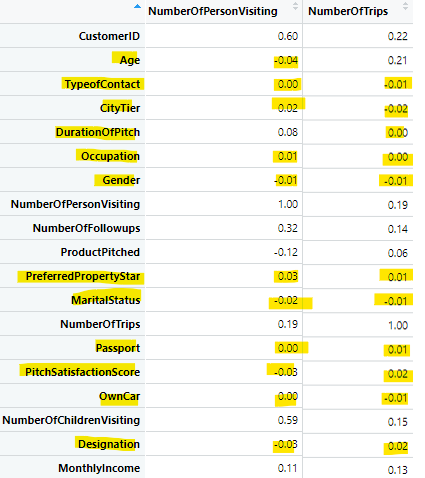
**Choosing Appropriate Statistical Analyses**

**DVs: Number of Trips and Visitors**

**For Prod Not Taken**

**Canonical Correlation Analysis **

|  |  |  |
| --- | --- | --- |
| **Continuous** | **Visitors** | **Trips** |
| Age | -0.04 |  |
| PropStar | 0.03 | 0.01 |
| Score | -0.03 | 0.02 |
| Duration |  | 0 |

**MANOVA**

|  |  |  |
| --- | --- | --- |
| **Categorical** | **Visitors** | **Trips** |
| TypeofContact | 0 | -0.01 |
| Occupation | 0.01 | 0 |
| Gender | -0.01 | -0.01 |
| MaritalStatus | -0.02 | -0.01 |
| Passport | 0 | 0.01 |
| OwnCar | 0 | -0.01 |
| Designation | -0.03 | -0.02 |

**MANOVA Analysis:**#IV1 = TypeofContact

#IV2 = Occupation

#IV3 = Gender

#IV4 = MaritalStatus

#IV5 = Passport

#DV1 = NumberOfPersonVisiting

#DV2 = NumberOfTrips

**Canonical Correlation Analysis**

#IV1 = Age

#IV2 = PreferredPropertyStar

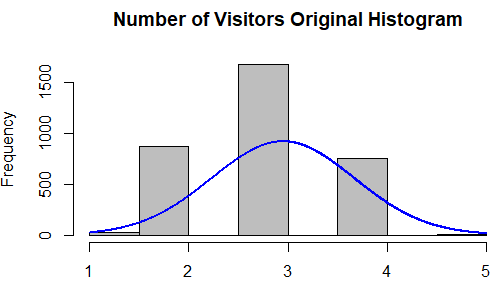
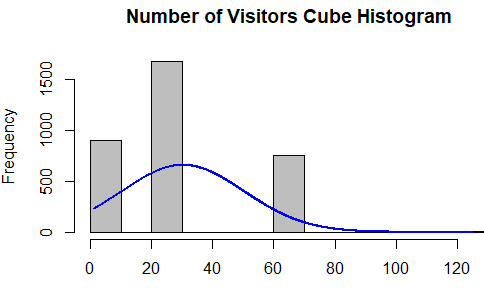
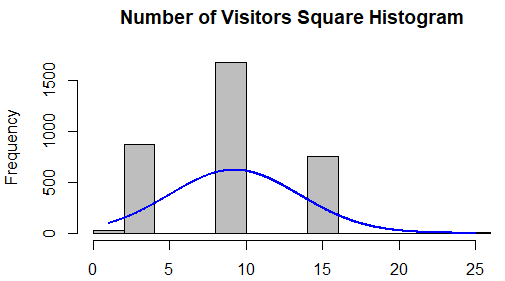
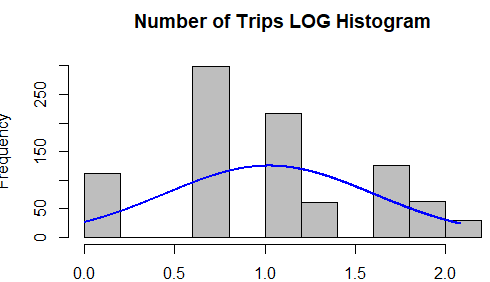
#IV3 = PitchSatisfactionScore

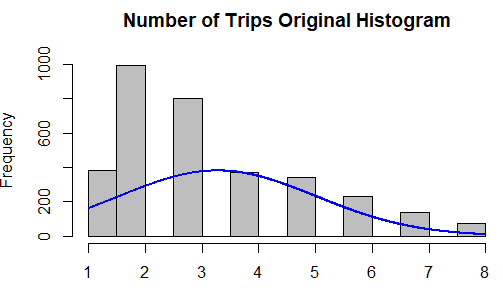
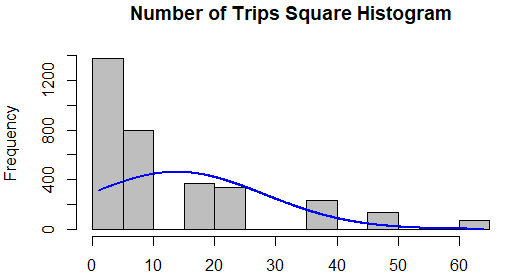
#IV4 = DurationOfPitch

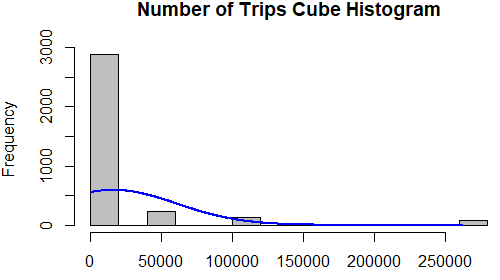
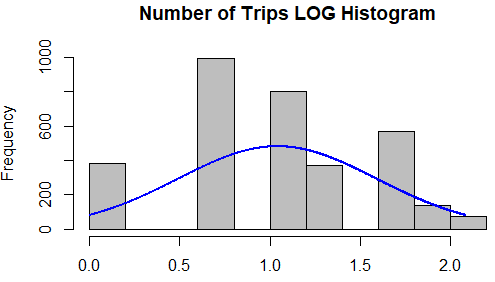
#DV1 = NumberOfPersonVisiting

#DV2 = NumberOfTrips

Check normality of DVs

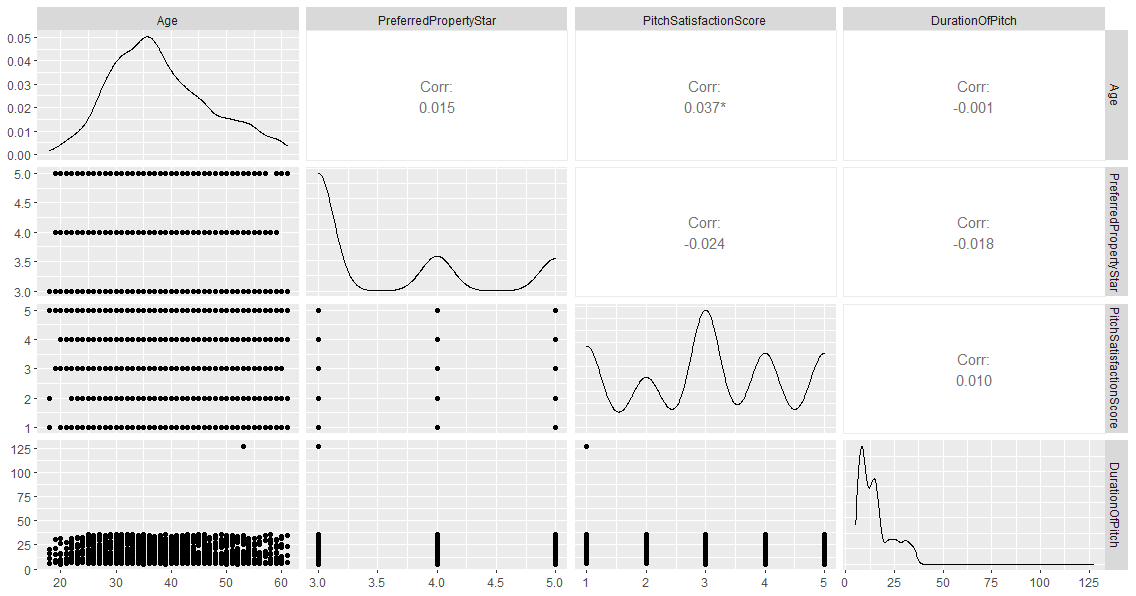
 

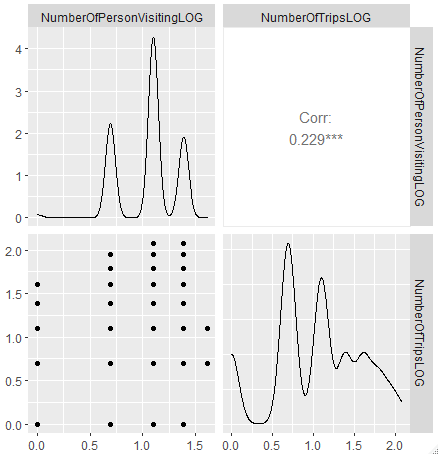
 

Exam the correlation within each subset:

By Graph:

ggpairs(IVs)



ggpairs(DVs)

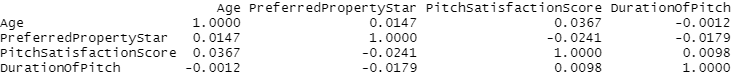
The associations between the two sets can be extracted as

#Exam the correlation between the two sets of variables using matcor from CCA

# correlations: cormat<-matcor(DVs, IVs)

#Extracting the within study correlations for IVs and DVs and between set cor

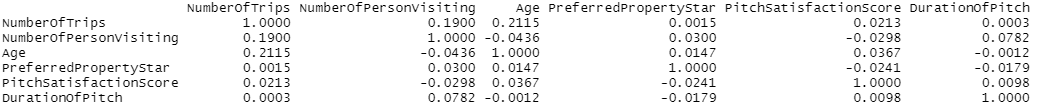
round(cormat$Ycor, 4)



#The associations between the two sets can be extracted as

#Between set associations: cormat<-matcor(DVs,IVs)

round(cormat$XYcor, 4)



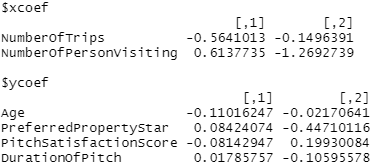
Obtain the canonical correlations

can\_cor1=cc(DVs, IVs)

can\_cor1$cor

[1] 0.2320303 0.0819885

Obtain raw canonical coefficients: can\_cor1[3:4]

Increase = +

Decrease = -

**No.Trip No. Visitor**

**+/increase 1 unit of =**

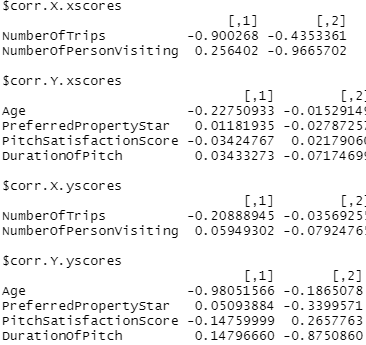
Usually, the number of canonical dimensions the same as the count of variables in the smaller set. The number of canonical dimensions that are significant in explaining the relationship between the 2 sets of variables may, however, be smaller than the number of variables in the smaller data set. In this case, there are 2 dimensions.

Computes the canonical loadings

can\_cor2=comput(DVs,IVs,can\_cor1)

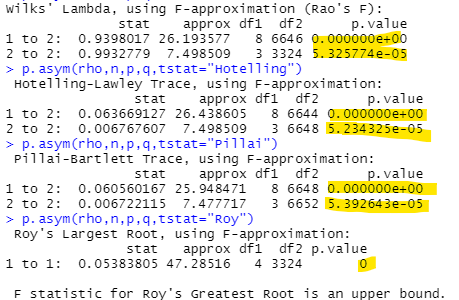
Displays the canonical loadings

can\_cor2[3:6]



Obtain the statistical significance of the dimensions

Test of canonical dimensions

rho=can\_cor1$cor

Defining the number of observations, no of variables in first set, and number of variables in second set

n=dim(DVs)[1]

p=length(DVs)

q=length(IVs)

Calculating the F approximations using different test statistics

p.asym(rho,n,p,q,tstat="Wilks")

p.asym(rho,n,p,q,tstat="Hotelling")

p.asym(rho,n,p,q,tstat="Pillai")

p.asym(rho,n,p,q,tstat="Roy")

The above test determines the significance all dimension 1 and 2 and combined are statistically significan since p<0.05.

Calculating standardized canonical coefficients using R

Standardizing the first set of canonical coefficients(DVs)

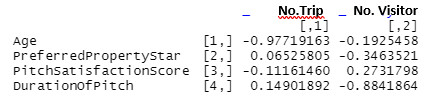
std\_coef1<-diag(sqrt(diag(cov(DVs))))

std\_coef1%\*%can\_cor1$xcoef

Standardizing the coeficents of the second set (IVs)

std\_coef2<-diag(sqrt(diag(cov(IVs))))

std\_coef2%\*%can\_cor1$ycoef



**+/increase 1 unit of =**

* Age has a strong negative correlation with number of trip and a weak negative correlation with number of visitors.
* Property Star has a weak positive correlation with number of trip and a strong positive correlation with number of visitors.
* Satisfaction Score has a negative correlation with number of trip and a positive correlation with number of visitors.
* Duration of pitch has a positive correlation with number of trip and strong negative correlation with number of visitors.