

Anova_Manova_Modelling.R

kriti

Sun Sep 30 12:05:36 2018

```
###ANOVA
```

```
#One way analysis of variance helps us understand the relationship between  
#one continuous dependent variable and one categorical independent variable.  
#When we have one continuous dependent variable and more than one independent  
#categorical variable we cannot use one way ANOVA. When we have two independent  
#categorical variable we need to use two way ANOVA. When we have more than two  
#categorical independent variables we need to use N way ANOVA.
```

```
#ONE way Anova
```

```
#Null Hypothesis-The mean average of the dependent variable is equal for all groups
```

```
#mu1=mu2=mu3
```

```
#this also implies that the categorical variable which will be the independent variable
```

```
#has no effect on the dependent continuous variable
```

```
## TWO WAY ANOVA- with two categorical variables
```

```
getwd()
```

```
## [1] "C:/Users/kriti/Documents"
```

```
data<-read.csv("moth-trap-experiment.csv")
```

```
summary(data)
```

```
##  number.of.moths  location  type.of.lure  
##  Min.    :12.00   Ground:15   Chemical:20  
##  1st Qu.:18.00   Lower :15   Scent   :20  
##  Median :27.00   Middle:15   Sugar   :20  
##  Mean   :26.68   Top    :15  
##  3rd Qu.:35.00  
##  Max.    :44.00
```

```
library(psych)
```

```
describe(data)
```

```
##
##      vars  n  mean   sd median trimmed   mad min max range skew
## number.of.moths    1 60 26.68 9.43   27.0   26.42 12.60  12 44   32 0.17
## location*          2 60  2.50 1.13    2.5    2.50  1.48   1  4    3 0.00
## type.of.lure*       3 60  2.00 0.82    2.0    2.00  1.48   1  3    2 0.00
##
##      kurtosis   se
## number.of.moths  -1.28 1.22
## location*        -1.41 0.15
## type.of.lure*    -1.55 0.11
```

```
table(data$location,data$type.of.lure)
```

```
##
##      Chemical Scent Sugar
## Ground      5      5      5
## Lower       5      5      5
## Middle      5      5      5
## Top         5      5      5
```

```
##getting descriptive statistics by location
```

```
describeBy(data$number.of.moths,data$location)
```

```
##
## Descriptive statistics by group
## group: Ground
##      vars  n  mean   sd median trimmed   mad min max range skew kurtosis   se
## X1      1 15 19.07 5.09    18   18.85 5.93  12 29   17 0.52   -1.06 1.31
## -----
## group: Lower
##      vars  n  mean   sd median trimmed   mad min max range skew kurtosis   se
## X1      1 15 33.33 7.5    34   33.77 7.41  17 44   27 -0.6   -0.5 1.94
## -----
## group: Middle
##      vars  n mean   sd median trimmed   mad min max range skew kurtosis
## X1      1 15  31 9.79    36   31.46 11.86  12 44   32 -0.39   -1.29
##      se
## X1 2.53
## -----
## group: Top
##      vars  n  mean   sd median trimmed   mad min max range skew kurtosis   se
## X1      1 15 23.33 7.41    21   23.23 8.9   13 35   22 0.24   -1.63 1.91
```

```
##getting descriptive statistics by location
```

```
describeBy(data$number.of.moths,data$type.of.lure)
```

```
##
## Descriptive statistics by group
## group: Chemical
##   vars  n mean   sd median trimmed  mad min max range  skew kurtosis  se
## X1    1 20 27.5 9.06  28.5  27.44 12.6  14  41   27 -0.01   -1.61 2.03
## -----
## group: Scent
##   vars  n mean   sd median trimmed  mad min max range  skew kurtosis
## X1    1 20 24.75 10.29    22  24.06 11.12  12  44   32 0.43    -1.2
##   se
## X1 2.3
## -----
## group: Sugar
##   vars  n mean   sd median trimmed  mad min max range  skew kurtosis  se
## X1    1 20 27.8 9.06    28  27.44 11.12  15  44   29 0.14   -1.35 2.03
```

```
##creating box plots
```

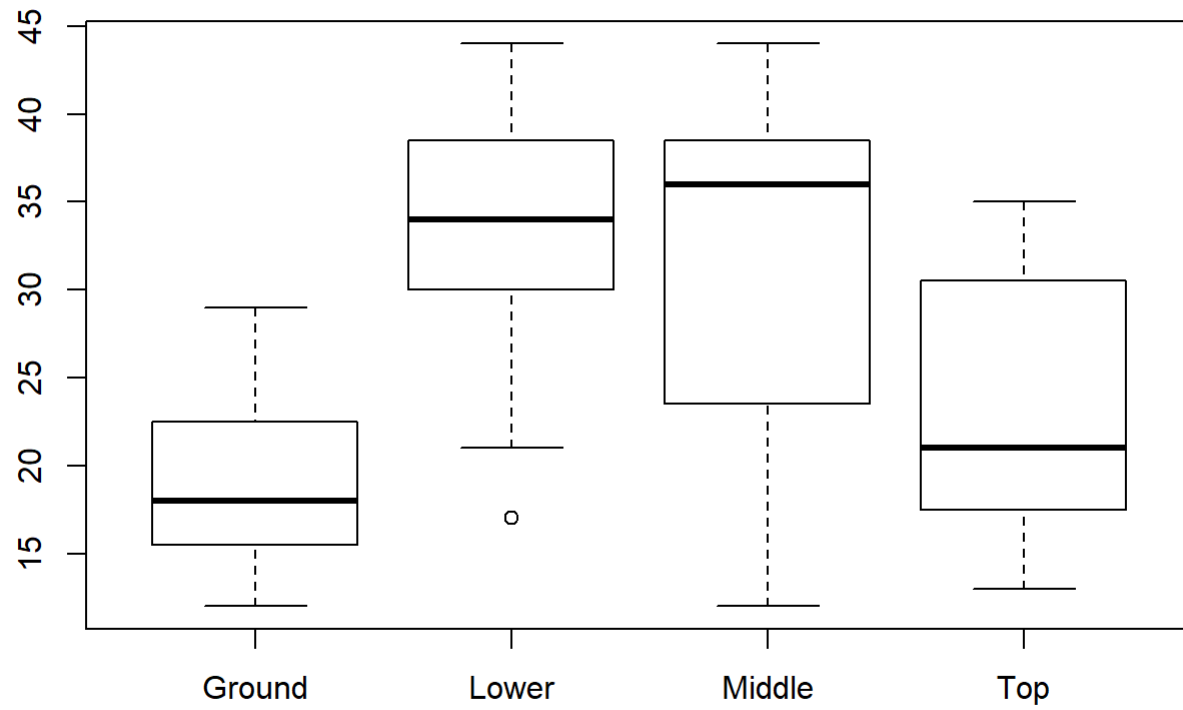
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

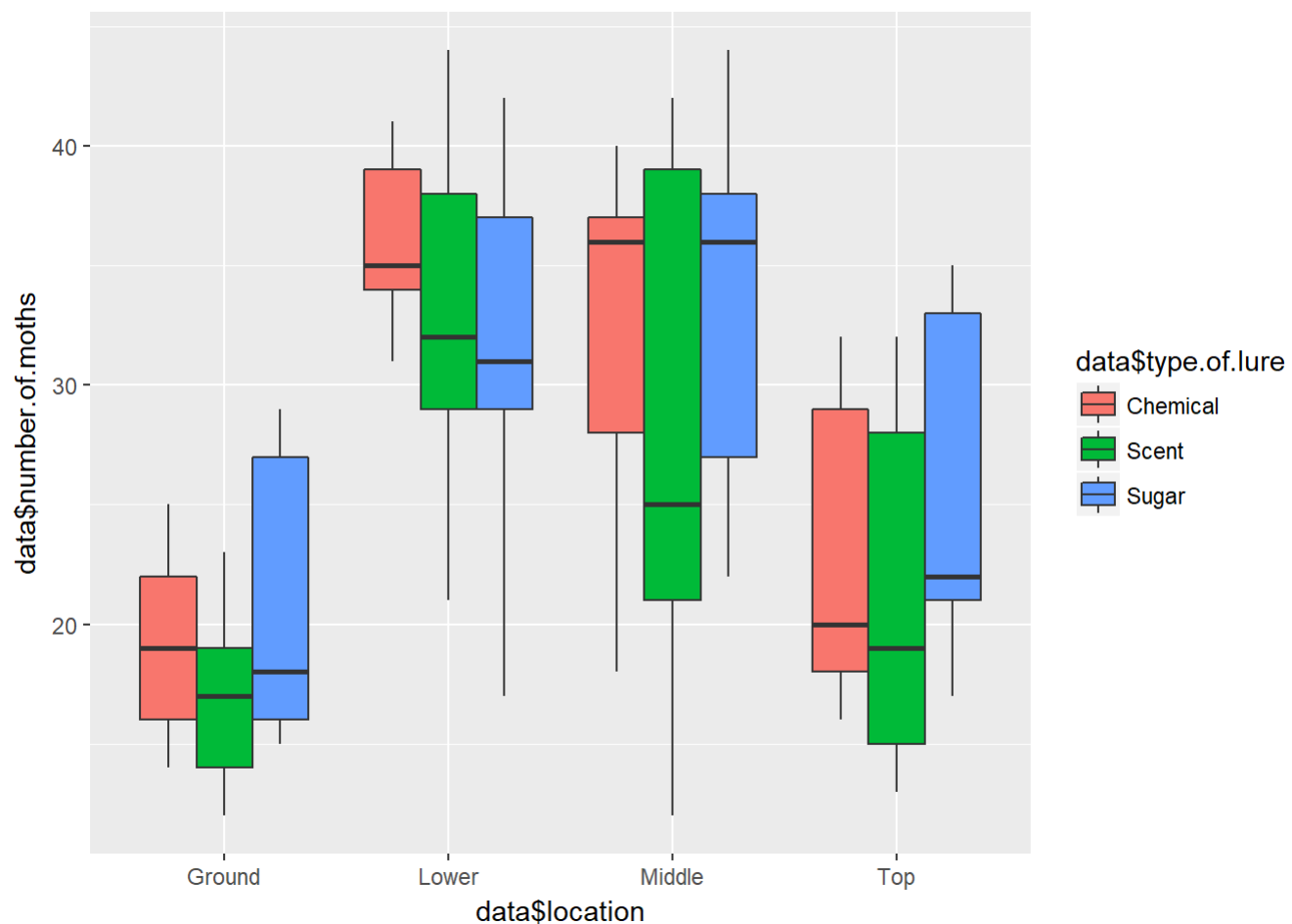
```
##
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
##   %+%, alpha
```

```
plot(x=data$location,y=data$number.of.moths)
```



```
ggplot(data,aes(y=data$number.of.moths,x=data$location,fill=data$type.of.lure))+geom_boxplot()
```



```
##checking for normality and equality of variance
```

```
shapiro.test(data$number.of.moths)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$number.of.moths
## W = 0.94533, p-value = 0.009448
```

```
##we reject the null hypothesis and hence it is not normally dist
```

```
library(car)
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':
##
##  logit
```

```
leveneTest(data$number.of.moths~data$location*data$type.of.lure)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 11  0.6377 0.7875
##      48
```

##the variances are equal

```
##doing a log transformation to make it normally dist
data$log_moth<-log(data$number.of.moths)
shapiro.test(data$log_moth)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$log_moth
## W = 0.94746, p-value = 0.01185
```

##the log tranformation is not very effective

##performing a power test

```
library(pwr)
```

```
## Warning: package 'pwr' was built under R version 3.4.4
```

```
#our design has 2 factors with 3 and 4 levels, we have 5 observations in each group
# our df for the mean squared term is 4*3(5-1)=48
#We choose a medium effect size of 0.25

pwr.f2.test(u=2,v=48,f2=(0.25*0.25))
```

```
##
##      Multiple regression power calculation
##
##              u = 2
##              v = 48
##              f2 = 0.0625
##      sig.level = 0.05
##              power = 0.3210203
```

#to check sample size

```
pwr.anova.test(k=2,f=0.25,sig.level=0.05,power=0.3210203)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##              k = 2
##              n = 18.86313
##              f = 0.25
##      sig.level = 0.05
##      power = 0.3210203
##
## NOTE: n is number in each group
```

```
##performing anova
```

```
anova_model<-aov(data$log_moth~data$location*data$type.of.lure)
summary(anova_model)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## data$location      3   2.966    0.9886    9.828 3.64e-05 ***
## data$type.of.lure    2   0.262    0.1308    1.301   0.282
## data$location:data$type.of.lure 6   0.196    0.0326    0.324   0.921
## Residuals          48   4.828    0.1006
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#NULL hypothesis
```

```
#the population means are equal for the first categorical variable
```

```
#the population means are equal for the second categorical variable
```

```
#The two categorical variable are independent of each other
```

```
#we see that we reject the first null hypothesis so Location has some effect
```

```
#we accept the second hypothesis hence lure has no effect
```

```
#we accept the third null hypothesis basically that the two variables are
```

```
#independent of each other
```

```
#####
```

```
#more on ANOVA-N way ANOVA
```

```
#reading data
```

```
data2<-read.csv("Quinn.csv")
```

```
library(car)
```

```
library(psych)
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:car':
##
##      recode
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
library(lattice)
library(ggplot2)
```

```
##checking data
```

```
summary(data2)
```

```
##  DENSITY    SEASON      EGGS
##  L1:6    spring:12  Min.   :0.3560
##  L2:6    summer:12  1st Qu.:0.9165
##  L3:6                Median :1.4330
##  L4:6                Mean   :1.4717
##                  3rd Qu.:1.9227
##                  Max.    :2.8750
```

```
table(data2$DENSITY,data2$SEASON)
```

```
##
##      spring summer
##  L1         3      3
##  L2         3      3
##  L3         3      3
##  L4         3      3
```



```
###we see that the data is balanced
```

```
##THE NULL Hypothesis in these cases are
```

```
#The season has no effect on eggs.The means of the dist is same across the groups based on season
```

```
#The density has no effect on eggs.The means of the dist is same across the groups based on season
```

```
#The two categorical variables do not effect together on the eggs
```

```
#descriptive statistic summaries and data visualization;
```

```
##checking dist of eggs across multiple densities
```

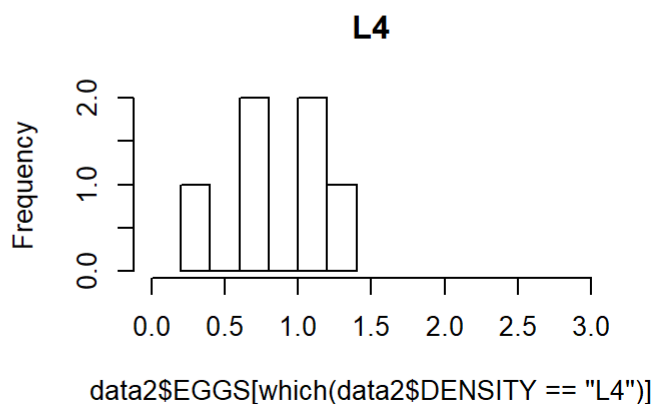
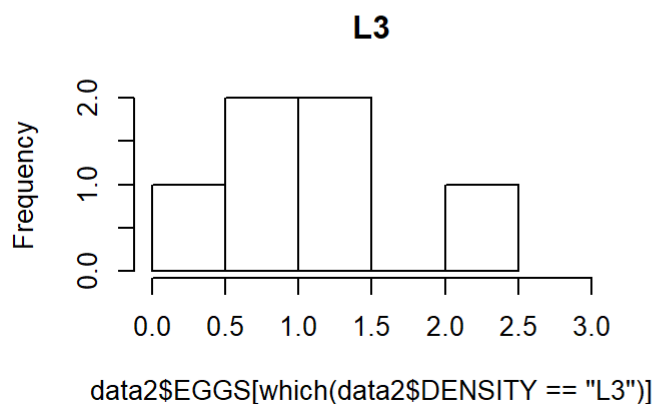
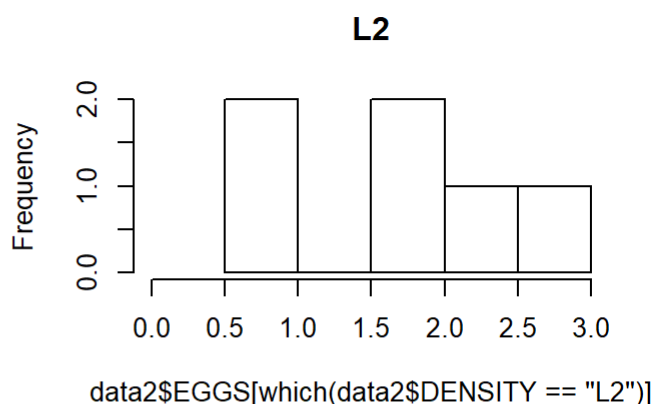
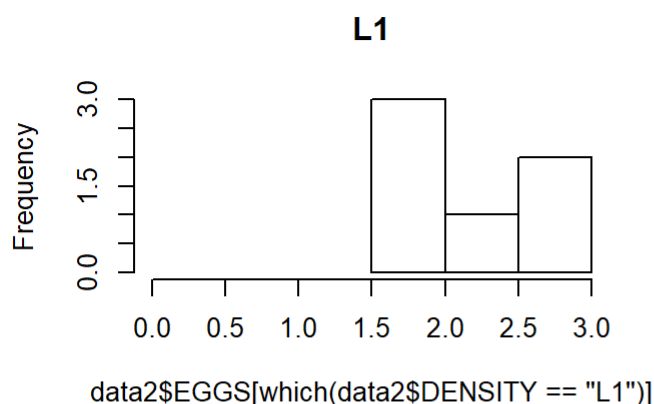
```
par(mfrow=c(2,2))
```

```
hist(data2$EGGS[which(data2$DENSITY=="L1")],xlim = c(0, 3),main="L1")
```

```
hist(data2$EGGS[which(data2$DENSITY=="L2")],xlim = c(0, 3),main="L2")
```

```
hist(data2$EGGS[which(data2$DENSITY=="L3")],xlim = c(0, 3),main="L3")
```

```
hist(data2$EGGS[which(data2$DENSITY=="L4")],xlim = c(0, 3),main="L4")
```

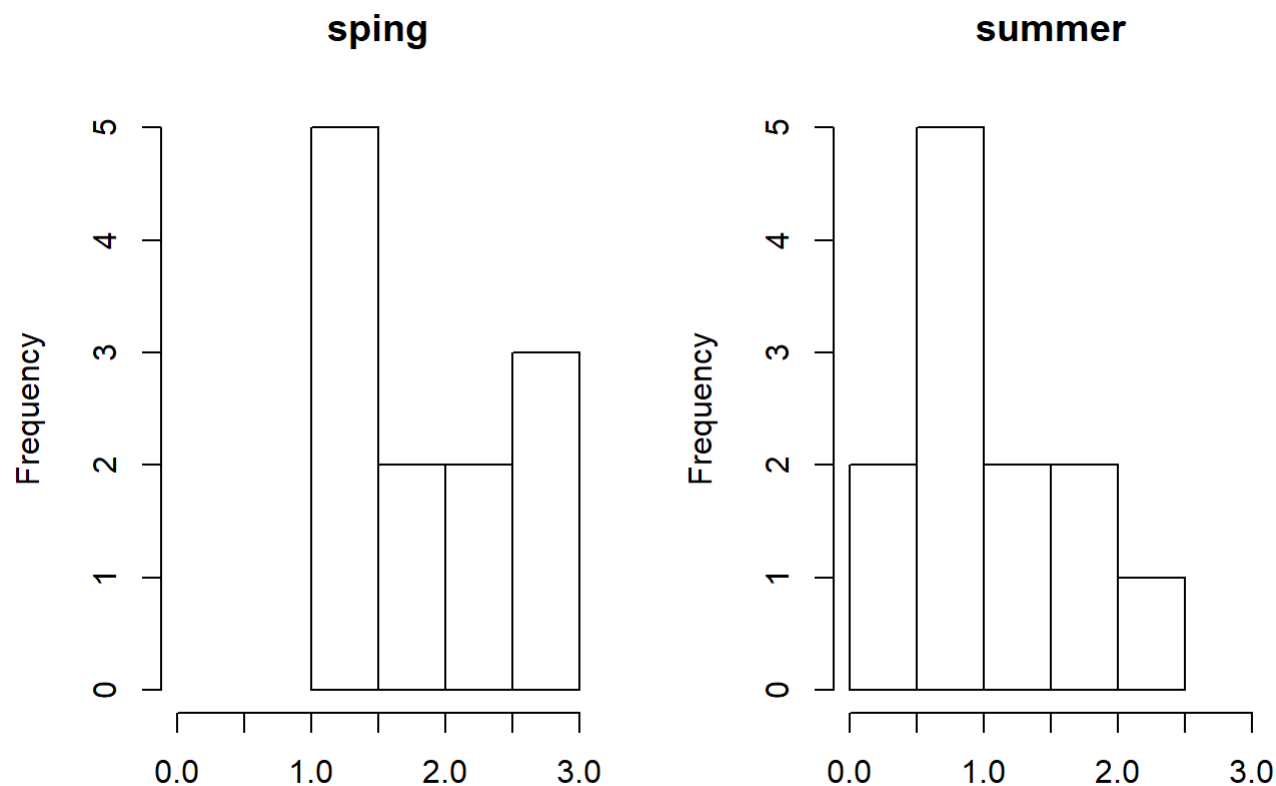


```

par(mfrow=c(1,2))
##checking dist of eggs across multiple season

hist(data2$EGGS[which(data2$SEASON=="spring")],xlim = c(0, 3),main="spring")
hist(data2$EGGS[which(data2$SEASON=="summer")],xlim = c(0, 3),main="summer")

```



```

data2$EGGS[which(data2$SEASON == "spring")]
data2$EGGS[which(data2$SEASON == "summer")]

```

```

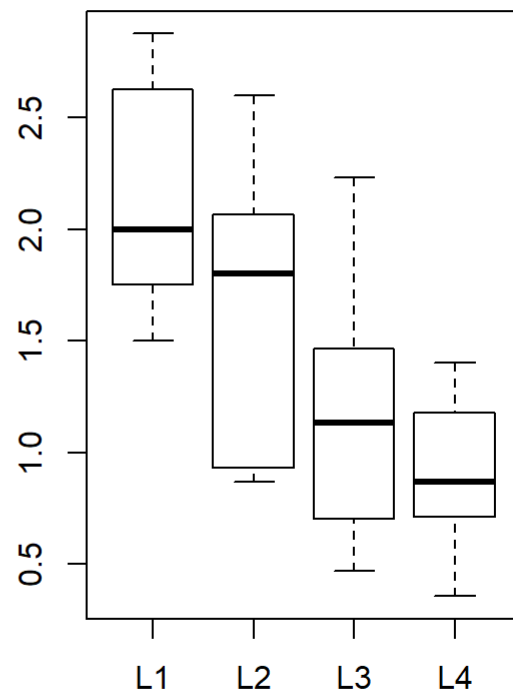
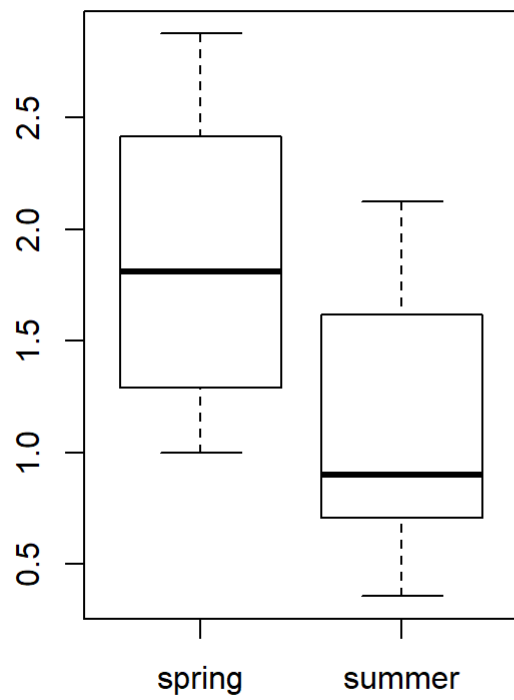
##creating boxplots

```

```

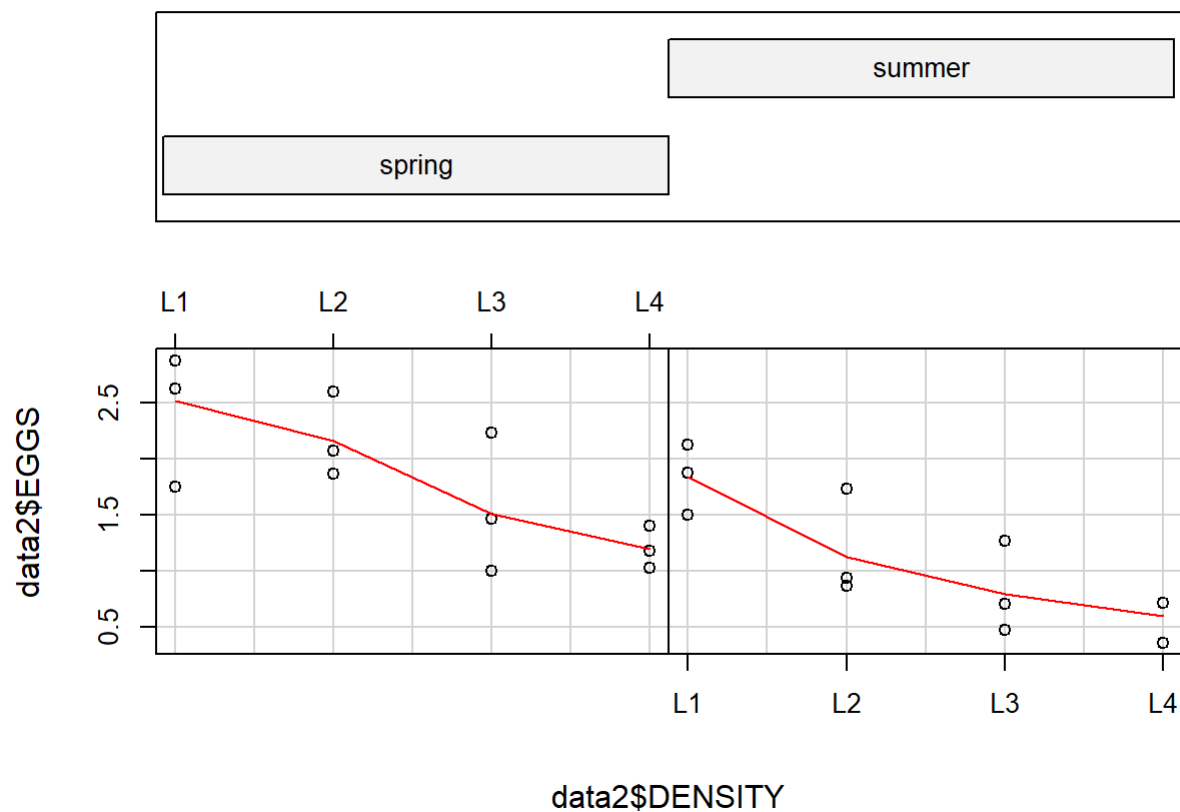
boxplot(data2$EGGS~data2$SEASON)
boxplot(data2$EGGS~data2$DENSITY)

```



```
par(mfrow=c(1,1))  
coplot(data2$EGGS~data2$DENSITY|data2$SEASON,panel=panel.smooth)
```

Given : data2\$SEASON



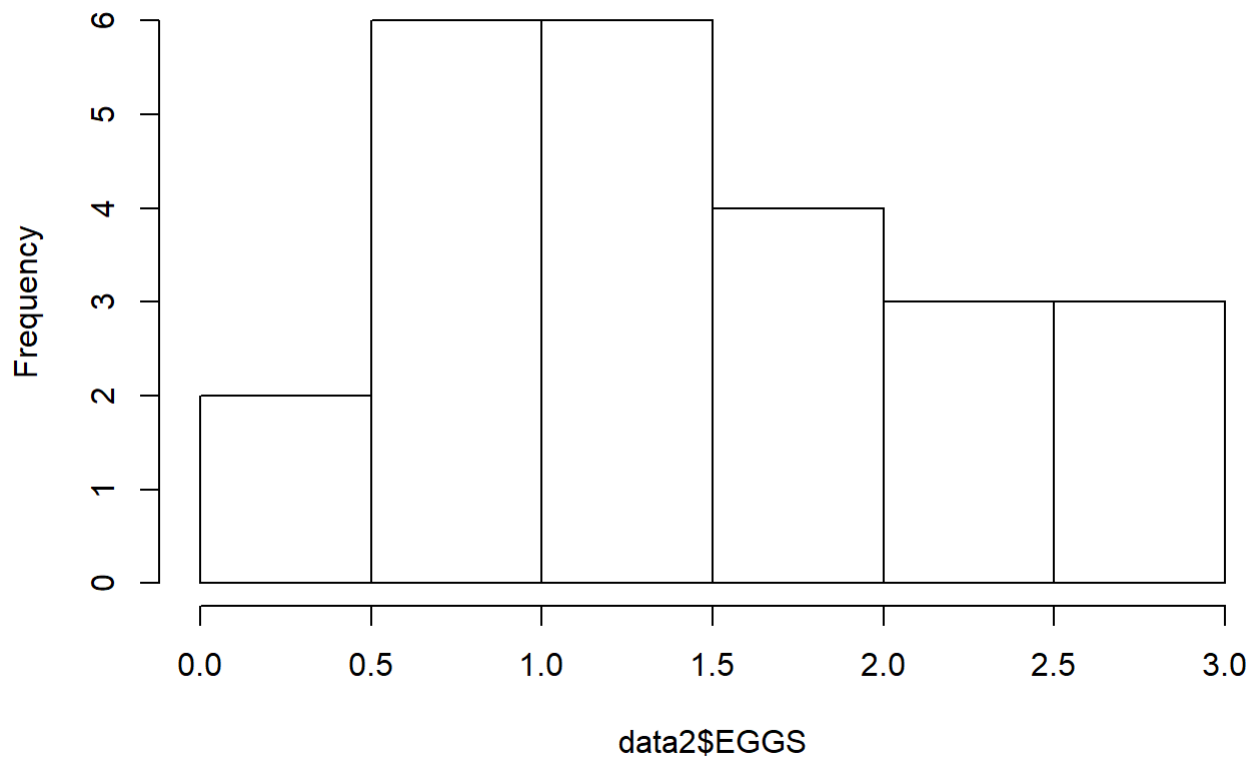
##checking if the continuiuos variable is normally distributed

```
shapiro.test(data2$EGGS)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data2$EGGS
## W = 0.96449, p-value = 0.5351
```

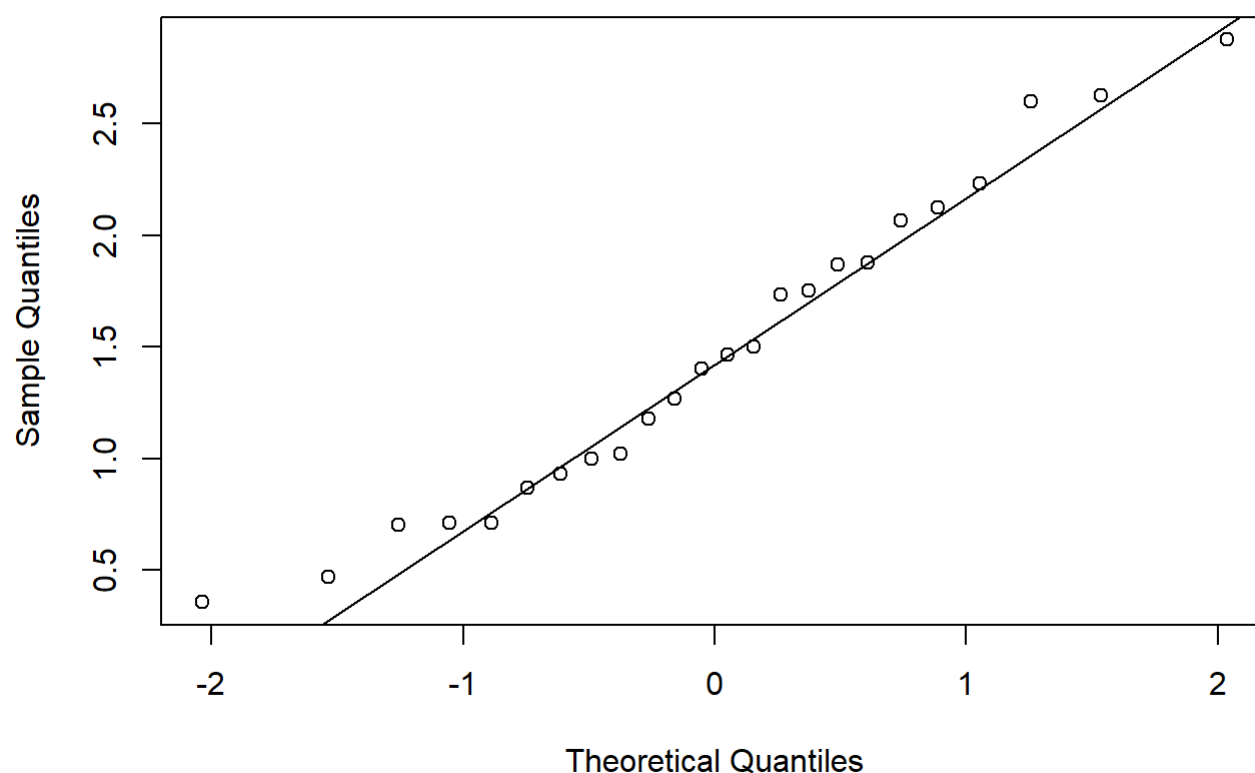
```
hist(data2$EGGS)
```

Histogram of data2\$EGGS



```
qqnorm(data2$EGGS)  
qqline(data2$EGGS)
```

Normal Q-Q Plot



```
##its normally dist
```

```
##checking equality of variances
```

```
leveneTest(data2$EGGS~data2$DENSITY)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##      Df F value Pr(>F)
```

```
## group 3  0.4125 0.7458
```

```
##      20
```

```
leveneTest(data2$EGGS~data2$SEASON)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##      Df F value Pr(>F)
```

```
## group 1  0.2252 0.6398
```

```
##      22
```

```
leveneTest(data2$EGGS~data2$DENSITY*data2$SEASON)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 7  0.3337 0.9268
##      16
```

##the is equality of variance in all cases

##we can perform parametric tests

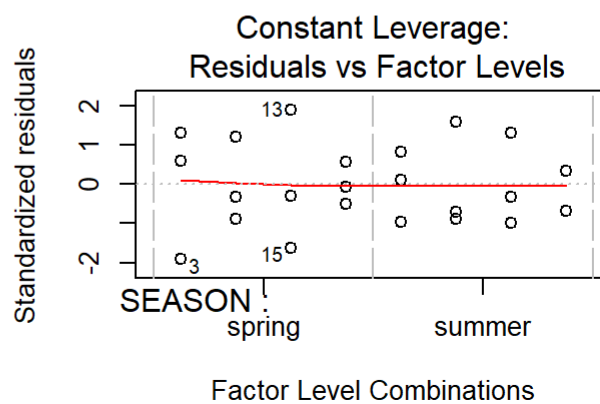
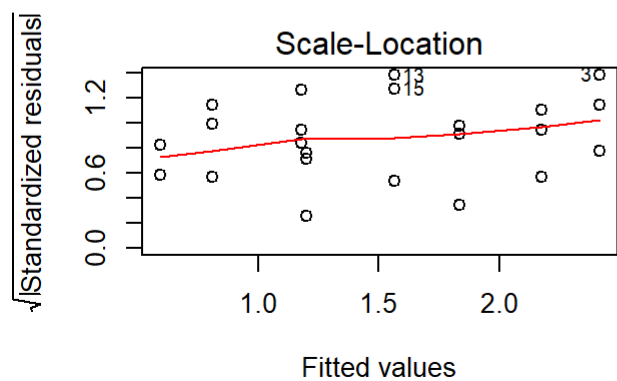
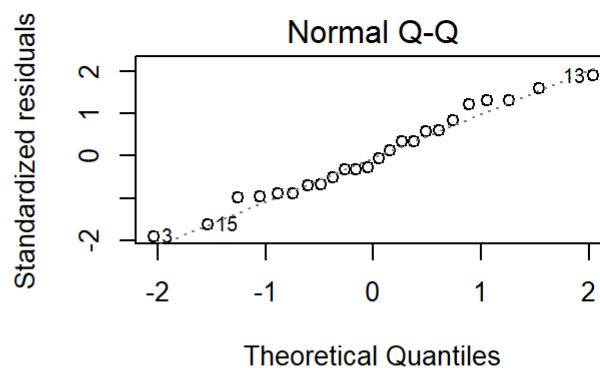
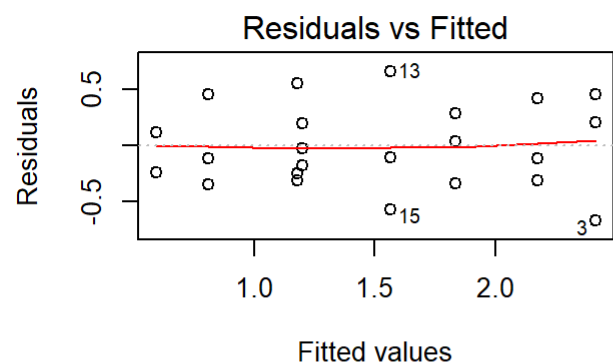
```
library(pwr)
anova1<-aov( EGGS ~ SEASON*DENSITY ,data=data2)
summary(anova1)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## SEASON          1  3.250    3.250   17.842 0.000645 ***
## DENSITY          3  5.284    1.761    9.669 0.000704 ***
## SEASON:DENSITY   3  0.165    0.055    0.301 0.823955
## Residuals       16  2.915    0.182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova2<-aov( EGGS ~ SEASON+DENSITY ,data=data2)
summary(anova2)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## SEASON          1  3.250    3.250   20.05 0.000258 ***
## DENSITY          3  5.284    1.761   10.87 0.000223 ***
## Residuals       19  3.079    0.162
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow=c(2,2))
plot(anova1)
```



###CONCLUSION

*#Season and density have an effect on the dependent variable eggs but their combined
#intererection term has no effect*

#####

#MANOVA _ MULTIVARIATE ANALYSIS OF VARIANCE

##Principle of MANOVA

*#In Multi-variate Analysis of Variance we have multiple dependent variables
#and one or more independent variable
#we basically test how the samples due a single categorical variables differ across
#the different dependent variable*

##NULL Hypothesis

*##Mean of all n-dimensional groups are equal...:Eg
##if we have k dependent variables and n groups in the categorical data then we have
mean $1*1=mean\ 1*2==mean\ n*k$*

*##The MANOVA test statistic compute the TOTAL sum of square by computing the
#sum of square within a group and sum of square between groups
##T=Within SS + Between SS*

##MANOVA using IRIS data

```
data<-data("iris")  
iris
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 1	5.1	3.5	1.4	0.2	setosa
## 2	4.9	3.0	1.4	0.2	setosa
## 3	4.7	3.2	1.3	0.2	setosa
## 4	4.6	3.1	1.5	0.2	setosa
## 5	5.0	3.6	1.4	0.2	setosa
## 6	5.4	3.9	1.7	0.4	setosa
## 7	4.6	3.4	1.4	0.3	setosa
## 8	5.0	3.4	1.5	0.2	setosa
## 9	4.4	2.9	1.4	0.2	setosa
## 10	4.9	3.1	1.5	0.1	setosa
## 11	5.4	3.7	1.5	0.2	setosa
## 12	4.8	3.4	1.6	0.2	setosa
## 13	4.8	3.0	1.4	0.1	setosa
## 14	4.3	3.0	1.1	0.1	setosa
## 15	5.8	4.0	1.2	0.2	setosa
## 16	5.7	4.4	1.5	0.4	setosa
## 17	5.4	3.9	1.3	0.4	setosa
## 18	5.1	3.5	1.4	0.3	setosa
## 19	5.7	3.8	1.7	0.3	setosa
## 20	5.1	3.8	1.5	0.3	setosa
## 21	5.4	3.4	1.7	0.2	setosa
## 22	5.1	3.7	1.5	0.4	setosa
## 23	4.6	3.6	1.0	0.2	setosa
## 24	5.1	3.3	1.7	0.5	setosa
## 25	4.8	3.4	1.9	0.2	setosa
## 26	5.0	3.0	1.6	0.2	setosa
## 27	5.0	3.4	1.6	0.4	setosa
## 28	5.2	3.5	1.5	0.2	setosa
## 29	5.2	3.4	1.4	0.2	setosa
## 30	4.7	3.2	1.6	0.2	setosa
## 31	4.8	3.1	1.6	0.2	setosa
## 32	5.4	3.4	1.5	0.4	setosa
## 33	5.2	4.1	1.5	0.1	setosa
## 34	5.5	4.2	1.4	0.2	setosa
## 35	4.9	3.1	1.5	0.2	setosa
## 36	5.0	3.2	1.2	0.2	setosa
## 37	5.5	3.5	1.3	0.2	setosa
## 38	4.9	3.6	1.4	0.1	setosa
## 39	4.4	3.0	1.3	0.2	setosa
## 40	5.1	3.4	1.5	0.2	setosa
## 41	5.0	3.5	1.3	0.3	setosa
## 42	4.5	2.3	1.3	0.3	setosa
## 43	4.4	3.2	1.3	0.2	setosa
## 44	5.0	3.5	1.6	0.6	setosa
## 45	5.1	3.8	1.9	0.4	setosa
## 46	4.8	3.0	1.4	0.3	setosa
## 47	5.1	3.8	1.6	0.2	setosa
## 48	4.6	3.2	1.4	0.2	setosa
## 49	5.3	3.7	1.5	0.2	setosa
## 50	5.0	3.3	1.4	0.2	setosa
## 51	7.0	3.2	4.7	1.4	versicolor
## 52	6.4	3.2	4.5	1.5	versicolor

## 53	6.9	3.1	4.9	1.5 versicolor
## 54	5.5	2.3	4.0	1.3 versicolor
## 55	6.5	2.8	4.6	1.5 versicolor
## 56	5.7	2.8	4.5	1.3 versicolor
## 57	6.3	3.3	4.7	1.6 versicolor
## 58	4.9	2.4	3.3	1.0 versicolor
## 59	6.6	2.9	4.6	1.3 versicolor
## 60	5.2	2.7	3.9	1.4 versicolor
## 61	5.0	2.0	3.5	1.0 versicolor
## 62	5.9	3.0	4.2	1.5 versicolor
## 63	6.0	2.2	4.0	1.0 versicolor
## 64	6.1	2.9	4.7	1.4 versicolor
## 65	5.6	2.9	3.6	1.3 versicolor
## 66	6.7	3.1	4.4	1.4 versicolor
## 67	5.6	3.0	4.5	1.5 versicolor
## 68	5.8	2.7	4.1	1.0 versicolor
## 69	6.2	2.2	4.5	1.5 versicolor
## 70	5.6	2.5	3.9	1.1 versicolor
## 71	5.9	3.2	4.8	1.8 versicolor
## 72	6.1	2.8	4.0	1.3 versicolor
## 73	6.3	2.5	4.9	1.5 versicolor
## 74	6.1	2.8	4.7	1.2 versicolor
## 75	6.4	2.9	4.3	1.3 versicolor
## 76	6.6	3.0	4.4	1.4 versicolor
## 77	6.8	2.8	4.8	1.4 versicolor
## 78	6.7	3.0	5.0	1.7 versicolor
## 79	6.0	2.9	4.5	1.5 versicolor
## 80	5.7	2.6	3.5	1.0 versicolor
## 81	5.5	2.4	3.8	1.1 versicolor
## 82	5.5	2.4	3.7	1.0 versicolor
## 83	5.8	2.7	3.9	1.2 versicolor
## 84	6.0	2.7	5.1	1.6 versicolor
## 85	5.4	3.0	4.5	1.5 versicolor
## 86	6.0	3.4	4.5	1.6 versicolor
## 87	6.7	3.1	4.7	1.5 versicolor
## 88	6.3	2.3	4.4	1.3 versicolor
## 89	5.6	3.0	4.1	1.3 versicolor
## 90	5.5	2.5	4.0	1.3 versicolor
## 91	5.5	2.6	4.4	1.2 versicolor
## 92	6.1	3.0	4.6	1.4 versicolor
## 93	5.8	2.6	4.0	1.2 versicolor
## 94	5.0	2.3	3.3	1.0 versicolor
## 95	5.6	2.7	4.2	1.3 versicolor
## 96	5.7	3.0	4.2	1.2 versicolor
## 97	5.7	2.9	4.2	1.3 versicolor
## 98	6.2	2.9	4.3	1.3 versicolor
## 99	5.1	2.5	3.0	1.1 versicolor
## 100	5.7	2.8	4.1	1.3 versicolor
## 101	6.3	3.3	6.0	2.5 virginica
## 102	5.8	2.7	5.1	1.9 virginica
## 103	7.1	3.0	5.9	2.1 virginica
## 104	6.3	2.9	5.6	1.8 virginica
## 105	6.5	3.0	5.8	2.2 virginica
## 106	7.6	3.0	6.6	2.1 virginica

## 107	4.9	2.5	4.5	1.7	virginica
## 108	7.3	2.9	6.3	1.8	virginica
## 109	6.7	2.5	5.8	1.8	virginica
## 110	7.2	3.6	6.1	2.5	virginica
## 111	6.5	3.2	5.1	2.0	virginica
## 112	6.4	2.7	5.3	1.9	virginica
## 113	6.8	3.0	5.5	2.1	virginica
## 114	5.7	2.5	5.0	2.0	virginica
## 115	5.8	2.8	5.1	2.4	virginica
## 116	6.4	3.2	5.3	2.3	virginica
## 117	6.5	3.0	5.5	1.8	virginica
## 118	7.7	3.8	6.7	2.2	virginica
## 119	7.7	2.6	6.9	2.3	virginica
## 120	6.0	2.2	5.0	1.5	virginica
## 121	6.9	3.2	5.7	2.3	virginica
## 122	5.6	2.8	4.9	2.0	virginica
## 123	7.7	2.8	6.7	2.0	virginica
## 124	6.3	2.7	4.9	1.8	virginica
## 125	6.7	3.3	5.7	2.1	virginica
## 126	7.2	3.2	6.0	1.8	virginica
## 127	6.2	2.8	4.8	1.8	virginica
## 128	6.1	3.0	4.9	1.8	virginica
## 129	6.4	2.8	5.6	2.1	virginica
## 130	7.2	3.0	5.8	1.6	virginica
## 131	7.4	2.8	6.1	1.9	virginica
## 132	7.9	3.8	6.4	2.0	virginica
## 133	6.4	2.8	5.6	2.2	virginica
## 134	6.3	2.8	5.1	1.5	virginica
## 135	6.1	2.6	5.6	1.4	virginica
## 136	7.7	3.0	6.1	2.3	virginica
## 137	6.3	3.4	5.6	2.4	virginica
## 138	6.4	3.1	5.5	1.8	virginica
## 139	6.0	3.0	4.8	1.8	virginica
## 140	6.9	3.1	5.4	2.1	virginica
## 141	6.7	3.1	5.6	2.4	virginica
## 142	6.9	3.1	5.1	2.3	virginica
## 143	5.8	2.7	5.1	1.9	virginica
## 144	6.8	3.2	5.9	2.3	virginica
## 145	6.7	3.3	5.7	2.5	virginica
## 146	6.7	3.0	5.2	2.3	virginica
## 147	6.3	2.5	5.0	1.9	virginica
## 148	6.5	3.0	5.2	2.0	virginica
## 149	6.2	3.4	5.4	2.3	virginica
## 150	5.9	3.0	5.1	1.8	virginica

```
summary(iris)
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300
## Median :5.800 Median :3.000 Median :4.350 Median :1.300
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500
## Species
## setosa :50
## versicolor:50
## virginica :50
##
##
##
```

```
library(psych)
describe(iris)
```

```
## vars n mean sd median trimmed mad min max range skew
## Sepal.Length 1 150 5.84 0.83 5.80 5.81 1.04 4.3 7.9 3.6 0.31
## Sepal.Width 2 150 3.06 0.44 3.00 3.04 0.44 2.0 4.4 2.4 0.31
## Petal.Length 3 150 3.76 1.77 4.35 3.76 1.85 1.0 6.9 5.9 -0.27
## Petal.Width 4 150 1.20 0.76 1.30 1.18 1.04 0.1 2.5 2.4 -0.10
## Species* 5 150 2.00 0.82 2.00 2.00 1.48 1.0 3.0 2.0 0.00
## kurtosis se
## Sepal.Length -0.61 0.07
## Sepal.Width 0.14 0.04
## Petal.Length -1.42 0.14
## Petal.Width -1.36 0.06
## Species* -1.52 0.07
```

```
##exploring the data
```

```
describeBy(iris[-5],iris$Species)
```

```
##
## Descriptive statistics by group
## group: setosa
##
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew
Sepal.Length	1	50	5.01	0.35	5.0	5.00	0.30	4.3	5.8	1.5	0.11
Sepal.Width	2	50	3.43	0.38	3.4	3.42	0.37	2.3	4.4	2.1	0.04
Petal.Length	3	50	1.46	0.17	1.5	1.46	0.15	1.0	1.9	0.9	0.10
Petal.Width	4	50	0.25	0.11	0.2	0.24	0.00	0.1	0.6	0.5	1.18

```
##
```

	kurtosis	se
Sepal.Length	-0.45	0.05
Sepal.Width	0.60	0.05
Petal.Length	0.65	0.02
Petal.Width	1.26	0.01

```
## -----
## group: versicolor
##
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew
Sepal.Length	1	50	5.94	0.52	5.90	5.94	0.52	4.9	7.0	2.1	0.10
Sepal.Width	2	50	2.77	0.31	2.80	2.78	0.30	2.0	3.4	1.4	-0.34
Petal.Length	3	50	4.26	0.47	4.35	4.29	0.52	3.0	5.1	2.1	-0.57
Petal.Width	4	50	1.33	0.20	1.30	1.32	0.22	1.0	1.8	0.8	-0.03

```
##
```

	kurtosis	se
Sepal.Length	-0.69	0.07
Sepal.Width	-0.55	0.04
Petal.Length	-0.19	0.07
Petal.Width	-0.59	0.03

```
## -----
## group: virginica
##
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew
Sepal.Length	1	50	6.59	0.64	6.50	6.57	0.59	4.9	7.9	3.0	0.11
Sepal.Width	2	50	2.97	0.32	3.00	2.96	0.30	2.2	3.8	1.6	0.34
Petal.Length	3	50	5.55	0.55	5.55	5.51	0.67	4.5	6.9	2.4	0.52
Petal.Width	4	50	2.03	0.27	2.00	2.03	0.30	1.4	2.5	1.1	-0.12

```
##
```

	kurtosis	se
Sepal.Length	-0.20	0.09
Sepal.Width	0.38	0.05
Petal.Length	-0.37	0.08
Petal.Width	-0.75	0.04

```
##checking if the distributions are normal
```

```
shapiro.test(iris$Sepal.Length)
```

```
##
## Shapiro-Wilk normality test
##
## data: iris$Sepal.Length
## W = 0.97609, p-value = 0.01018
```

```
shapiro.test(iris$Petal.Length)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: iris$Petal.Length  
## W = 0.87627, p-value = 7.412e-10
```

```
shapiro.test(iris$Sepal.Width)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: iris$Sepal.Width  
## W = 0.98492, p-value = 0.1012
```

```
shapiro.test(iris$Petal.Width)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: iris$Petal.Width  
## W = 0.90183, p-value = 1.68e-08
```

```
#to create a MANOVA Model we use species as the independent variable and the two  
#dependent variables are Petal Length and Sepal Length
```

```
##checking correlations  
cor(iris$Sepal.Length,iris$Petal.Length)
```

```
## [1] 0.8717538
```

```
cor(iris[-5])
```

```
##           Sepal.Length Sepal.Width Petal.Length Petal.Width  
## Sepal.Length    1.0000000 -0.1175698    0.8717538    0.8179411  
## Sepal.Width     -0.1175698    1.0000000   -0.4284401   -0.3661259  
## Petal.Length     0.8717538  -0.4284401    1.0000000    0.9628654  
## Petal.Width      0.8179411  -0.3661259    0.9628654    1.0000000
```

```
##checking homogeniaty of variances  
leveneTest(iris$Sepal.Length~iris$Species)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value    Pr(>F)
## group  2  6.3527 0.002259 **
##      147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
bartlett.test(iris$Sepal.Length~iris$Species)
```

```
##
## Bartlett test of homogeneity of variances
##
## data:  iris$Sepal.Length by iris$Species
## Bartlett's K-squared = 16.006, df = 2, p-value = 0.0003345
```

```
leveneTest(iris$Petal.Length~iris$Species)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value    Pr(>F)
## group  2  19.48 3.129e-08 ***
##      147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
bartlett.test(iris$Petal.Length~iris$Species)
```

```
##
## Bartlett test of homogeneity of variances
##
## data:  iris$Petal.Length by iris$Species
## Bartlett's K-squared = 55.423, df = 2, p-value = 9.229e-13
```

```
##we see that the homogeneity of variances condition is not satisfied

##some important things to note
#if sample size is greater than 30 then some assumptions can be ignored
#If group sizes are over 30, then the MANOVA is robust against violations of
#homogeneity of variance-covariance matrices assumption.
#we can report Pillai's trace statistic in this case instead of Wilk's Lambda

##performing MANOVA

manova_model<-manova(cbind(iris$Sepal.Length,iris$Petal.Length)~iris$Species)

manova_model
```



```
## Call:
##   manova(cbind(iris$Sepal.Length, iris$Petal.Length) ~ iris$Species)
##
## Terms:
##               iris$Species Residuals
## resp 1           63.2121    38.9562
## resp 2          437.1028    27.2226
## Deg. of Freedom      2        147
##
## Residual standard errors: 0.5147894 0.4303345
## Estimated effects may be unbalanced
```

```
summary(manova_model)
```

```
##              Df Pillai approx F num Df den Df    Pr(>F)
## iris$Species  2 0.9885   71.829      4   294 < 2.2e-16 ***
## Residuals    147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##the test is coming significant hence we reject the null hypothesis
##hence the species has certain effect on the petal and sepal length

```
summary.aov(manova_model)
```

```
## Response 1 :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## iris$Species  2 63.212  31.606  119.26 < 2.2e-16 ***
## Residuals    147 38.956   0.265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response 2 :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## iris$Species  2 437.10 218.551  1180.2 < 2.2e-16 ***
## Residuals    147  27.22   0.185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#the samples in petal length and sepal length are statistically
#different from each other
```

```
#The four test statistics that can be used are Wilk's Lambda, Pillai trace,
#Hotelling-Lawley trace and Roy's maximum root. Among the four test
#statistics Pillai is Least affected by any violations in assumptions
#but Wilk's is the most commonly used.
```

```
summary(manova_model,test="Wilks")
```

```
##              Df      Wilks approx F num Df den Df      Pr(>F)
## iris$Species  2 0.039878   292.56      4    292 < 2.2e-16 ***
## Residuals    147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(manova_model,test="Pillai")
```

```
##              Df Pillai approx F num Df den Df      Pr(>F)
## iris$Species  2 0.9885   71.829      4    294 < 2.2e-16 ***
## Residuals    147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(manova_model, test="Hotelling")
```

```
##              Df Hotelling-Lawley approx F num Df den Df      Pr(>F)
## iris$Species  2      23.365   846.97      4    290 < 2.2e-16 ***
## Residuals    147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(manova_model, test="Roy")
```

```
##              Df      Roy approx F num Df den Df      Pr(>F)
## iris$Species  2 23.334   1715.1      2    147 < 2.2e-16 ***
## Residuals    147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
head(manova_model$residuals)
```

```
##      [,1]  [,2]
## 1  0.094 -0.062
## 2 -0.106 -0.062
## 3 -0.306 -0.162
## 4 -0.406  0.038
## 5 -0.006 -0.062
## 6  0.394  0.238
```

##from all the four tests we can see that we fail to accept the null Hypothesis

#more information

#As you might expect, we use a multivariate analysis of variance (MANOVA) when we have one or more

#categorical independent variables with two or more treatment levels AND more than one continuous

#response variable (that's what makes it "multivariate"). MANOVA is commonly used species communities,

#where the frequency of each species represents a response variable.

#MANOVA is based on the same principles as a discriminant analysis, which is a rotational technique

#designed to maximise variance between groups, rather than across an entire data set (which would be a

#principal component analysis). Essentially, MANOVA examined the variance explained between your groups (treatment levels) by comparing components (called "eigenvectors") which separate the data.

#Generally, MANOVA is used for two things: 1) to implement a single inferential test on multiple dependent

#variables which may be correlated, and more frequently 2) determine the change in arrangement of

#dependent variables in response to the independent variable(s).