Manova Statistical Analysis

Faraz Younus

2023-05-27

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

attach(df)  
summary(df)

## cancer gender N/a\_col st1   
## Min. :1.000 Min. :1.000 Min. :38.00 Min. : 20.0   
## 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:56.50 1st Qu.: 119.5   
## Median :3.000 Median :1.000 Median :66.00 Median : 245.0   
## Mean :2.968 Mean :1.381 Mean :64.08 Mean : 475.6   
## 3rd Qu.:4.000 3rd Qu.:2.000 3rd Qu.:72.50 3rd Qu.: 452.5   
## Max. :6.000 Max. :2.000 Max. :93.00 Max. :4288.0   
## st2 st3 st4   
## Min. : 18.0 Min. : 8.0 Min. :10.00   
## 1st Qu.: 109.5 1st Qu.: 50.0 1st Qu.:18.50   
## Median : 199.0 Median : 124.0 Median :30.00   
## Mean : 270.1 Mean : 200.4 Mean :33.97   
## 3rd Qu.: 392.0 3rd Qu.: 260.0 3rd Qu.:41.00   
## Max. :1056.0 Max. :1267.0 Max. :91.00

## Including Plots

You can also embed plots, for example: Below we can see that the distribution of many varaibles isn’t normal.

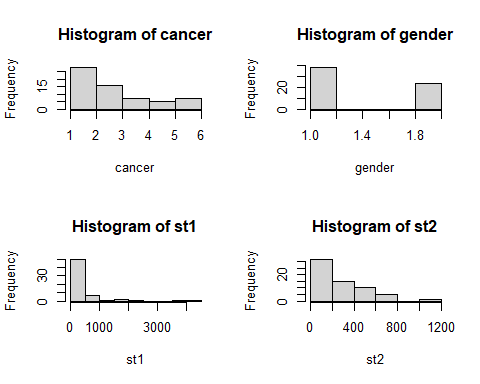
MANOVA in R can be used to determine whether there are significant differences between multiple groups on multiple dependent variables. However, it does not tell you which groups differ from the rest. This can be determined using a post-hoc test, such as Linear Discriminant Analysis (LDA).

MANOVA has several assumptions, including:

Multivariate normality: Each combination of independent and dependent variables should have a multivariate normal distribution. This can be verified using Shapiro-Wilk’s test. Linearity: Dependent variables should have a linear relationship with each group (factor) of the independent variable. No multicollinearity: Dependent variables should not have very high correlations. No outliers: There should not be any outliers in the dependent variables.

This isn’t important in our case.

par(mfrow = c(2, 2))  
# This code allows use to get histograms of the distributions of different variables  
hist(cancer)  
hist(gender)  
hist(st1)  
hist(st2)



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot. To conduct a manova in R we can simply

library(car)

## Loading required package: carData

# Create a factor variable for the independent variable  
  
group <- factor(cancer)  
  
#group<- levels(factor)  
  
# Perform the MANOVA  
result<-manova(cbind(st1, st2, st3, st4) ~ group, data = df)  
  
# Print the MANOVA results  
summary(result)

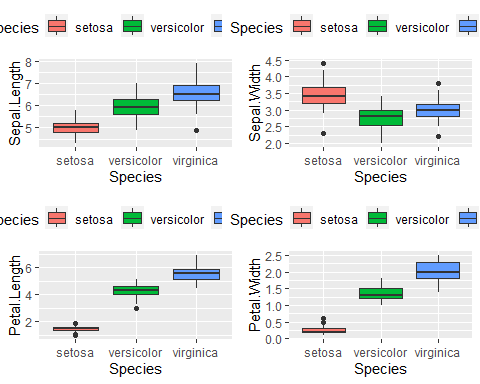
## Df Pillai approx F num Df den Df Pr(>F)   
## group 5 0.85343 3.092 20 228 2.189e-05 \*\*\*  
## Residuals 57   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Pillai: The Pillai’s trace statistic is a multivariate test statistic used in MANOVA. It measures the overall effect of the grouping variable on the dependent variables. The reported Pillai’s trace value is 0.85343.

Pr(>F): This column displays the p-value associated with the approximate F-test. It indicates the probability of observing the obtained F-statistic (or a more extreme value) under the null hypothesis of no group effect. In this case, the p-value is 2.189e-05, which is very small, suggesting strong evidence against the null hypothesis.

Overall, the MANOVA results suggest that there is a significant multivariate effect of the grouping variable on the dependent variables (st1, st2, st3, st4).

# Perform MANOVA  
library(ggplot2)  
library(gridExtra)  
  
box\_sl <- ggplot(iris, aes(x = Species, y = Sepal.Length, fill = Species)) +  
 geom\_boxplot() +  
 theme(legend.position = "top")  
box\_sw <- ggplot(iris, aes(x = Species, y = Sepal.Width, fill = Species)) +  
 geom\_boxplot() +  
 theme(legend.position = "top")  
box\_pl <- ggplot(iris, aes(x = Species, y = Petal.Length, fill = Species)) +  
 geom\_boxplot() +  
 theme(legend.position = "top")  
box\_pw <- ggplot(iris, aes(x = Species, y = Petal.Width, fill = Species)) +  
 geom\_boxplot() +  
 theme(legend.position = "top")  
  
grid.arrange(box\_sl, box\_sw, box\_pl, box\_pw, ncol = 2, nrow = 2)



dependent\_vars <- cbind(iris$Sepal.Length, iris$Sepal.Width, iris$Petal.Length, iris$Petal.Width)  
independent\_var <- iris$Species  
  
manova\_model <- manova(dependent\_vars ~ independent\_var, data = iris)  
summary(manova\_model)

## Df Pillai approx F num Df den Df Pr(>F)   
## independent\_var 2 1.1919 53.466 8 290 < 2.2e-16 \*\*\*  
## Residuals 147   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

library(effectsize)  
  
eta\_squared(manova\_model)

## # Effect Size for ANOVA (Type I)  
##   
## Parameter | Eta2 (partial) | 95% CI  
## -----------------------------------------------  
## independent\_var | 0.60 | [0.54, 1.00]  
##   
## - One-sided CIs: upper bound fixed at [1.00].

library(MASS)  
  
iris\_lda <- lda(independent\_var ~ dependent\_vars, CV = F)  
iris\_lda

## Call:  
## lda(independent\_var ~ dependent\_vars, CV = F)  
##   
## Prior probabilities of groups:  
## setosa versicolor virginica   
## 0.3333333 0.3333333 0.3333333   
##   
## Group means:  
## dependent\_vars1 dependent\_vars2 dependent\_vars3 dependent\_vars4  
## setosa 5.006 3.428 1.462 0.246  
## versicolor 5.936 2.770 4.260 1.326  
## virginica 6.588 2.974 5.552 2.026  
##   
## Coefficients of linear discriminants:  
## LD1 LD2  
## dependent\_vars1 0.8293776 0.02410215  
## dependent\_vars2 1.5344731 2.16452123  
## dependent\_vars3 -2.2012117 -0.93192121  
## dependent\_vars4 -2.8104603 2.83918785  
##   
## Proportion of trace:  
## LD1 LD2   
## 0.9912 0.0088

LDA\_Dataframe <- data.frame(  
 species = iris[, "Species"],  
 lda = predict(iris\_lda)$x  
)  
head(LDA\_Dataframe)

## species lda.LD1 lda.LD2  
## 1 setosa 8.061800 0.3004206  
## 2 setosa 7.128688 -0.7866604  
## 3 setosa 7.489828 -0.2653845  
## 4 setosa 6.813201 -0.6706311  
## 5 setosa 8.132309 0.5144625  
## 6 setosa 7.701947 1.4617210

ggplot(LDA\_Dataframe) +  
 geom\_point(aes(x = lda.LD1, y = lda.LD2, color = species), size = 4) +  
 theme\_classic()

