**Choosing Appropriate Statistical Analyses**

**DVs: Number of Trips and Visitors**

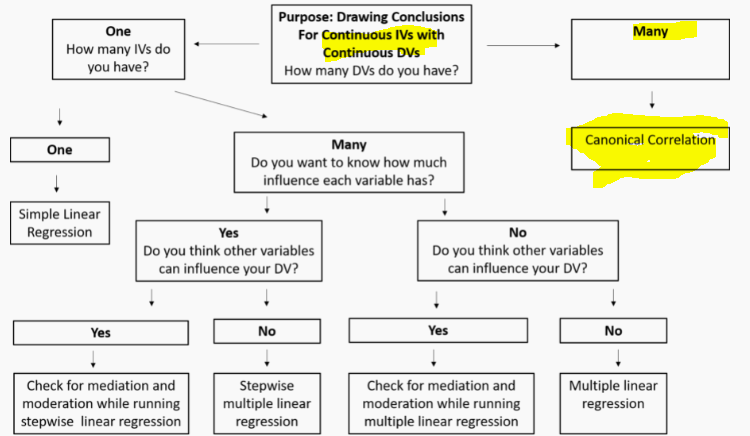
Continuous IVs correlated with both DVs:

PitchSatisfactionScore - or explore with just NumberOfTrips -> Simple Linear Regression.

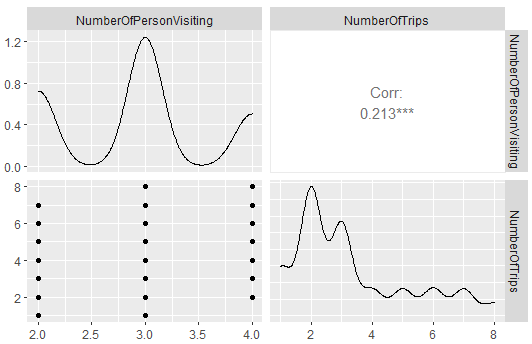
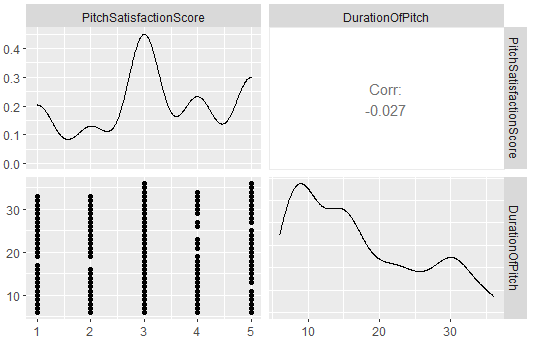
Using Canonical Correlation Analysis in R

<https://cmdlinetips.com/2020/12/canonical-correlation-analysis-in-r/#:~:text=Canonical%20Correlation%20Analysis%20%28CCA%29%20with%20cancor%20%28%29%20function,the%20linear%20projection%20of%20the%20second%20data%20matrix>.

<https://medium.com/analytics-vidhya/canonical-correlation-analysis-cca-in-r-a-non-technical-primer-b67d9bdeb9dd>



ggpairs(IVs) ggpairs(DVs)



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

Correlations cod: cormat<-matcor(IVs,DVs)

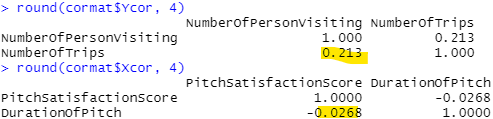
Extracting the within study correlations for set 1 and set 2 and between set cor

round(cormat$Ycor, 4)



* Not significant

round(cormat$Xcor, 4)

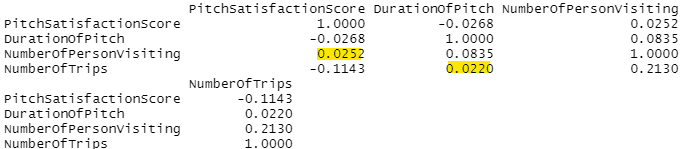


* Significant

Both sets of IVs and DVs;

cormat<-matcor(IVs,DVs)

round(cormat$XYcor, 4)

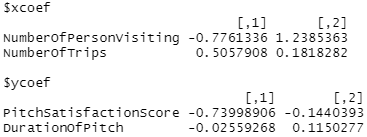


Obtain the canonical correlations

can\_cor1 <- cc(DVs, IVs)

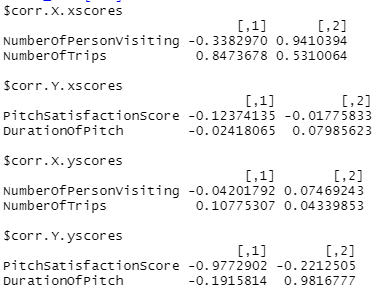
Display the canonical correlations

can\_cor1$cor 

The raw canonical coefficients can be obtained from can\_cor1 as

can\_cor1[3:4]

* A one unit increase in score would result in an decrease of 0.739 units in the value of the first canonical variate for the DVs set of variables, when the other variables are held constant.
* A one unit increase in duration would result in an decrease of 0.025 units in the value of the first canonical variate for the DVs set of variables, when the other variables are held constant.

Implement the compute function and load of the variables on the canonical dimensions

The number of canonical dimensions the same as the count of variables in the smaller set, which is 2.

Compute canonical loadings

can\_cor2 <- comput(DVs, IVs, can\_cor1)

Display canonical loadings

can\_cor2[3:6]

Obtain the statistical significance of the dimensions

Tests of canonical dimensions

rho <- can\_cor1$cor

Define number of observations, number of variables in first set, and number of variables in the second set.

n <- dim(DVs)[1]

p <- length(DVs)

q <- length(IVs)

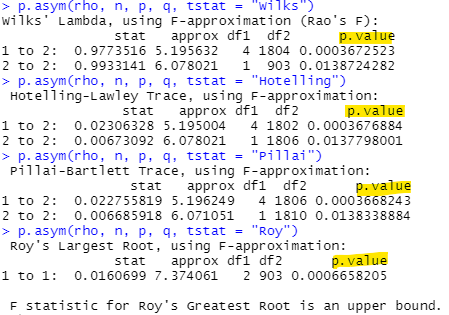
Calculate the F-approximations different test statistics

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")



With low p-value of the tests above, the combined dimensions from 1 to 2 as well as the 1st and 2nd dimensions are significant.

Calculating standardized canonical coefficients using R

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")

Standardized IVs canonical coefficients diagonal matrix of IVs sd's

s1 <- diag(sqrt(diag(cov(IVs))))

s1 %\*% cc1$xcoef

Standardized DVs canonical coefficients diagonal matrix of DVs sd's

s2 <- diag(sqrt(diag(cov(DVs))))

s2 %\*% cc1$ycoef

Calculating standardized canonical coefficients using R

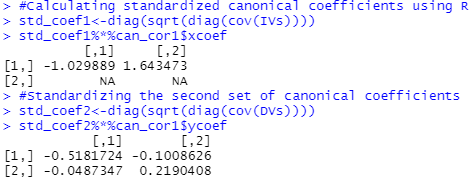
std\_coef1<-diag(sqrt(diag(cov(IVs))))

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

Standardizing the second set of canonical coefficients

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

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



DurationOfPitch

PitchSatisfactionScore

NumberOfTrips

NumberOfPersonVisiting

Interpreting the standardized canonical coefficients

in the DVs set of variables, a unit increase in NumberOfPersonVisiting value would result in a 0.52 unit standard deviation decrease on the score and duration when all the other variables in the model are held constant.

in the DVs set of variables, a unit increase in NumberOfTrip value would result in a 1 unit standard deviation decrease on the score and duration when all the other variables in the model are held constant.

* Duration and score has negative correlation with Number of trips and visitors.