UNIVERSIDADE FEDERAL FLUMINENSE

Programa de Mestrado e Doutorado em Engenharia de Produção

Multivariate Data Analysis

Correspondence Analysis

Professor: Valdecy Pereira, D. Sc.

email: valdecy.pereira@gmail.com

Outline

1. Definition
2. Terminology
3. CA
4. MCA
5. Bibliography

MVDA – *Correspondence Analysis*

The **CA** (Correspondence Analysis) is primarily a descriptive and visual technique designed to analyze contingency tables containing some measure of association between the rows and columns (variables) and for some situations this technique can take an exploratory character. Categorical data are the most common entries to start treatment and analysis, but continuous data can be used if they can be categorized (Ex: Age group - Age is a continuous variable, but can be grouped into categories) The **CA** is considered a special case of PCA and it can explore only two variables at a time, so this technique is also known as Simple Correspondence Analysis.

The **MCA** (Multiple Correspondence Analysis) is also a primarily descriptive and visual technique, which is used for the study of the relationship between two or more nominal or ordinal variables. A two-dimensional graphic (Biplot) typically allows a more comprehensive analysis of the data.

MVDA – *Correspondence Analysis*

- The **CA** and **MCA** can graphically represent:

- The preference of certain consumers for different brands;
- Psychological profiles and social behavior;
- Animal Species and habitats;
- Good management practices and organizational performance;
- Etc.

MVDA – *Correspondence Analysis*

- **Dimensions**: The maximum number of dimensions is given by the minimum between the number of lines and the number of columns, minus 1.
- **Mass**: Represent the marginal proportions of the row and column variables.
- **Scores**: Values that are used as coordinate points of the categories to plot a correspondence map.
- **Inertia**: It is an association measure (represented by the amount of variance) between two categorical variables. The total inertia (sum of inertia for each dimension) can be calculated as the value of chi-square statistic divided by the number of cases.
- **Eigenvalue**: The squared eigenvalue of a dimension represents its inertia.
- **Contribution of categories to dimensions**: The highest scores can be used to interpret a dimension, it is analogous to factor loadings.
- **Contribution of dimensions to categories**: They reflect how well a dimension may explain a particular category. Analogous to the coefficient of determination (proportion of explained variance).

Simple Correspondence Analysis

MVDA – *Correspondence Analysis*

In order to explain a **CA** approach, the following dataset will be used - A sample of 100 students was collected with the following information :

- Profile of Investment (categorical variable: Aggressive, Conservative or Moderate);
- Application Type (categorical variable: BDC (Bank Deposit Certificate), Savings, or Stock Shares).

Id	Student	Profile	Application
1	Gabriela	Conservative	Savings
2	Luiz_Felipe	Conservative	Savings
3	Patrícia	Conservative	Savings
4	Gustavo	Conservative	Savings
5	Leticia	Conservative	Savings
6	Ovidio	Conservative	Savings
7	Leonor	Conservative	Savings
8	Dalila	Conservative	Savings
9	Antonio	Conservative	BDC
10	Julia	Conservative	BDC
…			
34	Cintia	Moderate	BDC
…			
99	Leandro	Aggressive	Stock_shares
100	Estela	Aggressive	Stock_shares

MVDA – *Correspondence Analysis*

```
# Transforming data into a contingency table (if necessary!)
my_data <- table(my_data)
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	20	2	36
Conservative	4	8	5
Moderate	16	5	4

MVDA – *Correspondence Analysis*

```
library("FactoMineR")
ca <- CA(my_data, ncp = 2, graph = FALSE)

# Variance Percentage (Inertia)
ca$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.23321487	73.42075	73.42075
dim 2	0.08442678	26.57925	100.00000

```
# Trace (Correlation Between Rows and Columns)

trace <- sqrt(sum(ca$eig[,1]))
[1] 0.5635971

# Row Mass (Row Weigths)
ca$call$marge.row
```

Aggressive	Conservative	Moderate
0.58	0.17	0.25

```
# Column Mass (Column Weigths)
ca$call$marge.col
```

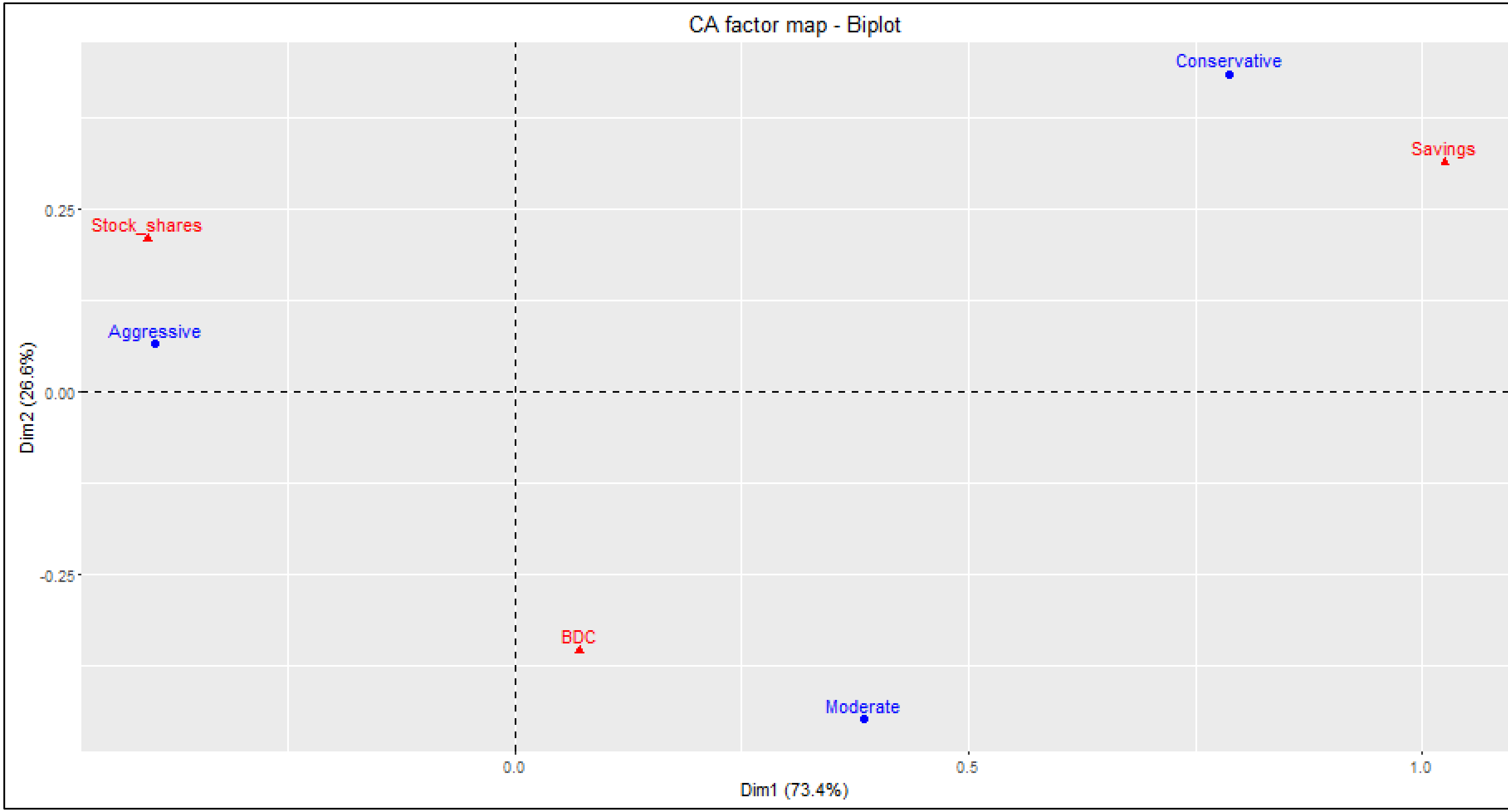
BDC	Savings	Stock_shares
0.40	0.15	0.45

MVDA – *Correspondence Analysis*

```
# Contribution of the categories in column to dimensions (Analogous to factor Loadings)
ca$col$contrib
                  Dim 1        Dim 2
BDC           0.8650818     59.13492
Savings      67.5336283     17.46637
Stock_shares 31.6012900     23.39871

# Contribution of the categories in row to dimensions (Analogous to factor Loadings)
ca$row$contrib
                 Dim 1        Dim 2
Aggressive    39.05149     2.948508
Conservative  45.10799    37.892011
Moderate      15.84052    59.159481
```

MVDA – *Correspondence Analysis*

```
# Coordinates of categories in row (Scores)
ca$row$coord

                  Dim 1         Dim 2
Aggressive   -0.3962625    0.06551296
Conservative  0.7866479    0.43379993
Moderate      0.3844084   -0.44697402

# Coordinates of categories in column (Scores)
ca$col$coord

                  Dim 1         Dim 2
BDC          0.07101935    -0.3532906
Savings      1.02469008     0.3135421
Stock_shares -0.40469167    0.2095221

# Biplot
library("factoextra")
fviz_ca_biplot(ca)
```


CA factor map - Biplot
Conservative
Savings
Stock_shares
Aggressive
BDC
Moderate
0.25
0.00
-0.25
0.0
0.5
1.0
Dim1 (73.4%)
Dim2 (26.6%)

Goodness of Fit

MVDA – *Correspondence Analysis*

Row and Column Sums (Weights)

	Application				
Profile	BDC	Savings	Stock Shares		Sum
Aggressive	20	2	36		58
Conservative	4	8	5		17
Moderate	16	5	4		25
Sum	40	15	45		**n = 100**

MVDA – *Correspondence Analysis*

```
# Expected Absolut Frequencies
P <- my_data
for (i in 1:3){
     for (j in 1:3){
         P[i, j] <- (sum(my_data[ ,j])*sum(my_data[i, ])/sum(my_data))
                 }
}
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	23.2	8.7	26.1
Conservative	6.8	2.55	7.65
Moderate	10	3.75	11.25

$$Absolut\ Fequencies = P = \frac{\sum l_i \times \sum C_j}{n}$$

MVDA – *Correspondence Analysis*

```
# Residuals
E <- my_data - P
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	-3.2	-6.7	9.9
Conservative	-2.8	5.45	-2.65
Moderate	6	1.25	-7.25

$$Residuals = E = Original\ Data\ - P$$

MVDA – *Correspondence Analysis*

```
# Chi-Squared Table
Chi_Sq_Table <- (E)^2/P
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	0.44	5.16	3.75
Conservative	1.15	11.65	0.92
Moderate	3.60	0.42	4.67

$$Chi\ Squared\ Table = \chi^2 = \frac{E^2}{P}$$

MVDA – *Correspondence Analysis*

```
# Chi-Squared Table
Chi_Sq_Table <- (E)^2/P
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	0.44	5.16	3.75
Conservative	1.15	11.65	0.92
Moderate	3.60	0.42	4.67

The following hypothesis is made to test the model adequacy:

$$H_0: The\ association\ between\ both\ variables\ is\ random$$
$$H_1: The\ association\ between\ both\ variables\ is\ not\ random$$

```
# Chi-Squared Value
Chi_Sq <- sum(Chi_Sq_Table)

[1]  31.76416
```

Reject the null hypothesis H_0 if ${\chi^2}_{test} > {\chi^2}_{(rows-1)(columns-1)}$

```
# Chi-Squared Test
qchisq(0.95, df = 2*2)

[1] 9.487729
```

MVDA – *Correspondence Analysis*

```
# Beta Test  (Association between variable is significant for values greater than 3)
Beta <- (Chi_Sq - (3 - 1)*(3 - 1))/((3 - 1)*(3- 1))^(1/2)

[1] 13.88208
```

$$Beta = \frac{{\chi^2}_{test} - (l-1) \times (c-1)}{\sqrt{(l-1) \times (c-1)}}$$

MVDA – *Correspondence Analysis*

```
# Standartized Residuals
E_st <- E/(P)^(0.5)
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	-0.66	-2.27	1.94
Conservative	-1.07	3.41	-0.96
Moderate	1.90	0.64	-2.16

$$Standartized\ Residuals = E_{st} = \frac{E}{\sqrt{P}}$$

MVDA – *Correspondence Analysis*

```
# Adjusted Standartized Residuals
E_st_adj <- E_st
for (i in 1:3){
     for (j in 1:3){
         E_st_adj[i, j] <- E_st[i, j]/((1 - sum(my_data[ ,j])/sum(my_data))*(1 - sum(my_data[i, ])/sum(my_data)))^(0.5)
                 }
}
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	-1.32	-3.80	4.03
Conservative	-1.52	4.06	-1.42
Moderate	2.83	0.81	-3.36

$$Adjusted\ Standartized\ Residuals = E_{adj} = \frac{E_{st}}{\sqrt{\left(\frac{1-\sum l_i}{n}\right) \times \left(\frac{1-\sum C_j}{n}\right)}}$$

MVDA – *Correspondence Analysis*

```
# Adjusted Standartized Residuals
E_st_adj <- E_st
for (i in 1:3){
     for (j in 1:3){
         E_st_adj[i, j] <- E_st[i, j]/((1 - sum(my_data[ ,j])/sum(my_data))*(1 - sum(my_data[i, ])/sum(my_data)))^(0.5)
                 }
}
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	-1.32	-3.80	4.03
Conservative	-1.52	4.06	-1.42
Moderate	2.83	0.81	-3.36

Values greater than 1.96 (5% of significance level), shows association between the category in row with the category in column.

MVDA – *Correspondence Analysis*

Quality of representation (0 = worst, 1 = Perfect) of the categories in column for each dimension (Sum of Both Dimensions)
cacolcos2

	Dim 1	Dim 2
BDC	0.03884049	0.96115951
Savings	0.91438756	0.08561244
Stock_shares	0.78861426	0.21138574

Quality of representation (0 = worst, 1 = Perfect) of the categories in row for each dimension (Sum of Both Dimensions)
carowcos2

	Dim 1	Dim 2
Aggressive	0.9733941	0.02660586
Conservative	0.7668116	0.23318836
Moderate	0.4251688	0.57483119

Multiple Correspondence Analysis

MVDA – *Correspondence Analysis*

In order to explain a **MCA** approach, the following dataset will be used - A sample of 100 students was collected with the following information:

- Profile of Investment (categorical variable: Aggressive, Conservative or Moderate);
- Application Type (categorical variable: BDC (Bank Deposit Certificate), Savings, or Stock Shares).
- Marital Status (categorical variable: Married or Single).

Id	Student	Profile	Application	Marital Status
1	Gabriela	Conservative	Savings	Married
2	Luiz_Felipe	Conservative	Savings	Married
3	Patrícia	Conservative	Savings	Married
4	Gustavo	Conservative	Savings	Single
5	Leticia	Conservative	Savings	Married
6	Ovidio	Conservative	Savings	Married
7	Leonor	Conservative	Savings	Married
8	Dalila	Conservative	Savings	Married
9	Antonio	Conservative	BDC	Married
10	Julia	Conservative	BDC	Married
…				
34	Cintia	Moderate	BDC	Married
…				
99	Leandro	Aggressive	Stock_shares	Single
100	Estela	Aggressive	Stock_shares	Single

MVDA – *Correspondence Analysis*

```
# Transforming data into a contingency table
cont_my_data2 <- table(my_data2)
```

Married	**Application**		
Profile	BDC	Savings	Stock Shares
Aggressive	11	0	6
Conservative	3	7	2
Moderate	10	3	1

Single	**Application**		
Profile	BDC	Savings	Stock Shares
Aggressive	9	2	30
Conservative	1	1	3
Moderate	6	2	3

MVDA – *Correspondence Analysis*

```
library("FactoMineR")
mca <- MCA(my_data2, ncp = 2, graph = FALSE)

# Variance Percentage (Inertia)
mca$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.6023045	36.13827	36.13827
dim 2	0.4359878	26.15927	62.29754
dim 3	0.2764728	16.58837	78.88591
dim 4	0.1798371	10.79022	89.67613
dim 5	0.1720645	10.32387	100.00000

```
# Column Mass (Column Weigths)
mca$call$marge.col
```

Aggressive	Conservative	Moderate	BDC	Savings	Stock_shares
0.19333333	0.05666667	0.08333333	0.13333333	0.05000000	0.15000000

Married	Single
0.14333333	0.19000000

MVDA – *Correspondence Analysis*

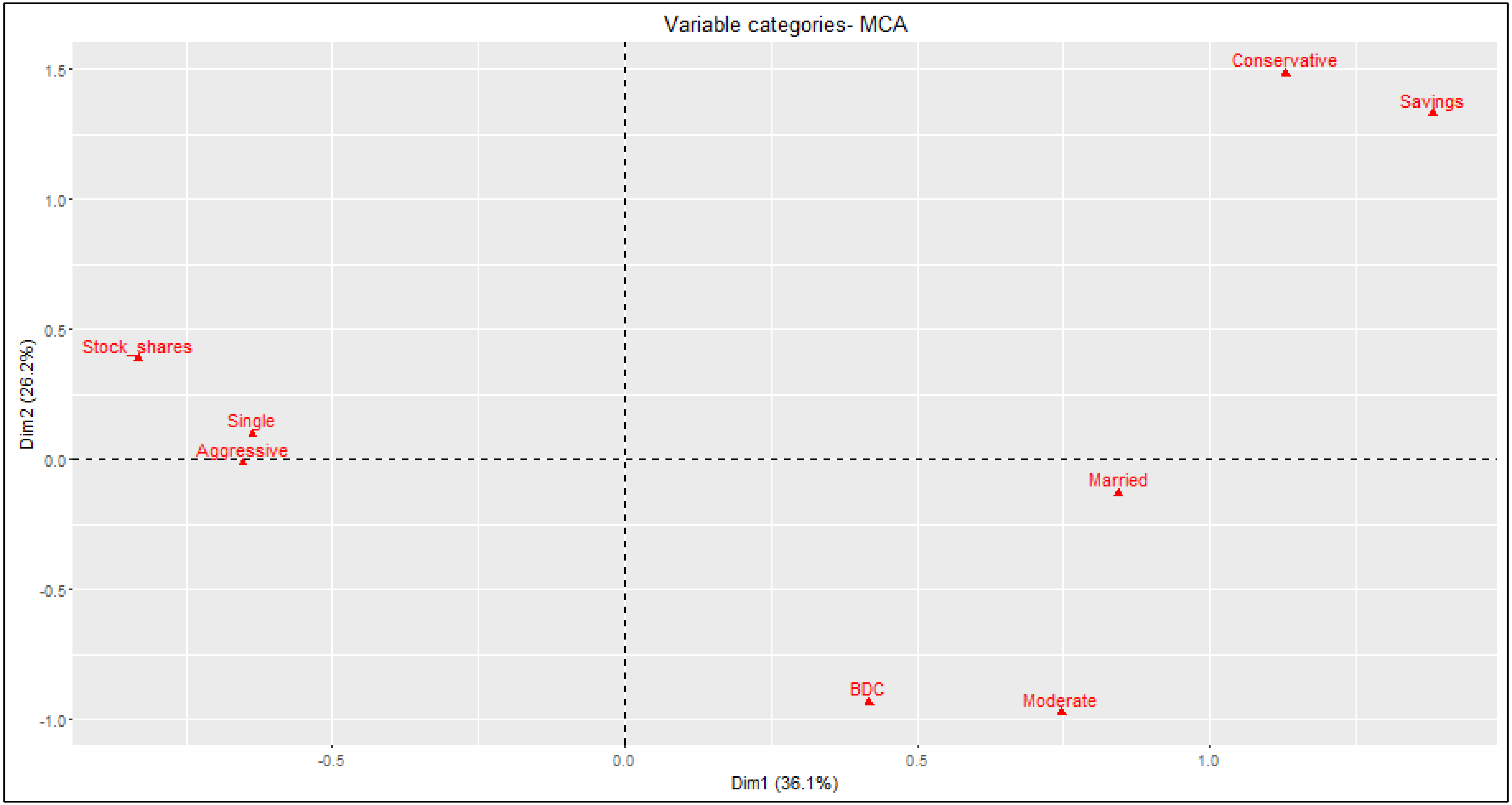
```
# Row Mass (Row Weigths)
mca$call$marge.row

# Contribution of the categories in column to dimensions (Analogous to factor Loadings)
mca$var$contrib
```

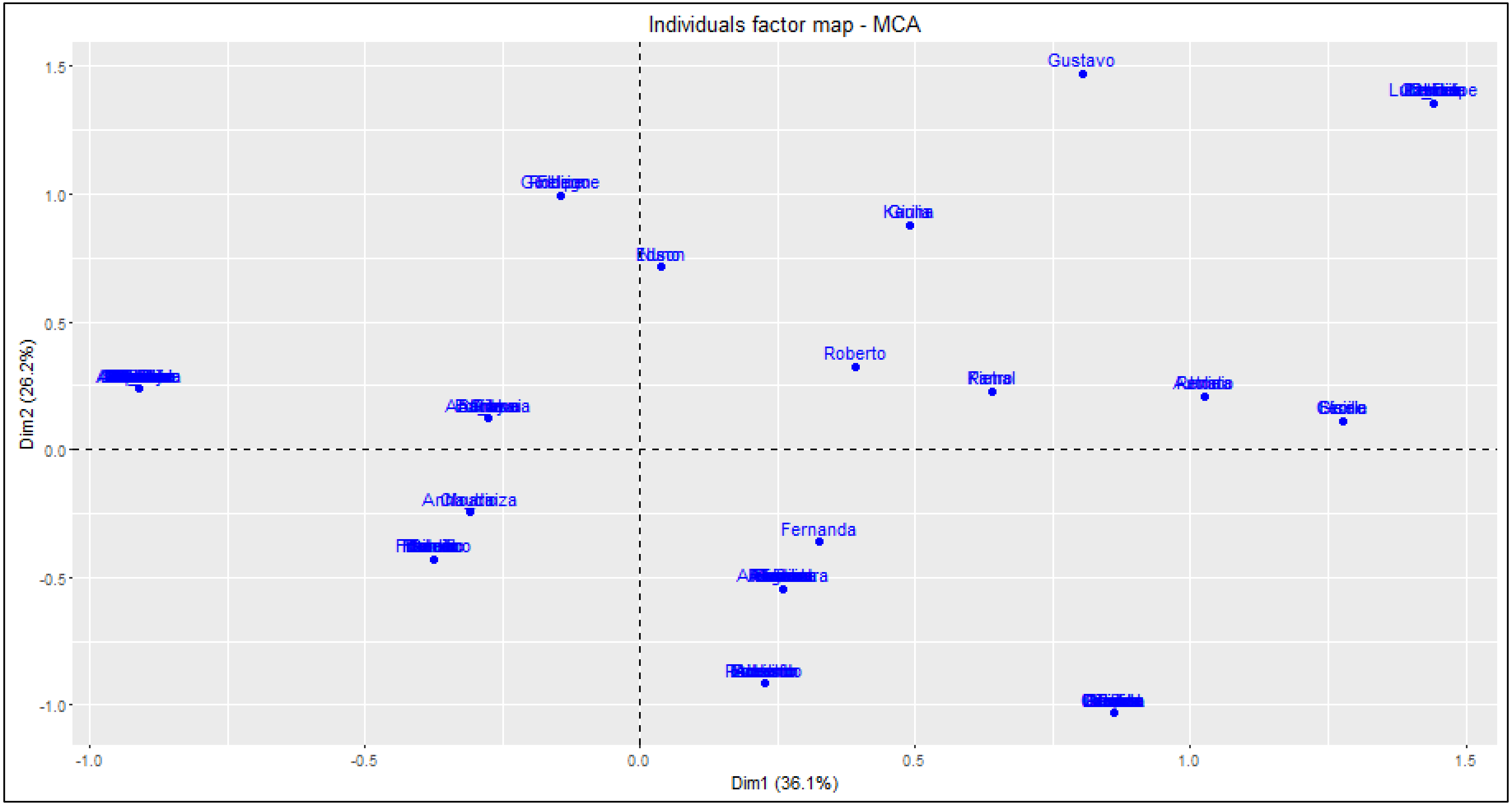
	Dim 1	Dim 2
Aggressive	13.690180	0.009624243
Conservative	12.012834	28.608515192
Moderate	7.715121	18.157985111
BDC	3.852805	26.743624674
Savings	15.838730	20.321796679
Stock_shares	17.208656	5.166471138
Married	16.918554	0.565430289
Single	12.763120	0.426552674

```
# Contribution of the categories in row to dimensions (Analogous to factor Loadings)
mca$ind$contrib
```

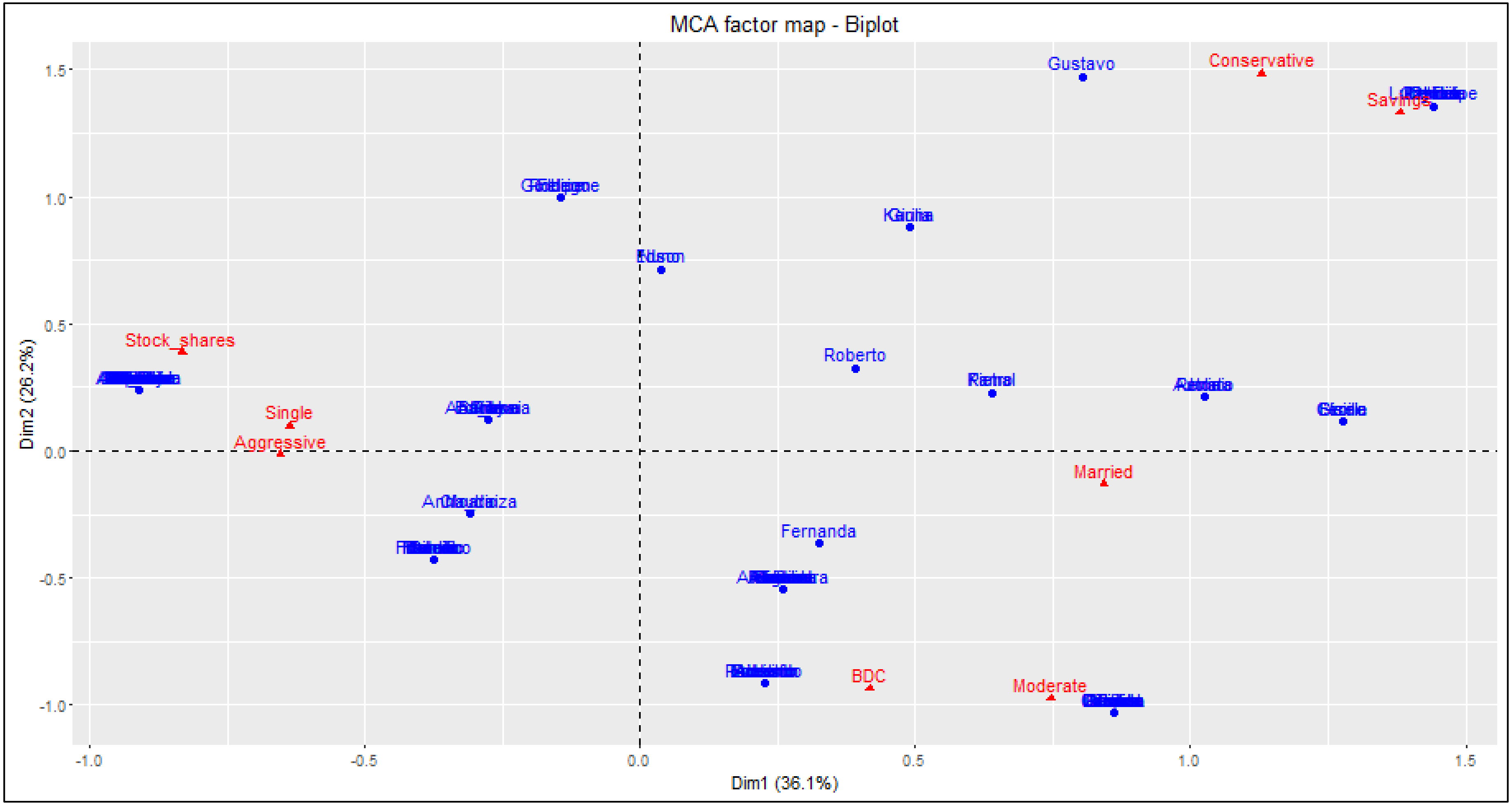
MVDA – *Correspondence Analysis*

```
# Coordinates of categories in row (Scores)
ca$row$coord

                  Dim 1          Dim 2
Aggressive   -0.3962625     0.06551296
Conservative  0.7866479     0.43379993
Moderate      0.3844084    -0.44697402

# Coordinates of categories in column (Scores)
ca$col$coord

                  Dim 1          Dim 2
BDC           0.07101935    -0.3532906
Savings       1.02469008     0.3135421
Stock_shares -0.40469167     0.2095221

# Biplot
library("factoextra")
fviz_mca_var(mca)
fviz_mca_ind(mca)
fviz_mca_biplot(mca)
```


Variable categories- MCA
Conservative
Savings
Stock_shares
Single
Aggressive
Married
BDC
Moderate
Dim1 (36.1%)
Dim2 (26.2%)
-0.5
0.0
0.5
1.0
1.5
-1.0



Individuals factor map - MCA
Gustavo
Roberto
Fernanda
Dim1 (36.1%)
Dim2 (26.2%)
-1.0
-0.5
0.0
0.5
1.0
1.5


MCA factor map - Biplot
Gustavo
Conservative
Savings
Roberto
Fernanda
Stock_shares
Single
Aggressive
Married
BDC
Moderate
Dim1 (36.1%)
Dim2 (26.2%)

Goodness of Fit

MVDA – *Correspondence Analysis*

```
P <- table(my_data2[,c(1,2)])
temp <- table(my_data2[,c(1,2)])
for (i in 1:3){
     for (j in 1:3){
         P[i, j] <- (sum(temp[ ,j])*sum(temp[i, ])/sum(temp))
                 }
}
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	23.2	8.7	26.1
Conservative	6.8	2.55	7.65
Moderate	10	3.75	11.25

MVDA – *Correspondence Analysis*

```
E <- temp - P
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	-3.2	-6.7	9.9
Conservative	-2.8	5.45	-2.65
Moderate	6	1.25	-7.25

MVDA – *Correspondence Analysis*

```
Chi_Sq_Table <- (E)^2/P
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	0.44	5.16	3.75
Conservative	1.15	11.65	0.92
Moderate	3.60	0.42	4.67

MVDA – *Correspondence Analysis*

```
Chi_Sq_Table <- (E)^2/P
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	0.44	5.16	3.75
Conservative	1.15	11.65	0.92
Moderate	3.60	0.42	4.67

The following hypothesis is made to test the model adequacy:

$$H_0: The\ association\ between\ both\ variables\ is\ random$$
$$H_1: The\ association\ between\ both\ variables\ is\ not\ random$$

```
Chi_Sq <- sum(Chi_Sq_Table)

[1]  31.76416
```

Reject the null hypothesis H_0 if $\chi^2{}_{test} > \chi^2{}_{(rows-1)(columns-1)}$

```
qchisq(0.95, df = 2*2)

[1] 9.487729
```

MVDA – *Correspondence Analysis*

```
Beta <- (Chi_Sq - (3 - 1)*(3 - 1))/((3 - 1)*(3- 1))^(1/2)

[1] 13.88208
```

$$Beta = \frac{{\chi^2}_{test} - (l-1)\times(c-1)}{\sqrt{(l-1)\times(c-1)}}$$

MVDA – *Correspondence Analysis*

```
E_st <- E/(P)^(0.5)
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	-0.66	-2.27	1.94
Conservative	-1.07	3.41	-0.96
Moderate	1.90	0.64	-2.16

MVDA – *Correspondence Analysis*

```
E_st_adj <- E_st
for (i in 1:3){
    for (j in 1:3){
        E_st_adj[i, j] <- E_st[i, j]/((1 - sum(temp[ ,j])/sum(temp))*(1 - sum(temp[i, ])/sum(temp)))^(0.5)
            }
}
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	-1.32	-3.80	4.03
Conservative	-1.52	4.06	-1.42
Moderate	2.83	0.81	-3.36

MVDA – *Correspondence Analysis*

```
E_st_adj <- E_st
for (i in 1:3){
     for (j in 1:3){
          E_st_adj[i, j] <- E_st[i, j]/((1 - sum(temp[ ,j])/sum(temp))*(1 - sum(temp[i, ])/sum(temp)))^(0.5)
                 }
}
```

	Application		
Profile	BDC	Savings	Stock Shares
Aggressive	-1.32	-3.80	4.03
Conservative	-1.52	4.06	-1.42
Moderate	2.83	0.81	-3.36

MVDA – *Correspondence Analysis*

```
P <- table(my_data2[,c(1,3)])
temp <- table(my_data2[,c(1,3)])
for (i in 1:3){
     for (j in 1:2){
          P[i, j] <- (sum(temp[ ,j])*sum(temp[i, ])/sum(temp))
                    }
}
```

	Marital Status	
Profile	Single	Married
Aggressive	24.94	33.06
Conservative	7.31	9.69
Moderate	10.75	14.25

MVDA – *Correspondence Analysis*

```
E <- temp - P
```

	Marital Status	
Profile	Single	Married
Aggressive	-7.94	7.94
Conservative	4.69	-4.69
Moderate	3.25	-3.25

MVDA – *Correspondence Analysis*

```
Chi_Sq_Table <- (E)^2/P
```

	Marital Status	
Profile	Single	Married
Aggressive	2.53	1.91
Conservative	3.01	2.27
Moderate	0.98	0.74

MVDA – *Correspondence Analysis*

```
Chi_Sq_Table <- (E)^2/P
```

	Marital Status	
Profile	Single	Married
Aggressive	2.53	1.91
Conservative	3.01	2.27
Moderate	0.98	0.74

The following hypothesis is made to test the model adequacy:

$$H_0: The\ association\ between\ both\ variables\ is\ random$$
$$H_1: The\ association\ between\ both\ variables\ is\ not\ random$$

```
Chi_Sq <- sum(Chi_Sq_Table)

[1] 11.43756
```

Reject the null hypothesis H_0 if ${\chi^2}_{test} > {\chi^2}_{(rows-1)(columns-1)}$

```
qchisq(0.95, df = 2*1)

[1] 5.991465
```

MVDA – *Correspondence Analysis*

```
Beta <- (Chi_Sq - (3 - 1)*(2 - 1))/((3 - 1)*(2- 1))^(1/2)

[1] 6.673365
```

$$Beta = \frac{{\chi^2}_{test} - (l-1)\times(c-1)}{\sqrt{(l-1)\times(c-1)}}$$

MVDA – *Correspondence Analysis*

```
E_st <- E/(P)^(0.5)
```

	Marital Status	
Profile	Single	Married
Aggressive	-1.59	1.38
Conservative	1.73	-1.51
Moderate	0.99	-0.86

MVDA – *Correspondence Analysis*

```
E_st_adj <- E_st
for (i in 1:3){
     for (j in 1:2){
          E_st_adj[i, j] <- E_st[i, j]/((1 - sum(temp[ ,j])/sum(temp))*(1 - sum(temp[i, ])/sum(temp)))^(0.5)
                 }
}
```

	Marital Status	
Profile	Single	Married
Aggressive	-3.25	3.25
Conservative	2.52	-2.52
Moderate	1.52	-1.52

MVDA – *Correspondence Analysis*

```
E_st_adj <- E_st
for (i in 1:3){
     for (j in 1:2){
          E_st_adj[i, j] <- E_st[i, j]/((1 - sum(temp[ ,j])/sum(temp))*(1 - sum(temp[i, ])/sum(temp)))^(0.5)
                  }
}
```

	Marital Status	
Profile	Single	Married
Aggressive	-3.25	3.25
Conservative	2.52	-2.52
Moderate	1.52	-1.52

MVDA – *Correspondence Analysis*

```
P <- table(my_data2[,c(2,3)])
temp <- table(my_data2[,c(2,3)])
for (i in 1:3){
     for (j in 1:2){
         P[i, j] <- (sum(temp[ ,j])*sum(temp[i, ])/sum(temp))
                 }
}
```

	Marital Status	
Aplication	Single	Married
BDC	17.20	22.80
Savings	6.45	8.55
Stock Shares	19.35	25.65

MVDA – *Correspondence Analysis*

```
E <- temp - P
```

	Marital Status	
Aplication	Single	Married
BDC	6.80	-6.80
Savings	3.55	-3.55
Stock Shares	-10.35	10.35

MVDA – *Correspondence Analysis*

```
Chi_Sq_Table <- (E)^2/P
```

	Marital Status	
Aplication	Single	Married
BDC	2.69	2.03
Savings	1.95	1.47
Stock Shares	5.53	4.17

MVDA – *Correspondence Analysis*

```
Chi_Sq_Table <- (E)^2/P
```

	Marital Status	
Aplication	Single	Married
BDC	2.69	2.03
Savings	1.95	1.47
Stock Shares	5.53	4.17

The following hypothesis is made to test the model adequacy:

$$H_0: The\ association\ between\ both\ variables\ is\ random$$
$$H_1: The\ association\ between\ both\ variables\ is\ not\ random$$

```
Chi_Sq <- sum(Chi_Sq_Table)

[1] 17.85666
```

Reject the null hypothesis H_0 if $\chi^2{}_{test} > \chi^2{}_{(rows-1)(columns-1)}$

```
qchisq(0.95, df = 2*1)

[1] 5.991465
```

MVDA – *Correspondence Analysis*

```
Beta <- (Chi_Sq - (3 - 1)*(2 - 1))/((3 - 1)*(2- 1))^(1/2)

[1] 11.21235
```

$$Beta = \frac{{\chi^2}_{test} - (l-1) \times (c-1)}{\sqrt{(l-1) \times (c-1)}}$$

MVDA – *Correspondence Analysis*

```
E_st <- E/(P)^(0.5)
```

	Marital Status	
Aplication	Single	Married
BDC	1.64	-1.42
Savings	1.39	-1.21
Stock Shares	-2.35	2.04

MVDA – *Correspondence Analysis*

```
E_st_adj <- E_st
for (i in 1:3){
     for (j in 1:2){
          E_st_adj[i, j] <- E_st[i, j]/((1 - sum(temp[ ,j])/sum(temp))*(1 - sum(temp[i, ])/sum(temp)))^(0.5)
                }
}
```

	Marital Status	
Aplication	Single	Married
BDC	2.80	-2.80
Savings	2.01	-2.01
Stock Shares	-4.20	4.20

MVDA – *Correspondence Analysis*

```
E_st_adj <- E_st
for (i in 1:3){
     for (j in 1:2){
          E_st_adj[i, j] <- E_st[i, j]/((1 - sum(temp[ ,j])/sum(temp))*(1 - sum(temp[i, ])/sum(temp)))^(0.5)
                 }
}
```

	Marital Status	
Aplication	Single	Married
BDC	2.80	-2.80
Savings	2.01	-2.01
Stock Shares	-4.20	4.20

MVDA – *Correspondence Analysis*

Quality of representation (0 = worst, 1 = Perfect) of the categories in column for each dimension (Sum of Both Dimensions)
mcavarcos

	Dim 1	Dim 2
Aggressive	0.5889755	0.000299718
Conservative	0.2615199	0.450830033
Moderate	0.1858741	0.316666416
BDC	0.1160281	0.582994735
Savings	0.3366967	0.312707855
Stock_shares	0.5653555	0.122864646
Married	0.5363222	0.012974775
Single	0.5363222	0.012974775

Quality of representation (0 = worst, 1 = Perfect) of the categories in row for each dimension (Sum of Both Dimensions)
mcaindcos

Squared correlation between each variable and the dimensions. (Coefficient of Determination)
mcavareta2

	Dim 1	Dim 2
Profile	0.6038368	0.61181462
Aplication	0.6667546	0.68317407
Marital_Status	0.5363222	0.01297477

MVDA – *Correspondence Analysis*

```
#  If the absolute value of the v.test is superior to 2, then the category coordinate is significantly different from 0.
mca$var$v.test
```

	Dim 1	Dim 2
Aggressive	-7.636005	-0.1722559
Conservative	5.088268	6.6807315
Moderate	4.289701	-5.5991049
BDC	3.389216	-7.5971362
Savings	5.773471	5.5639983
Stock_shares	-7.481323	3.4876353
Married	7.286693	-1.1333590
Single	-7.286693	1.1333590

MVDA – *Correspondence Analysis*

If the absolute value of the v.test is superior to 2, then the category coordinate is significantly different from 0.
mcavarv.test

	Dim 1	Dim 2
Aggressive	-7.636005	**-0.1722559**
Conservative	5.088268	6.6807315
Moderate	4.289701	-5.5991049
BDC	3.389216	-7.5971362
Savings	5.773471	5.5639983
Stock_shares	-7.481323	3.4876353
Married	7.286693	**-1.1333590**
Single	-7.286693	**1.1333590**

MVDA – *Correspondence Analysis*

```
# Dimension Description (This function can be used to identify the most correlated variables with a given dimension)
dimdesc(mca, axes = 1, proba = 0.05)

$`Dim 1`$quali
                      R2       p.value
Aplication     0.6667546  7.146026e-24
Profile        0.6038368  3.139751e-20
Marital_Status 0.5363222  4.802063e-18

$`Dim 1`$category
                 Estimate      p.value
Married        0.57400977 4.802063e-18
Savings        0.82177962 2.496427e-10
Conservative   0.56040122 5.414449e-08
Moderate       0.26298378 7.536668e-06
BDC            0.07355727 5.247853e-04
Single        -0.57400977 4.802063e-18
Stock_shares  -0.89533689 1.969490e-19
Aggressive    -0.82338500 1.249415e-20
```

MVDA – *Correspondence Analysis*

```
dimdesc(mca, axes = 2, proba = 0.05)

$`Dim 2`$quali
                    R2          p.value
Aplication     0.6831741     6.163177e-25
Profile        0.6118146     1.170577e-20

$`Dim 2`$category
                   Estimate       p.value
Conservative    0.87084816   2.083068e-14
Savings         0.70650682   1.474001e-09
Stock_shares    0.08341794   3.498686e-04
Moderate       -0.75234753   1.104083e-09
BDC            -0.78992476   2.548404e-20
```

MVDA

https://github.com/Valdecy/Multivariate_Data_Analysis

##
##

Created by: Prof. Valdecy Pereira, D.Sc.
UFF - Universidade Federal Fluminense (Brazil)
email: valdecypereira@yahoo.com.br
Course: Multivariate Data Analysis
Lesson: Correspondence Analysis

Citation:
PEREIRA, V. (2016). Project: Multivariate Data Analysis, File: R-MVDA-05-CA.pdf, GitHub repository: <https://github.com/Valdecy/Multivariate_Data_Analysis>

##
##

Bibliography

CORRAR, L.J.; PAULO, E.; DIAS FILHO, J. M. **Análise Multivariada para Cursos de Administração, Ciências Contábeis e Economia**. ATLAS, 2009.

FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. **Análise de Dados: Modelagem Multivariada para Tomada de Decisões**. CAMPUS, 2009.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise Multivariada de Dados**. BOOKMAN, 2009.

LATTIN, J.; CARROLL, J. D.; GREEN, P. E. **Análise de Dados Multivariados**. CENGAGE Learning, 2011.

LEVINE, D. M.; STEPHAN, D. F.; KREHBIEL, T. C.; BERENSON, M. L. **Estatística - Teoria e Aplicações - Usando Microsoft Excel**. LTC, 2012.