CANONICAL CORRELATION ANALYSIS

HARSHAL SHIVAJI HOLAM
2102451

What is canonical correlation analysis ? :-

Canonical correlation analysis is used to identify and measure the associations among two sets of variables. Canonical correlation is appropriate in the same situations where multiple regression would be, but where are there are multiple intercorrelated outcome variables. Canonical correlation analysis determines a set of canonical variates, orthogonal linear combinations of the variables within each set that best explain the variability both within and between sets.

Dataset:-

Here we have dataset of financial dataset of particular shop on particular object.

There are 4 variables with 9995 rows. We have to find canonical correlation pairs and also we have to find canonical correlation between this pairs.

The link of dataset is "https://github.com/Harsh8793/Cannonical-correlation-analysis/blob/main/financial.xlsx"

Analysis:-

We have dataset of financial dataset of particular shop on particular object. So first we find correlation in variables in R studio.

```
> library(readxl)
> data <- read_excel("E:/MSC/MSC SEM 2/A_multivariate/multivariate project/ca.xlsx")
> View(data)
> x1=data$Sales
> x2=data$Quantity
> x3=data$Discount
> x4=data$Profit
> cor(data)
               Sales   Quantity    Discount      Profit
Sales     1.00000000 0.20079477 -0.02819012  0.47906435
Quantity  0.20079477 1.00000000  0.00862297  0.06625319
Discount -0.02819012 0.00862297  1.00000000 -0.21948746
Profit    0.47906435 0.06625319 -0.21948746  1.00000000
> |
```

We get correlation between variables. Variable are sales, quantity , discount , profit.

Now we analyse correlation matrix

```
>> S=[1 0.20079 -0.02819 0.479064;0.20079 1 0.008622 0.066253;-0.02819 0.008622 1 -0.21948;0.4790 0.06625 -0.21948 1]
S =

   1.0000e+00   2.0079e-01  -2.8190e-02   4.7906e-01
   2.0079e-01   1.0000e+00   8.6220e-03   6.6253e-02
  -2.8190e-02   8.6220e-03   1.0000e+00  -2.1948e-01
   4.7900e-01   6.6250e-02  -2.1948e-01   1.0000e+00

>> #now we partition the variance covariance matrix
>> s11=S(1:2,1:2)
s11 =

   1.0000   0.2008
   0.2008   1.0000

>> s12=S(1:2,3:4)
s12 =

  -2.8190e-02   4.7906e-01
   8.6220e-03   6.6253e-02

>> s21=S(3:4,1:2)
s21 =

  -2.8190e-02   8.6220e-03
   4.7900e-01   6.6250e-02

>> s22=S(3:4,3:4)
s22 =

   1.0000  -0.2195
  -0.2195   1.0000
```

```
>> A=(s11^(-1/2)*s12*(inv(s22))*s21*s11^(-1/2))
A =

   2.3611e-01   9.8481e-03
   9.8450e-03   5.7742e-04

>> [v d]=eig(A)
v =

   0.999131  -0.041702
   0.041689   0.999130

d =

Diagonal Matrix

   2.3652e-01            0
            0   1.6651e-04

>> #according to eigen value select eigen vector which has highest eigen value
>> e1=v(:,1)
e1 =

   0.999131
   0.041689
```

```
>> a=e1'*s11^(-1/2)
a =

   1.010401  -0.060580

>> #U1=1.010401*x1(1)-0.060580*x2(2)
>> B=(s22^(-1/2)*s21*(inv(s11))*s12*s22^(-1/2))
B =

   7.7989e-04   1.2024e-02
   1.2026e-02   2.3591e-01

>> [v1 d1]=eig(B)
v1 =

  -0.998701  -0.050938
   0.050947  -0.998702

d1 =

Diagonal Matrix

   1.6651e-04            0
            0   2.3652e-01
```

```
>> #according to eigen value select eigen vector which has highest eigen value
>> e2=v1(:,2)
e2 =

  -0.050938
  -0.998702

>> b=e2'*s22^(-1/2)
b =

  -0.1649  -1.0232

>> #v1=-0.1649*x1(1)-1.0232*x2(2)
##now canonical pairs are
#U1=1.010401*x1(1)-0.060580*x2(2)
#v1=-0.16490*x1(1)-1.023200*x2(2)
>> #now cannonical correlation between pair is
>> rho=d(1,1)
rho = 0.2365
>> cor=sqrt(rho)
cor = 0.4863
>> |
```

Conclusion :-

1. First canonical pair is given by
2. U1=1.010401*x1^(1)-0.060580*x2^(2) and v1=-0.1649*x1(1)-1.0232*x2(2)
3. Here we get 48% correlation among first canonical pair