Tut7_memo

2023-05-22

Q1. Consider the banking data uploaded in Blackboard. Use the dataset to find how the gender and education (independent variables) influence the multivariate response variables, which are age, income, and loan amount. Use the 'car' package for this MANOVA analysis.

Ans.

Store the dataset into your working directory. Find the working directory by getwd() function is you are not sure about the location of it.

```
# Load required libraries
library(car)

## Loading required package: carData

banking_data = read.csv("banking_data.csv",header = T)

# Perform MANOVA analysis
manova_result <- manova(cbind(Age, Income, LoanAmount) ~ Gender + Education,
data = banking_data)

# Print the results
summary(manova_result)

##           Df   Pillai approx F num Df den Df Pr(>F)
## Gender     1 0.027406  0.89233      3     95 0.4481
## Education  1 0.034598  1.13485      3     95 0.3390
## Residuals 97
```

The p-values associated with the independent variables (gender and education) in the MANOVA analysis are not significant, it indicates that there is no statistically significant evidence to conclude that gender or education has a significant multivariate effect on the combined dependent variables (age, income, and loan amount).

Q2. Let us consider the biotech_data. It contains variables such as SampleID, Gene1, Gene2, Gene3, Treatment, and Outcome. The Gene1, Gene2, and Gene3 variables represent gene expression levels, while Treatment and Outcome are factors representing different treatment groups and outcomes, respectively. Find out if the treatment and outcome is jointly affecting the gene expression levels. Perform a MANOVA analysis accordingly.

```
# Load required libraries
library(car)

biotech_data = read.csv("biotech_data.csv")
```

```
# Perform MANOVA analysis
manova_result <- manova(cbind(Gene1, Gene2, Gene3) ~ Treatment * Outcome,
data = biotech_data)

# Print the results
summary(manova_result)
```

```
##                   Df   Pillai approx F num Df den Df   Pr(>F)
## Treatment          1 0.041851   1.3686      3     94 0.257192
## Outcome            1 0.053827   1.7825      3     94 0.155812
## Treatment:Outcome  1 0.137079   4.9774      3     94 0.003001 **
## Residuals         96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the above example the p-values for treatment and outcome individually are not significant, but the p-value for the joint effect of treatment and outcome is significant, it indicates that while there may not be a significant individual effect of treatment or outcome on the gene expression levels, there is evidence of a significant combined effect when considering both variables together.

Q3. Consider another banking data uploaded in Blackboard. Use the dataset to find how the marital status and education (independent variables) influence the multivariate response variables, which are age, income, and account balance. Use the 'car' package for this MANOVA analysis.

```
# Load required libraries
library(car)

library(readr)
banking_data2 = read_csv("banking_data2.csv")
```

```
## New names:
## Rows: 100 Columns: 7
## ── Column specification
## ──────────────────────────────────────────────────────── Delimiter: ","
chr
## (2): Education, MaritalStatus dbl (5): ...1, CustomerID, Age, Income,
## AccountBalance
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ
## Specify the column types or set `show_col_types = FALSE` to quiet this
message.
## • `` -> `...1`
```

```
# Perform MANOVA analysis
manova_result <- manova(cbind(Age, Income, AccountBalance) ~ Education +
MaritalStatus, data = banking_data2)
```

```r
# Print the results
summary(manova_result)
```

```
##               Df   Pillai approx F num Df den Df  Pr(>F)
## Education      2 0.159087  2.70775      6    188 0.01518 *
## MaritalStatus  2 0.030615  0.48709      6    188 0.81749
## Residuals     95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-values associated with the independent variable Education in the MANOVA analysis is significant, it indicates that there is statistically significant evidence to conclude that education has a significant effect on the combined dependent variables (age, income, and account balance).