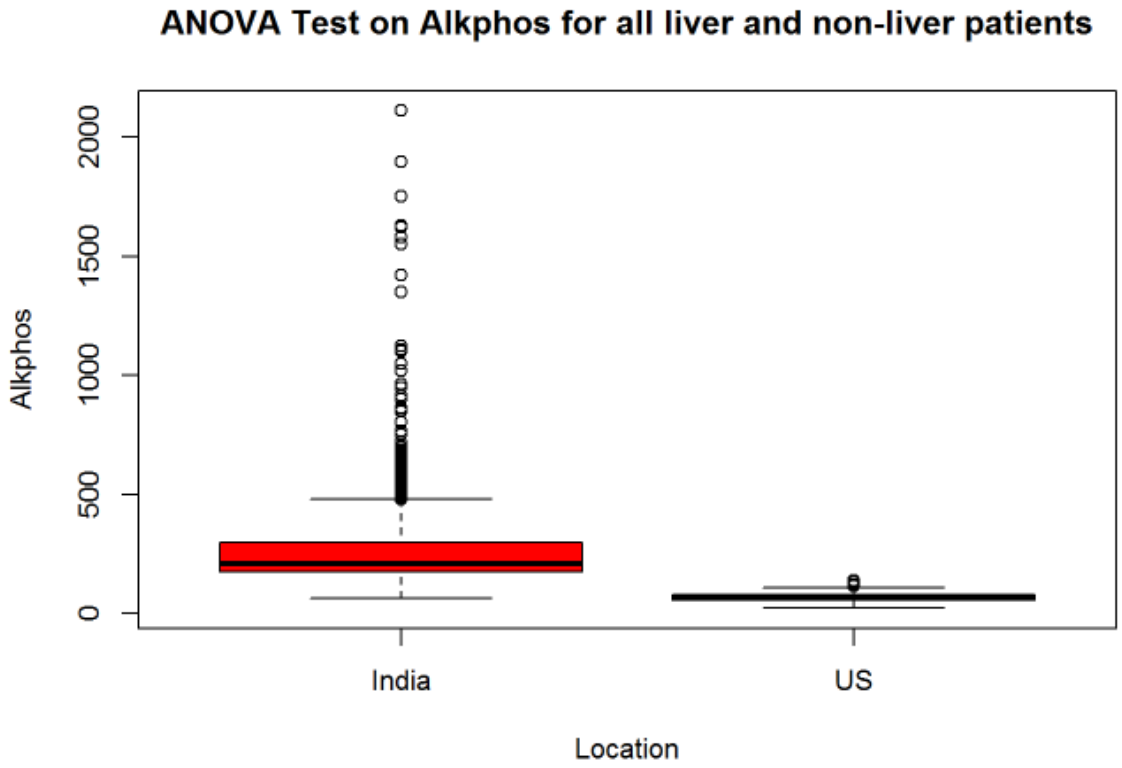
After knowing the value of the mean square between two groups and the mean square within two groups we can calculate the F-ratio which is equal to mean square between two groups divided by mean square within two groups, and it is equal to 283.665. We can use the numerator degree of freedom and the denominator degree of freedom to look for the corresponding critical value in the F distribution table. The F distribution table we use is for 0.001 significance level. We find that the critical value for a numerator degree of freedom of 1 and a denominator degree of freedom of 926 and the 0.001 significance level is 10.82756617. Obviously, the F-ratio is much larger than the critical value. Hence, we safely conclude that the significant difference on the mean value of the Alkphos attribute between Indian patients group and US patients group exist.

----------ANOVA Test on Alkphos in Experiment 1 ----------

```{r}
india <- read.csv("~/Indian Liver Patient Dataset (ILPD).csv")
US <- read.csv("~/bupa.csv")


library(sqldf)
library(ggplot2)

India_SGPT <- sqldf("SELECT SGPT FROM india")
India_SGOT <- sqldf("SELECT SGOT FROM india")
India_Alkphos <- sqldf("SELECT Alkphos FROM india")
```

```
US_SGPT <- sqldf("SELECT SGPT FROM US")
US_SGOT <- sqldf("SELECT SGOT FROM US")
US_Alkphos <- sqldf("SELECT Alkphos FROM US")

India_Alkphos2 <- data.frame(cbind(India_Alkphos,"India"))
colnames(India_Alkphos2) <- c("Alkphos", "Location")

US_Alkphos2 <- data.frame(cbind(US_Alkphos,"US"))
colnames(US_Alkphos2) <- c("Alkphos", "Location")

Alkphos <- data.frame((rbind(India_Alkphos2, US_Alkphos2)))

Anova_Exp1_Alkphos <- aov(Alkphos~Location, data=Alkphos)
summary(Anova_Exp1_Alkphos)

boxplot(Alkphos$Alkphos ~ Alkphos$Location, main="ANOVA Test on Alkphos for all
records", xlab="Location", ylab="Alkphos", col=rainbow(2))
```

```
##              Df   Sum Sq  Mean Sq F value Pr(>F)
## Location      1 10557740 10557740   283.7 <2e-16 ***
## Residuals   926 34464783    37219
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

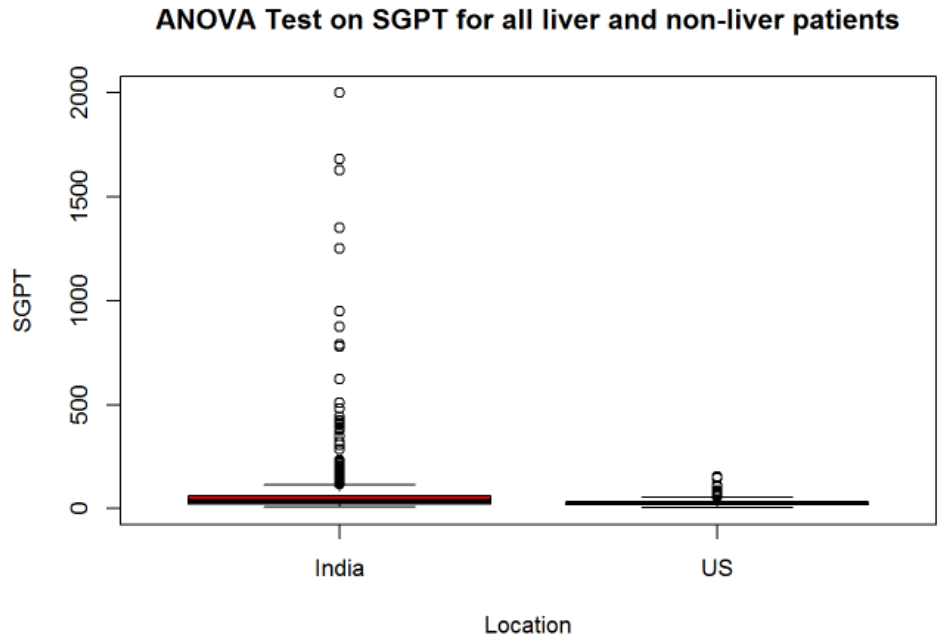


Next we compare Indian and US patients on SGPT using ANOVA. The critical value for 0.001 significance level is the same the critical value in the first comparison which is 10.82756617. The calculated F-ratio of this comparison is 25.994 which is greater than the critical value for 0.001 significance level. Hence, we safely conclude that the significant difference on the mean value of the SGPT attribute between Indian patients group and US patients group exist.

----------ANOVA Test on SGPT in Experiment 1 ----------

```{r}
India_SGPT2 <- data.frame(cbind(India_SGPT,"India"))
colnames(India_SGPT2) <- c("SGPT", "Location")

US_SGPT2 <- data.frame(cbind(US_SGPT,"US"))
colnames(US_SGPT2) <- c("SGPT", "Location")

SGPT <- data.frame((rbind(India_SGPT2, US_SGPT2)))

Anova_Exp1_SGPT <- aov(SGPT~Location, data=SGPT)
summary(Anova_Exp1_SGPT)

boxplot(SGPT$SGPT ~ SGPT$Location, main="ANOVA Test on SGPT for all liver and non-liver patients", xlab="Location", ylab="SGPT", col=rainbow(2))
```

```
##              Df   Sum Sq Mean Sq F value   Pr(>F)
## Location      1   548542  548542   25.99 4.15e-07 ***
## Residuals   926 19540784   21102
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

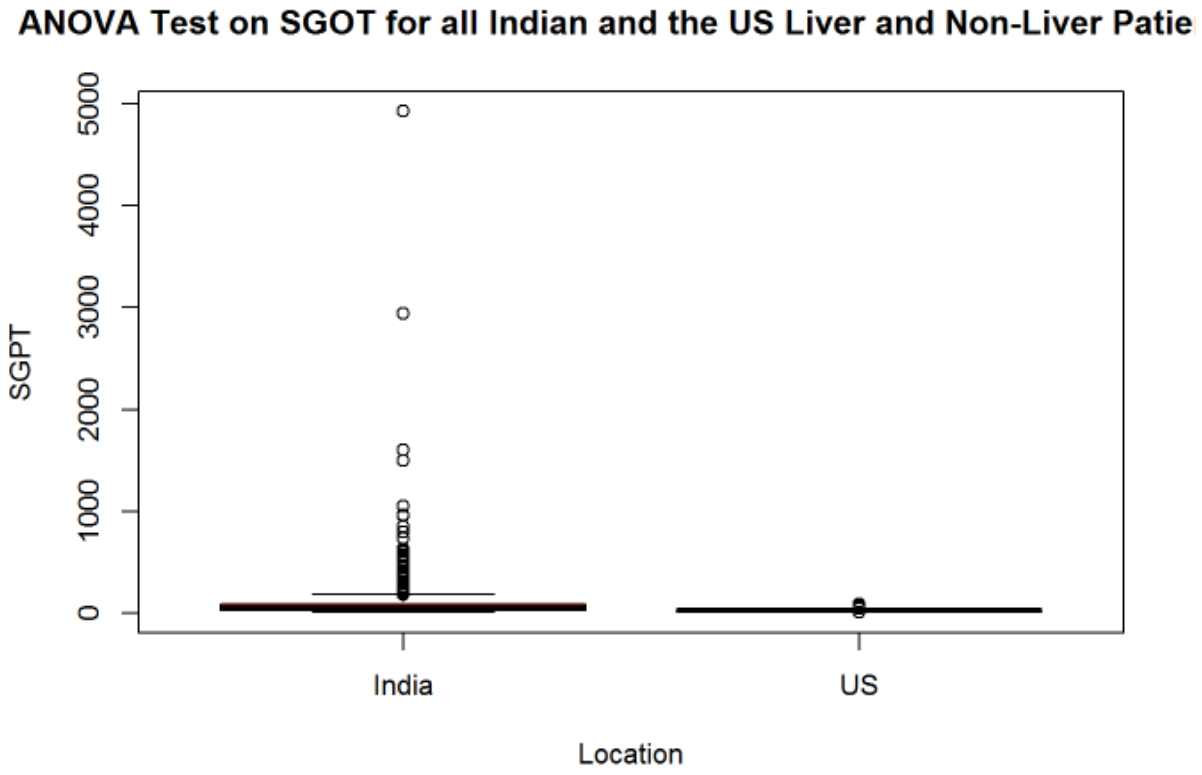


We compare Indian and US patients on SGOT using ANOVA. The critical value for 0.001 significance level is the same the critical value in the first comparison which is 10.82756617. The calculated F-ratio of this comparison is 30.014 which is greater than the critical value for 0.001 significance level. Hence, we safely conclude that the significant difference on the mean value of the SGOT attribute between Indian patients group and US patients group exist.

----------ANOVA Test on SGOT in Experiment 1 ----------

```{r}
India_SGOT2 <- data.frame(cbind(India_SGOT, "India"))
colnames(India_SGOT2) <- c("SGOT", "Location")
US_SGOT2 <- data.frame(cbind(US_SGOT, "US"))
colnames(US_SGOT2) <- c("SGOT", "Location")
SGOT <- data.frame((rbind(India_SGOT2, US_SGOT2)))
Anova_Exp1_SGOT <- aov(SGOT~Location, data=SGOT)
summary(Anova_Exp1_SGOT)
boxplot(SGOT$SGOT ~ SGOT$Location, main="ANOVA Test on SGOT for all Indian and the US Liver and Non-Liver Patients", xlab="Location", ylab="SGPT", col=rainbow(2))
```

```
##              Df   Sum Sq Mean Sq F value   Pr(>F)
## Location      1  1575814 1575814   30.01 5.53e-08 ***
## Residuals   926 48616665   52502
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



----------MANOVA Test on Alkphos&SGPT in Experiment 1----------

```{r}
```

```
Exp1_Alkphos_SGPT_India <- data.frame(cbind(India_Alkphos, India_SGPT,"India"))
colnames(Exp1_Alkphos_SGPT_India) <- c("Alkphos", "SGPT", " Location")

Exp1_Alkphos_SGPT_US <- data.frame(cbind(US_Alkphos, US_SGPT,"US"))
colnames(Exp1_Alkphos_SGPT_US) <- c("Alkphos", "SGPT", " Location")

Exp_Com1 <- data.frame((rbind(Exp1_Alkphos_SGPT_India, Exp1_Alkphos_SGPT_US)))
summary(Exp_Manova1, test="Pillai", intercept = TRUE)
summary(Exp_Manova1, test="Wilks", intercept = TRUE)
summary(Exp_Manova1, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp_Manova1, test="Roy", intercept = TRUE)
```

```
##                       Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.55698   581.47      2    925 < 2.2e-16 ***
## as.factor(X.Location)  1 0.24019   146.20      2    925 < 2.2e-16 ***
## Residuals            926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                       Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.44302   581.47      2    925 < 2.2e-16 ***
## as.factor(X.Location)  1 0.75981   146.20      2    925 < 2.2e-16 ***
## Residuals            926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                       Df Hotelling-Lawley approx F num Df den Df
## (Intercept)            1          1.25723   581.47      2    925
## as.factor(X.Location)  1          0.31612   146.20      2    925
## Residuals            926
##                          Pr(>F)
## (Intercept)           < 2.2e-16 ***
## as.factor(X.Location) < 2.2e-16 ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                       Df     Roy approx F num Df den Df    Pr(>F)
## (Intercept)            1 1.25723   581.47      2    925 < 2.2e-16 ***
## as.factor(X.Location)  1 0.31612   146.20      2    925 < 2.2e-16 ***
## Residuals            926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The fifth, sixth, and seventh comparison are on the combination of Alkphos and SGOT, the combination of SGOT and SGPT, and the combination of Alkphos, SGOT and SGPT respectively using ANOVA. All the statistics in these comparisons show the 0.000 significance level. Hence we reject the null hypothesis and conclude that the differences between Indian patients and US patients on the mean value of the above combinations exist.

From experiment 1 we find that there are significant differences on the mean value of either individual common attribute or all the possible combinations of the common attributes. The differences in the mean values suggest that there are significant differences between Indian and US patients.

3.2. Experiment 2

After showing that there are differences between general US and Indian patients, we want to find out whether there are differences between US and Indian liver patients. For experiment 2, we will choose liver patients in both US data set and Indian data set as the sample. We will still conduct seven comparisons for this experiment. There are 145 out of 345 liver patients in the US data set and 416 out of 583 liver patients in the Indian data set.

We use ANOVA for the first three comparisons in this experiment. The three comparisons are between US and Indian patients and they are on Alkphos, SGOT, SGPT respectively. From the test statistics, we find that the significant values of all three tests are at 0.000 level, from which we reject the null hypothesis and conclude that there are significant differences between US and Indian liver patients on the mean values of the three individual common attributes Alkphos, SGOT, SGPT.

We use MANOVA for the rest of comparisons, which are the comparison on Alkphos & SGPT, the comparison on Alkphos & SGOT, the comparison on SGPT & SGOT, and the comparison on Alkphos, SGPT & SGOT. We find that the significant values of all four tests are at 0.000 level. It shows that there are significant differences between US and Indian liver patients on Alkphos & SGPT, Alkphos & SGOT, SGPT & SGOT, Alkphos, SGPT & SGPT.

In conclusion, experiment 2 shows that there are significant differences between US and Indian liver patients. It is proved by the statistical differences between US and Indian liver patients' data on their common attributes and the all combinations of their common attributes.

3.3. Experiment 3

For experiment 3, we want to find out whether there are significant differences between non-liver patients in US and India. 200 records in US data set are non-liver patients and 167 records in Indian data set are non-liver patients.

The first three comparisons are between US and Indian non-liver patients on Alkphos, SGPT, and SGOT using ANOVA. Test results on Alkphos and SGOT shows the significance level of 0.000, while the test result on SGPT shows the significance value of 0.119 which is larger than 0.05. It shows that US and Indian non-liver patients have no difference on the attribute SGPT. The differences exist on Alkphos and SGOT.

The next four comparisons are on Alkphos & SGPT, Alkphos & SGOT, SGPT & SGOT, Alkphos, SGPT & SGOT. All test statistics show the significance value of 0.000. It shows that US and Indian non-liver patients have no difference on all combinations of their common attributes.

In conclusion, except the case of SGPT, experiment 3 shows that there are significant differences between US and Indian non-liver patients.

All the three experiments show the differences between US and Indian liver patients as well as the differences between US and Indian non-liver patients. This explains the existence of differences in classification accuracy when analyzing the US data set and Indian data set.

4. Limitation

Several limitations exist during the research. The analysis results may be affected due to these limitations. In order to get more accurate results, limitations should be specified for further improvement. During this study, the limitations should mainly include 3 areas, data size, common attributes and data preparation.

4.1. Data Size

One of the project goals is to adjust the localization settings of the automatic medical diagnosis system by finding the significant differences of two data sets from different countries. Both of the data sets only contain several hundred records, while India has over 1 billion population and the US has around 300 million. The sample size is not big enough to fully represent two countries. More data should be collected for further accuracy improvement.

4.2. Common Attributes

The common attributes of the two data sets are Alkphos, SGPT and SGOT. These three attributes and their combinations are tested as the dependent variables, which means these attributes are considered as the main factors of the liver diseases. Actually, more attributes should be tested for the purpose of automatic medical diagnosis system development. In addition to Alkphos, SGPT and SGOT, 5' nucleotidase and gamma-glutamyl transpeptidase (GGT) are a few of the other enzymes located in the liver and should also be tested. So the attributes of data sets are too limited for the development of the system. More broad attributes should be collected for diagnosis accuracy improvements.

4.3. Data Preparation

During the study, no data preparation is mentioned. Good data sets are the basises for the meaningful research outcome. Accurate and effective results can be achieved only when the input data are reliable. Before the study, outliers should be detected and removed. It is observed that some Indian patients' SGOT are over 5,000 while most of patients' SGOT values are below 1,000. Outliers also exist in the SGPT and Alkphos. All these should be detected before the data analysis for higher accuracy.

5. Conclusion

In this study, One-way Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA) are the two techniques used to examine the significant differences of the common attributes of the US and India data sets. Three experiments are performed. Both of the experiment 1 and 2 show that significant differences in the groups with all the possible attribute combinations exist. Only the analysis on SGPT of Experiment 3 indicates that there is no significant difference between groups on SGPT for non liver patients from the US and India, while other attributes and all combinations still exist the significance. The results of the study confirm the difference in liver patients from the US and India, which is very important for the development of automatic medical diagnosis system and its localization settings according to different geographical regions. Drugs for the liver patients from different locations should also be prescribed accordingly.

References:

[1]. Bendi Venkata Ramana, M.Surendra Prasad Babu, N.B.Venkateswarlu, " A Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysis(May, 2012)", IJCSI International Journal of Computer Science Issues

[2]. Bendi Venkata Ramana, M.Surendra Prasad Babu, N.B.Venkateswarlu, "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis.(May, 2011)", IJCSI International Journal of Computer Science Issues

[3]. Bendi Venkata Ramana, M.Surendra Prasad Babu, B.R.Sarath Kumar, "New Automatic Diagnosis of Liver Status Using Bayesian Classification"(Nov, 2010), IEEE International Conference on Intelligent Network and Computing

[4]. BUPA Liver Disorders Data Set. UCI Repository of Machine Learning Databases ,

Retrived from "https://archive.ics.uci.edu/ml/datasets/Liver+Disorders".

[5]. F-Distribution Tables. (n.d.). Retrieved April 05, 2016, from "http://www.socr.ucla.edu/applets.dir/f_table.html"

[6]. (n.d.). Retrieved April 05, 2016, from https://en.wikipedia.org/wiki/F-test

[7]. S. (2012). How To Calculate and Understand Analysis of Variance (ANOVA) F Test. Retrieved April 05, 2016, from "https://www.youtube.com/watch?v=-yQb_ZJnFXw".

[8]. Statistical Soup: ANOVA, ANCOVA, MANOVA, & MANCOVA. (n.d.). Retrieved April 05, 2016, from "http://www.statsmakemecry.com/smmctheblog/stats-soup-anova-ancova-manova-mancova".

[9]. Zaiontz, C. (n.d.). MANOVA Basic Concepts. Retrieved April 05, 2016, from "http://www.real-statistics.com/multivariate-statistics/multivariate-analysis-of-variance-manova/manova-basic-concepts/"

[10]. 13.3 - Test Statistics for MANOVA. (n.d.). Retrieved April 05, 2016, from "https://onlinecourses.science.psu.edu/stat505/node/163"

Appendix

Experiment 1-5: MANOVA Test of Alkphos & SGOT of All Liver and Non-Liver Patients

```{r}
Exp1_Alkphos_SGOT_India <- data.frame(cbind(India_Alkphos, India_SGOT, "India"))
colnames(Exp1_Alkphos_SGOT_India) <- c("Alkphos", "SGOT", "Location")

Exp1_Alkphos_SGOT_US <- data.frame(cbind(US_Alkphos, US_SGOT, "US"))
colnames(Exp1_Alkphos_SGOT_US) <- c("Alkphos", "SGOT", "Location")

Exp1_Com2 <- data.frame((rbind(Exp1_Alkphos_SGOT_India, Exp1_Alkphos_SGOT_US)))

Exp1_Manova2 <- manova(cbind(Alkphos, SGOT) ~ as.factor(Location), data=Exp1_Com2)

summary(Exp1_Manova2, test="Pillai", intercept = TRUE)
summary(Exp1_Manova2, test="Wilks", intercept = TRUE)
summary(Exp1_Manova2, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp1_Manova2, test="Roy", intercept = TRUE)
```

```
summary(Exp1_Manova2, test="Pillai", intercept = TRUE)
```

```
##                      Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)           1 0.54493   553.82      2    925 < 2.2e-16 ***
## as.factor(Location)   1 0.23909   145.33      2    925 < 2.2e-16 ***
## Residuals           926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp1_Manova2, test="Wilks", intercept = TRUE)
```

```
##                      Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)           1 0.45507   553.82      2    925 < 2.2e-16 ***
## as.factor(Location)   1 0.76091   145.33      2    925 < 2.2e-16 ***
## Residuals           926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp1_Manova2, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                      Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)           1          1.19745   553.82      2    925 < 2.2e-16
## as.factor(Location)   1          0.31422   145.33      2    925 < 2.2e-16
## Residuals           926
##
## (Intercept)         ***
## as.factor(Location) ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp1_Manova2, test="Roy", intercept = TRUE)
```

```
##                      Df     Roy approx F num Df den Df    Pr(>F)
## (Intercept)           1 1.19745   553.82      2    925 < 2.2e-16 ***
## as.factor(Location)   1 0.31422   145.33      2    925 < 2.2e-16 ***
## Residuals           926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Experiment 1-6: MANOVA Test of SGOT & SGPT of All Liver and Non-Liver Patients

```{r}
Exp1_SGPT_SGOT_India <- data.frame(cbind(India_SGPT, India_SGOT, "India"))
colnames(Exp1_SGPT_SGOT_India) <- c("SGPT", "SGOT", "Location")

Exp1_SGPT_SGOT_US <- data.frame(cbind(US_SGPT, US_SGOT, "US"))
colnames(Exp1_SGPT_SGOT_US) <- c("SGPT", "SGOT", "Location")

Exp1_Com3 <- data.frame((rbind(Exp1_SGPT_SGOT_India, Exp1_SGPT_SGOT_US)))

Exp1_Manova3 <- manova(cbind(SGPT,SGOT) ~ as.factor(Location), data = Exp1_Com3)

summary(Exp1_Manova3, test="Pillai", intercept = TRUE)
summary(Exp1_Manova3, test="Wilks", intercept = TRUE)
summary(Exp1_Manova3, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp1_Manova3, test="Roy", intercept = TRUE)
```

```r
summary(Exp1_Manova3, test="Pillai", intercept = TRUE)
```

```
##                       Df   Pillai approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.154445   84.478      2    925 < 2.2e-16 ***
## as.factor(Location)    1 0.032982   15.775      2    925 1.834e-07 ***
## Residuals            926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp1_Manova3, test="Wilks", intercept = TRUE)
```

```
##                       Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.84555   84.478      2    925 < 2.2e-16 ***
## as.factor(Location)    1 0.96702   15.775      2    925 1.834e-07 ***
## Residuals            926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp1_Manova3, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                       Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)            1         0.182656   84.478      2    925 < 2.2e-16
## as.factor(Location)    1         0.034107   15.775      2    925 1.834e-07
## Residuals            926
##
## (Intercept)         ***
## as.factor(Location) ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp1_Manova3, test="Roy", intercept = TRUE)
```

```
##                       Df      Roy approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.182656   84.478      2    925 < 2.2e-16 ***
## as.factor(Location)    1 0.034107   15.775      2    925 1.834e-07 ***
## Residuals            926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Experiment 1-7: MANOVA Test of Alkphos, SGPT, SGOT of All Liver and Non-Liver Patients

```{r}
Exp1_Alkphos_SGPT_SGOT_India <- data.frame(cbind(India_Alkphos, India_SGPT, India_SGOT, "India"))
colnames(Exp1_Alkphos_SGPT_SGOT_India) <- c("Alkphos", "SGPT", "SGOT", "Location")

Exp1_Alkphos_SGPT_SGOT_US <- data.frame(cbind(US_Alkphos, US_SGPT, US_SGOT, "US"))
colnames(Exp1_Alkphos_SGPT_SGOT_US) <- c("Alkphos", "SGPT", "SGOT", "Location")

Exp1_Com4 <- data.frame((rbind(Exp1_Alkphos_SGPT_SGOT_India, Exp1_Alkphos_SGPT_SGOT_US)))

Exp1_Manova4 <- manova(cbind(Alkphos, SGPT, SGOT) ~ as.factor(Location), data = Exp1_Com4)

summary(Exp1_Manova4, test="Pillai", intercept = TRUE)
summary(Exp1_Manova4, test="Wilks", intercept = TRUE)
summary(Exp1_Manova4, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp1_Manova4, test="Roy", intercept = TRUE)
```

```
summary(Exp1_Manova4, test="Pillai", intercept = TRUE)
```

```
##                      Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)           1 0.55936   390.98      3    924 < 2.2e-16 ***
## as.factor(Location)   1 0.24037    97.46      3    924 < 2.2e-16 ***
## Residuals           926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp1_Manova4, test="Wilks", intercept = TRUE)
```

```
##                      Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)           1 0.44064   390.98      3    924 < 2.2e-16 ***
## as.factor(Location)   1 0.75963    97.46      3    924 < 2.2e-16 ***
## Residuals           926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp1_Manova4, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                      Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)           1          1.26940   390.98      3    924 < 2.2e-16
## as.factor(Location)   1          0.31643    97.46      3    924 < 2.2e-16
## Residuals           926
##
## (Intercept)         ***
## as.factor(Location) ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp1_Manova4, test="Roy", intercept = TRUE)
```

```
##                      Df     Roy approx F num Df den Df    Pr(>F)
## (Intercept)           1 1.26940   390.98      3    924 < 2.2e-16 ***
## as.factor(Location)   1 0.31643    97.46      3    924 < 2.2e-16 ***
## Residuals           926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Experiment 2-1: ANOVA Test on Alkphos of all Liver Patients

```{r}
India_Liver_SGPT <- sqldf("SELECT SGPT FROM india WHERE Selector = 1")
India_Liver_SGOT <- sqldf("SELECT SGOT FROM india WHERE Selector = 1")
India_Liver_Alkphos <- sqldf("SELECT Alkphos FROM india WHERE Selector = 1")

US_Liver_SGPT <- sqldf("SELECT SGPT FROM US WHERE Selector = 1")
US_Liver_SGOT <- sqldf("SELECT SGOT FROM US WHERE Selector = 1")
US_Liver_Alkphos <- sqldf("SELECT Alkphos FROM US WHERE Selector =1")

India_Liver_Alkphos2 <- data.frame(cbind(India_Liver_Alkphos, "India"))
colnames(India_Liver_Alkphos2) <- c("Alkphos", "Location")

US_Liver_Alkphos2 <- data.frame(cbind(US_Liver_Alkphos, "US"))
colnames(US_Liver_Alkphos2) <- c("Alkphos", "Location")

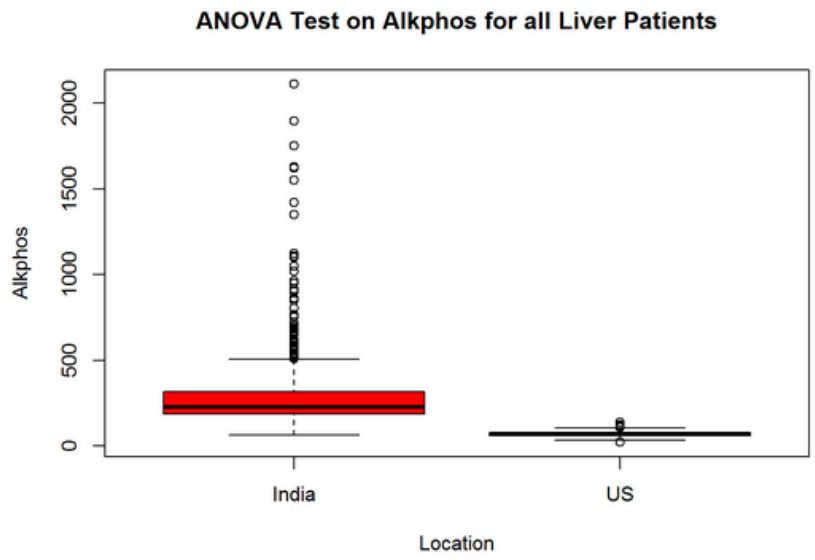
Liver_Alkphos <- data.frame((rbind(India_Liver_Alkphos2, US_Liver_Alkphos2)))
Exp2_Anova_Alkphos <- aov(Alkphos~Location, data=Liver_Alkphos)

summary(Exp2_Anova_Alkphos)
```

boxplot(Liver_Alkphos$Alkphos ~ Liver_Alkphos$Location, main="ANOVA Test on Alkphos for all Liver Patients", xlab="Location", ylab="Alkphos", col=rainbow(2))
```

```
##              Df   Sum Sq Mean Sq F value Pr(>F)
## Location      1  6561309 6561309   122.6 <2e-16 ***
## Residuals   559 29925260   53534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
boxplot(Liver_Alkphos$Alkphos ~ Liver_Alkphos$Location, main="ANOVA Test on Alkphos for all Liver Patients", xlab="Location", ylab="Alkphos", col=rainbow(2))
```



Experiment 2-2 ANOVA Test on SGPT for All Liver Patients

```{r}
India_Liver_SGPT2 <- data.frame(cbind(India_Liver_SGPT, "India"))
colnames(India_Liver_SGPT2) <- c("SGPT", "Location")

US_Liver_SGPT2 <- data.frame(cbind(US_Liver_SGPT, "US"))
colnames(US_Liver_SGPT2) <- c("SGPT", "Location")

Liver_SGPT <- data.frame((rbind(India_Liver_SGPT2, US_Liver_SGPT2)))
Exp2_Anova_SGPT <- aov(SGPT~Location, data=Liver_SGPT)

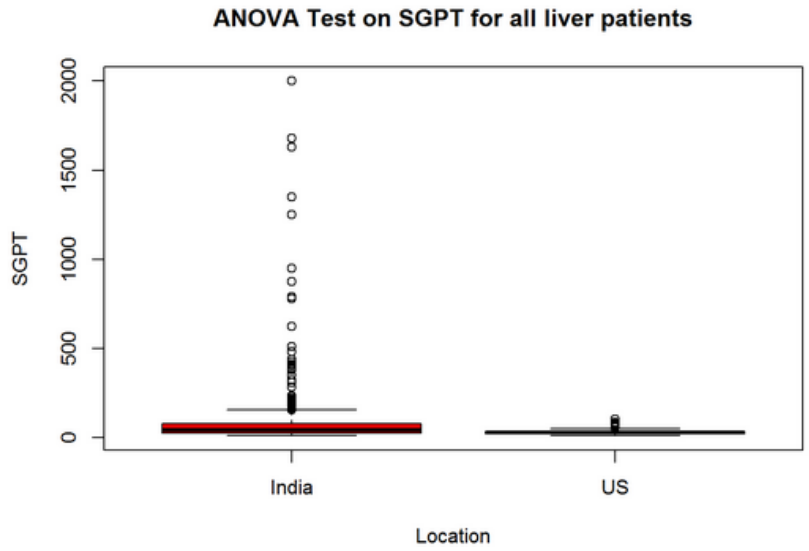
summary(Exp2_Anova_SGPT)

boxplot(Liver_SGPT$SGPT ~ Liver_SGPT$Location, main="ANOVA Test on SGPT for all liver patients", xlab="Location", ylab="SGPT", col=rainbow(2))
```
```

```
summary(Exp2_Anova_SGPT)
```

```
##              Df   Sum Sq Mean Sq F value   Pr(>F)
## Location      1   503033  503033   14.94 0.000124 ***
## Residuals   559 18823073   33673
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
boxplot(Liver_SGPT$SGPT ~ Liver_SGPT$Location, main="ANOVA Test on SGPT for all liver patients", xlab="Location", ylab="SGPT", col=rainbow(2))
```



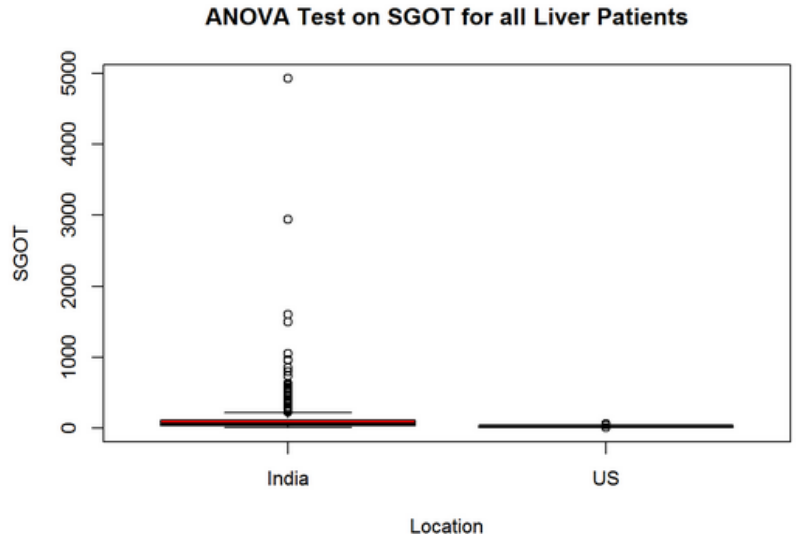
Experiment 2-3 ANOVA Test on SGOT for All Liver Patients

```{r}
India_Liver_SGOT2 <- data.frame(cbind(India_Liver_SGOT, "India"))
colnames(India_Liver_SGOT2) <- c ("SGOT", "Location")

US_Liver_SGOT2 <- data.frame(cbind(US_Liver_SGOT, "US"))
colnames(US_Liver_SGOT2) <- c("SGOT", "Location")

Liver_SGOT <- data.frame((rbind(India_Liver_SGOT2, US_Liver_SGOT2)))
Exp2_Anova_SGOT <- aov(SGOT~Location, data=Liver_SGOT)

summary(Exp2_Anova_SGOT)

boxplot(Liver_SGOT$SGOT ~ Liver_SGOT$Location, main="ANOVA Test on SGOT for all Liver Patients", xlab="Location", ylab="SGOT", col=rainbow(2))
```

```
summary(Exp2_Anova_SGOT)
```

```
##              Df   Sum Sq Mean Sq F value   Pr(>F)
## Location      1  1419839 1419839    16.8 4.77e-05 ***
## Residuals   559 47248902   84524
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
boxplot(Liver_SGOT$SGOT ~ Liver_SGOT$Location, main="ANOVA Test on SGOT for all Liver Patients", xlab="Location", ylab="SGOT", col=rainbow(2))
```



Experiment 2-4 MANOVA Test on Alkphos & SGPT for All Liver Patients

```{r}
Exp2_Alkphos_SGPT_India <- data.frame(cbind(India_Liver_Alkphos, India_Liver_SGPT, "India"))
colnames(Exp2_Alkphos_SGPT_India) <- c("Alkphos", "SGPT", "Location")

Exp2_Alkphos_SGPT_US <- data.frame(cbind(US_Liver_Alkphos, US_Liver_SGPT, "US"))
colnames(Exp2_Alkphos_SGPT_US) <- c("Alkphos", "SGPT", "Location")

Exp2_Com1 <- data.frame((rbind(Exp2_Alkphos_SGPT_India, Exp2_Alkphos_SGPT_US)))

Exp2_Manova1 <- manova(cbind(Alkphos,SGPT) ~ as.factor(Location), data=Exp2_Com1)

summary(Exp2_Manova1, test="Pillai", intercept = TRUE)
summary(Exp2_Manova1, test="Wilks", intercept = TRUE)
summary(Exp2_Manova1, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp2_Manova1, test="Roy", intercept = TRUE)
```

```r
summary(Exp2_Manova1, test="Pillai", intercept = TRUE)
```

```
##                       Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.57249   373.61      2    558 < 2.2e-16 ***
## as.factor(Location)    1 0.18936    65.17      2    558 < 2.2e-16 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp2_Manova1, test="Wilks", intercept = TRUE)
```

```
##                       Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.42751   373.61      2    558 < 2.2e-16 ***
## as.factor(Location)    1 0.81064    65.17      2    558 < 2.2e-16 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp2_Manova1, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                       Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)            1          1.33910   373.61      2    558 < 2.2e-16
## as.factor(Location)    1          0.23359    65.17      2    558 < 2.2e-16
## Residuals            559
##
## (Intercept)         ***
## as.factor(Location) ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp2_Manova1, test="Roy", intercept = TRUE)
```

```
##                       Df     Roy approx F num Df den Df    Pr(>F)
## (Intercept)            1 1.33910   373.61      2    558 < 2.2e-16 ***
## as.factor(Location)    1 0.23359    65.17      2    558 < 2.2e-16 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Experiment 2-5 MANOVA Test on Alkphos & SGOT for All Liver Patients

```{r}
Exp2_Alkphos_SGOT_India <- data.frame(cbind(India_Liver_Alkphos, India_Liver_SGOT, "India"))
colnames(Exp2_Alkphos_SGOT_India) <- c("Alkphos", "SGOT", "Location")
```

```
Exp2_Alkphos_SGOT_US <- data.frame(cbind(US_Liver_Alkphos, US_Liver_SGOT, "US"))
colnames(Exp2_Alkphos_SGOT_US) <- c("Alkphos", "SGOT", "Location")

Exp2_Com2 <- data.frame((rbind(Exp2_Alkphos_SGOT_India, Exp2_Alkphos_SGOT_US)))
Exp2_Manova2 <- manova(cbind(Alkphos, SGOT) ~ as.factor(Location), data=Exp2_Com2)

summary(Exp2_Manova2, test="Pillai", intercept = TRUE)
summary(Exp2_Manova2, test="Wilks", intercept = TRUE)
summary(Exp2_Manova2, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp2_Manova2, test="Roy", intercept = TRUE)
```

```
summary(Exp2_Manova2, test="Pillai", intercept = TRUE)
```

```
##                       Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.55876   353.30      2    558 < 2.2e-16 ***
## as.factor(Location)    1 0.18739    64.34      2    558 < 2.2e-16 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp2_Manova2, test="Wilks", intercept = TRUE)
```

```
##                       Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.44124   353.30      2    558 < 2.2e-16 ***
## as.factor(Location)    1 0.81261    64.34      2    558 < 2.2e-16 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp2_Manova2, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                       Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)            1           1.2663   353.30      2    558 < 2.2e-16
## as.factor(Location)    1           0.2306    64.34      2    558 < 2.2e-16
## Residuals            559
##
## (Intercept)          ***
## as.factor(Location)  ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp2_Manova2, test="Roy", intercept = TRUE)
```

```
##                       Df    Roy approx F num Df den Df    Pr(>F)
## (Intercept)            1 1.2663   353.30      2    558 < 2.2e-16 ***
## as.factor(Location)    1 0.2306    64.34      2    558 < 2.2e-16 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Experiment 2-6 MANOVA Test on SGOT & SGPT for All Liver Patients

```{r}
Exp2_SGOT_SGPT_India <- data.frame(cbind(India_Liver_SGOT, India_Liver_SGPT, "India"))
colnames(Exp2_SGOT_SGPT_India) <- c("SGOT", "SGPT", "Location")

Exp2_SGOT_SGPT_US <- data.frame(cbind(US_Liver_SGOT, US_Liver_SGPT, "US"))
colnames(Exp2_SGOT_SGPT_US) <- c("SGOT", "SGPT", "Location")

Exp2_Com3 <- data.frame((rbind(Exp2_SGOT_SGPT_India, Exp2_SGOT_SGPT_US)))
Exp2_Manova3 <- manova(cbind(SGOT, SGPT) ~ as.factor(Location), data=Exp2_Com3)

summary(Exp2_Manova3, test="Pillai", intercept = TRUE)
summary(Exp2_Manova3, test="Wilks", intercept = TRUE)
summary(Exp2_Manova3, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp2_Manova3, test="Roy", intercept = TRUE)
```

```r
summary(Exp2_Manova3, test="Pillai", intercept = TRUE)
```

```
##                       Df   Pillai approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.167438   56.110      2    558 < 2.2e-16 ***
## as.factor(Location)    1 0.030983    8.921      2    558 0.0001536 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp2_Manova3, test="Wilks", intercept = TRUE)
```

```
##                       Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.83256   56.110      2    558 < 2.2e-16 ***
## as.factor(Location)    1 0.96902    8.921      2    558 0.0001536 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp2_Manova3, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                       Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)            1         0.201112   56.110      2    558 < 2.2e-16
## as.factor(Location)    1         0.031973    8.921      2    558 0.0001536
## Residuals            559
##
## (Intercept)          ***
## as.factor(Location)  ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp2_Manova3, test="Roy", intercept = TRUE)
```

```
##                       Df      Roy approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.201112   56.110      2    558 < 2.2e-16 ***
## as.factor(Location)    1 0.031973    8.921      2    558 0.0001536 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Experiment 2-7 MANOVA Test on Alkphos $ SGOT & SGPT for All Liver Patients

```{r}
Exp2_Alkphos_SGPT_SGOT_India <- data.frame(cbind(India_Liver_Alkphos, India_Liver_SGPT, India_Liver_SGOT, "India"))
colnames(Exp2_Alkphos_SGPT_SGOT_India) <- c("Alkphos", "SGPT", "SGOT", "Location")
```

```
Exp2_Alkphos_SGPT_SGOT_US <- data.frame(cbind(US_Liver_Alkphos, US_Liver_SGPT, US_Liver_SGOT, "US"))
colnames(Exp2_Alkphos_SGPT_SGOT_US) <- c("Alkphos", "SGPT", "SGOT", "Location")

Exp2_Com4 <- data.frame((rbind(Exp2_Alkphos_SGPT_SGOT_India, Exp2_Alkphos_SGPT_SGOT_US)))
Exp2_Manova4 <- manova(cbind(Alkphos, SGPT, SGOT) ~ as.factor(Location), data=Exp2_Com4)

summary(Exp2_Manova4, test="Pillai", intercept = TRUE)
summary(Exp2_Manova4, test="Wilks", intercept = TRUE)
summary(Exp2_Manova4, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp2_Manova4, test="Roy", intercept = TRUE)
```

```
summary(Exp2_Manova4, test="Pillai", intercept = TRUE)
```

```
##                       Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.57399  250.162      3    557 < 2.2e-16 ***
## as.factor(Location)    1 0.18963   43.446      3    557 < 2.2e-16 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp2_Manova4, test="Wilks", intercept = TRUE)
```

```
##                       Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.42601  250.162      3    557 < 2.2e-16 ***
## as.factor(Location)    1 0.81037   43.446      3    557 < 2.2e-16 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp2_Manova4, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                       Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)            1           1.3474  250.162      3    557 < 2.2e-16
## as.factor(Location)    1           0.2340   43.446      3    557 < 2.2e-16
## Residuals            559
##
## (Intercept)          ***
## as.factor(Location)  ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp2_Manova4, test="Roy", intercept = TRUE)
```

```
##                       Df    Roy approx F num Df den Df    Pr(>F)
## (Intercept)            1 1.3474  250.162      3    557 < 2.2e-16 ***
## as.factor(Location)    1 0.2340   43.446      3    557 < 2.2e-16 ***
## Residuals            559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

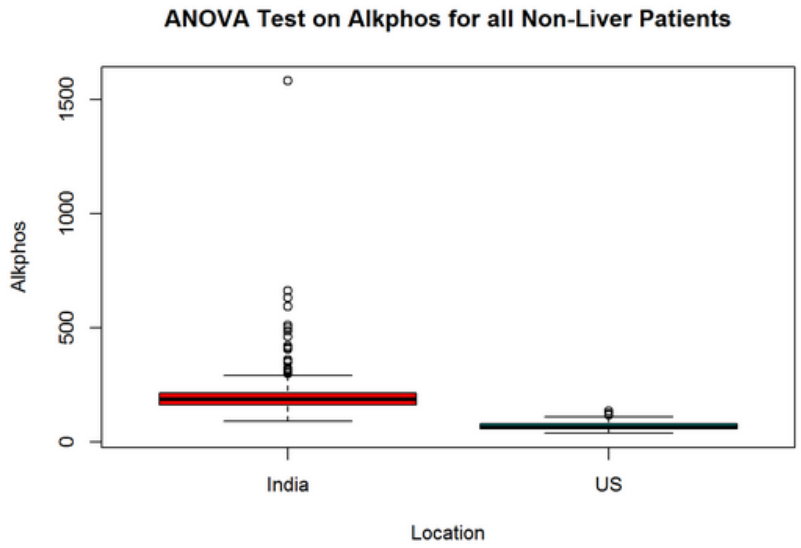
Experiment 3-1: ANOVA Test on Alkphos of all Non-Liver Patients

```{r}
India_NonLiver_SGPT <- sqldf("SELECT SGPT FROM india WHERE Selector = 2")
India_NonLiver_SGOT <- sqldf("SELECT SGOT FROM india WHERE Selector = 2")
India_NonLiver_Alkphos <- sqldf("SELECT Alkphos FROM india WHERE Selector = 2")

US_NonLiver_SGPT <- sqldf("SELECT SGPT FROM US WHERE Selector = 2")
US_NonLiver_SGOT <- sqldf("SELECT SGOT FROM US WHERE Selector = 2")
US_NonLiver_Alkphos <- sqldf("SELECT Alkphos FROM US WHERE Selector = 2")

India_NonLiver_Alkphos2 <- data.frame(cbind(India_NonLiver_Alkphos, "India"))
colnames(India_NonLiver_Alkphos2) <- c("Alkphos", "Location")

US_NonLiver_Alkphos2 <- data.frame(cbind(US_NonLiver_Alkphos, "US"))
colnames(US_NonLiver_Alkphos2) <- c("Alkphos", "Location")

NonLiver_Alkphos <- data.frame((rbind(India_NonLiver_Alkphos2, US_NonLiver_Alkphos2)))
Exp3_Anova_Alkphos <- aov(Alkphos~Location, data=NonLiver_Alkphos)

summary(Exp3_Anova_Alkphos)

boxplot(NonLiver_Alkphos$Alkphos ~ NonLiver_Alkphos$Location, main="ANOVA Test on
Alkphos for all Non-Liver Patients", xlab="Location", ylab="Alkphos", col=rainbow(2))
```

```
summary(Exp3_Anova_Alkphos)
```

```
##              Df  Sum Sq Mean Sq F value Pr(>F)
## Location      1 2086485 2086485   226.4 <2e-16 ***
## Residuals   365 3364524    9218
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
boxplot(NonLiver_Alkphos$Alkphos ~ NonLiver_Alkphos$Location, main="ANOVA Test on Alkphos for all Non-Liver Patients", xlab="Location", ylab="Alkphos", col=rainbow(2))
```



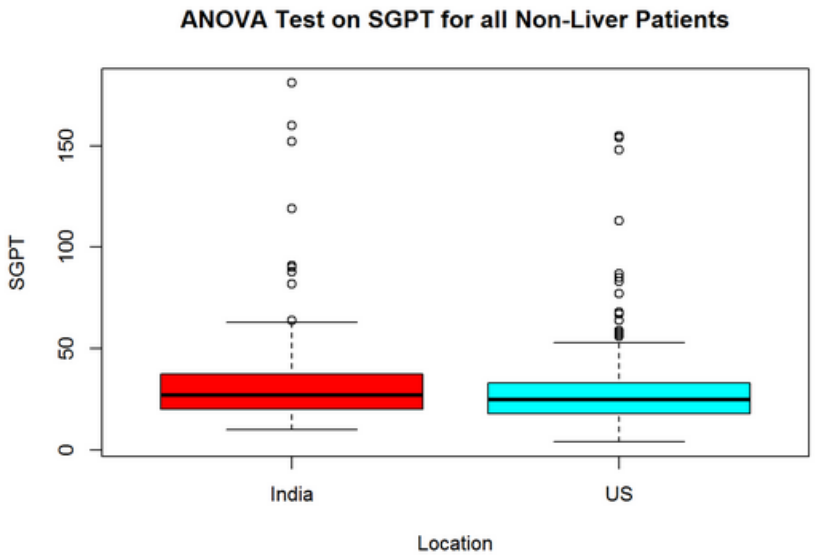
Experiment 3-2 ANOVA Test on SGPT for All Non-Liver Patients

```{r}
India_NonLiver_SGPT2 <- data.frame(cbind(India_NonLiver_SGPT, "India"))
colnames(India_NonLiver_SGPT2) <- c("SGPT", "Location")

US_NonLiver_SGPT2 <- data.frame(cbind(US_NonLiver_SGPT, "US"))
colnames(US_NonLiver_SGPT2) <- c("SGPT", "Location")

NonLiver_SGPT <- data.frame((rbind(India_NonLiver_SGPT2, US_NonLiver_SGPT2)))
Exp3_Anova_SGPT <- aov(SGPT~Location, data=NonLiver_SGPT)

summary(Exp3_Anova_SGPT)

boxplot(NonLiver_SGPT$SGPT ~ NonLiver_SGPT$Location, main="ANOVA Test on SGPT for all Non-Liver Patients", xlab="Location", ylab="SGPT", col=rainbow(2))

```

```
summary(Exp3_Anova_SGPT)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Location      1   1333  1333.4   2.443  0.119
## Residuals   365 199215   545.8
```

```
boxplot(NonLiver_SGPT$SGPT ~ NonLiver_SGPT$Location, main="ANOVA Test on SGPT for all Non-Liver Patients", xlab="Location"
, ylab="SGPT", col=rainbow(2))
```



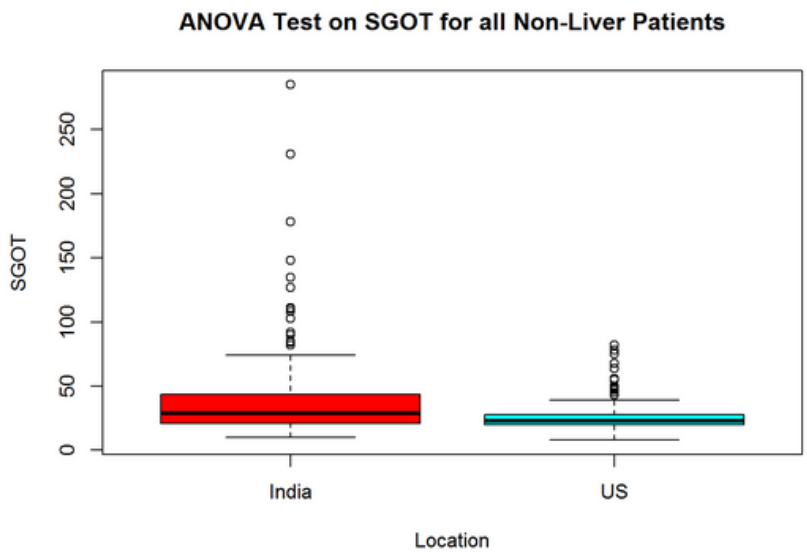
Experiment 3-3 ANOVA Test on SGOT for All Non-Liver Patients

```{r}
India_NonLiver_SGOT2 <- data.frame(cbind(India_NonLiver_SGOT, "India"))
colnames(India_NonLiver_SGOT2) <- c("SGOT", "Location")

US_NonLiver_SGOT2 <- data.frame(cbind(US_NonLiver_SGOT, "US"))
colnames(US_NonLiver_SGOT2) <- c("SGOT", "Location")

NonLiver_SGOT <- data.frame((rbind(India_NonLiver_SGOT2, US_NonLiver_SGOT2)))
Exp3_Anova_SGOT <- aov (SGOT~Location, data=NonLiver_SGOT)

summary(Exp3_Anova_SGOT)

boxplot(NonLiver_SGOT$SGOT ~ NonLiver_SGOT$Location, main="ANOVA Test on
SGOT for all Non-Liver Patients", xlab="Location", ylab="SGOT", col=rainbow(2))
```

```
summary(Exp3_Anova_SGOT)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Location      1  19662   19662   29.24 1.16e-07 ***
## Residuals   365 245444     672
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
boxplot(NonLiver_SGOT$SGOT ~ NonLiver_SGOT$Location, main="ANOVA Test on SGOT for all Non-Liver Patients", xlab="Location"
, ylab="SGOT", col=rainbow(2))
```



Experiment 3-4 MANOVA Test on Alkphos & SGPT for All Non-Liver Patients

```{r}
Exp3_Alkphos_SGPT_India_Non <- data.frame(cbind(India_NonLiver_Alkphos, India_NonLiver_SGPT, "India"))
colnames(Exp3_Alkphos_SGPT_India_Non) <- c("Alkphos", "SGPT", "Location")
Exp3_Alkphos_SGPT_US_Non <- data.frame(cbind(US_NonLiver_Alkphos, US_NonLiver_SGPT, "US"))
colnames(Exp3_Alkphos_SGPT_US_Non) <- c("Alkphos", "SGPT", "Location")
Exp3_Com1 <- data.frame((rbind(Exp3_Alkphos_SGPT_India_Non, Exp3_Alkphos_SGPT_US_Non)))
Exp3_Manova1 <- manova(cbind(Alkphos,SGPT) ~ as.factor(Location), data=Exp3_Com1)

summary(Exp3_Manova1, test="Pillai", intercept = TRUE)
summary(Exp3_Manova1, test="Wilks", intercept = TRUE)
summary(Exp3_Manova1, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp3_Manova1, test="Roy", intercept = TRUE)
```

```r
summary(Exp3_Manova1, test="Pillai", intercept = TRUE)
```

```
##                       Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.75236   552.94      2    364 < 2.2e-16 ***
## as.factor(Location)    1 0.39074   116.72      2    364 < 2.2e-16 ***
## Residuals            365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp3_Manova1, test="Wilks", intercept = TRUE)
```

```
##                       Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.24764   552.94      2    364 < 2.2e-16 ***
## as.factor(Location)    1 0.60926   116.72      2    364 < 2.2e-16 ***
## Residuals            365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp3_Manova1, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                       Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)            1          3.03814   552.94      2    364 < 2.2e-16
## as.factor(Location)    1          0.64133   116.72      2    364 < 2.2e-16
## Residuals            365
##
## (Intercept)          ***
## as.factor(Location)  ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp3_Manova1, test="Roy", intercept = TRUE)
```

```
##                       Df     Roy approx F num Df den Df    Pr(>F)
## (Intercept)            1 3.03814   552.94      2    364 < 2.2e-16 ***
## as.factor(Location)    1 0.64133   116.72      2    364 < 2.2e-16 ***
## Residuals            365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Experiment 3-5 MANOVA Test on Alkphos & SGOT for All Non-Liver Patients

```{r}
Exp3_Alkphos_SGOT_India_Non <- data.frame(cbind(India_NonLiver_Alkphos, India_NonLiver_SGOT, "India"))
colnames(Exp3_Alkphos_SGOT_India_Non) <- c("Alkphos", "SGOT", "Location")

Exp3_Alkphos_SGOT_US_Non <- data.frame(cbind(US_NonLiver_Alkphos, US_NonLiver_SGOT, "US"))
colnames(Exp3_Alkphos_SGOT_US_Non) <- c("Alkphos", "SGOT", "Location")
```

```
Exp3_Com2 <- data.frame((rbind(Exp3_Alkphos_SGOT_India_Non,
Exp3_Alkphos_SGOT_US_Non)))
Exp3_Manova2 <- manova(cbind(Alkphos, SGOT) ~ as.factor(Location), data=Exp3_Com2)

summary(Exp3_Manova2, test="Pillai", intercept = TRUE)
summary(Exp3_Manova2, test="Wilks", intercept = TRUE)
summary(Exp3_Manova2, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp3_Manova2, test="Roy", intercept = TRUE)
```

```
summary(Exp3_Manova2, test="Pillai", intercept = TRUE)
```

```
##                      Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)           1 0.74443   530.14      2    364 < 2.2e-16 ***
## as.factor(Location)   1 0.38542   114.14      2    364 < 2.2e-16 ***
## Residuals           365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp3_Manova2, test="Wilks", intercept = TRUE)
```

```
##                      Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)           1 0.25557   530.14      2    364 < 2.2e-16 ***
## as.factor(Location)   1 0.61458   114.14      2    364 < 2.2e-16 ***
## Residuals           365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp3_Manova2, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                      Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)           1          2.91285   530.14      2    364 < 2.2e-16
## as.factor(Location)   1          0.62712   114.14      2    364 < 2.2e-16
## Residuals           365
##
## (Intercept)          ***
## as.factor(Location)  ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp3_Manova2, test="Roy", intercept = TRUE)
```

```
##                      Df     Roy approx F num Df den Df    Pr(>F)
## (Intercept)           1 2.91285   530.14      2    364 < 2.2e-16 ***
## as.factor(Location)   1 0.62712   114.14      2    364 < 2.2e-16 ***
## Residuals           365
## ---
```

Experiment 3-6 MANOVA Test on SGPT & SGOT for All Non-Liver Patients

```{r}
Exp3_SGPT_SGOT_India_Non <- data.frame(cbind(India_NonLiver_SGPT,
India_NonLiver_SGOT, "India"))
```

```
colnames(Exp3_SGPT_SGOT_India_Non) <- c("SGPT", "SGOT", "Location")

Exp3_SGPT_SGOT_US_Non <- data.frame(cbind(US_NonLiver_SGPT,
US_NonLiver_SGOT, "US"))
colnames(Exp3_SGPT_SGOT_US_Non) <- c("SGPT", "SGOT", "Location")

Exp3_Com3 <- data.frame((rbind(Exp3_SGPT_SGOT_India_Non,
Exp3_SGPT_SGOT_US_Non)))
Exp3_Manova3 <- manova(cbind(SGPT, SGOT) ~ as.factor(Location), data=Exp3_Com3)

summary(Exp3_Manova3, test="Pillai", intercept = TRUE)
summary(Exp3_Manova3, test="Wilks", intercept = TRUE)
summary(Exp3_Manova3, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp3_Manova3, test="Roy", intercept = TRUE)
```

```
summary(Exp3_Manova3, test="Pillai", intercept = TRUE)
```

```
##                       Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.67919   385.31      2    364 < 2.2e-16 ***
## as.factor(Location)    1 0.08701    17.34      2    364 6.383e-08 ***
## Residuals            365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp3_Manova3, test="Wilks", intercept = TRUE)
```

```
##                       Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.32081   385.31      2    364 < 2.2e-16 ***
## as.factor(Location)    1 0.91299    17.34      2    364 6.383e-08 ***
## Residuals            365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp3_Manova3, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                       Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)            1           2.1171   385.31      2    364 < 2.2e-16
## as.factor(Location)    1           0.0953    17.34      2    364 6.383e-08
## Residuals            365
##
## (Intercept)          ***
## as.factor(Location)  ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Exp3_Manova3, test="Roy", intercept = TRUE)
```

```
##                       Df    Roy approx F num Df den Df    Pr(>F)
## (Intercept)            1 2.1171   385.31      2    364 < 2.2e-16 ***
## as.factor(Location)    1 0.0953    17.34      2    364 6.383e-08 ***
## Residuals            365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Experiment 3-7 MANOVA Test on Alkphos & SGPT & SGOT for All Non-Liver Patients

```{r}
Exp3_Alkphos_SGPT_SGOT_India_Non <- data.frame(cbind(India_NonLiver_Alkphos, India_NonLiver_SGPT, India_NonLiver_SGOT, "India"))
colnames(Exp3_Alkphos_SGPT_SGOT_India_Non) <- c("Alkphos", "SGPT", "SGOT", "Location")

Exp3_Alkphos_SGPT_SGOT_US_Non <- data.frame(cbind(US_NonLiver_Alkphos, US_NonLiver_SGPT, US_NonLiver_SGOT, "US"))
colnames(Exp3_Alkphos_SGPT_SGOT_US_Non) <- c("Alkphos", "SGPT", "SGOT", "Location")

Exp3_Com4 <- data.frame((rbind(Exp3_Alkphos_SGPT_SGOT_India_Non, Exp3_Alkphos_SGPT_SGOT_US_Non)))
Exp3_Manova4 <- manova(cbind(Alkphos, SGPT, SGOT) ~ as.factor(Location), data=Exp3_Com4)

summary(Exp3_Manova4, test="Pillai", intercept = TRUE)
summary(Exp3_Manova4, test="Wilks", intercept = TRUE)
summary(Exp3_Manova4, test="Hotelling-Lawley", intercept = TRUE)
summary(Exp3_Manova4, test="Roy", intercept = TRUE)


```
```
```

```r
summary(Exp3_Manova4, test="Pillai", intercept = TRUE)
```

```
##                       Df  Pillai approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.76249   388.45      3    363 < 2.2e-16 ***
## as.factor(Location)    1 0.40716    83.10      3    363 < 2.2e-16 ***
## Residuals            365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp3_Manova4, test="Wilks", intercept = TRUE)
```

```
##                       Df   Wilks approx F num Df den Df    Pr(>F)
## (Intercept)            1 0.23751   388.45      3    363 < 2.2e-16 ***
## as.factor(Location)    1 0.59284    83.10      3    363 < 2.2e-16 ***
## Residuals            365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp3_Manova4, test="Hotelling-Lawley", intercept = TRUE)
```

```
##                       Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## (Intercept)            1           3.2103   388.45      3    363 < 2.2e-16
## as.factor(Location)    1           0.6868    83.10      3    363 < 2.2e-16
## Residuals            365
##
## (Intercept)          ***
## as.factor(Location)  ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Exp3_Manova4, test="Roy", intercept = TRUE)
```

```
##                       Df    Roy approx F num Df den Df    Pr(>F)
## (Intercept)            1 3.2103   388.45      3    363 < 2.2e-16 ***
## as.factor(Location)    1 0.6868    83.10      3    363 < 2.2e-16 ***
## Residuals            365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```