**MODERN COLLEGE OF ARTS,SCI. & COMM. PUNE-05.**

**DEPARTMENT OF STATISTICS.**

**ST- 28 M.Sc.( I )**

**Date:**

**Practical No. 13 Submission date:**

**FACTOR ANALYSIS**

Q.1 In a consumer preference study a random sample of customers were asked to rate several attributes of a new product. The response on a 7-point somatic differential scale, were tabulated and the attribute correlation matrix constructed in given as follows.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Attribute (Variable) |  | 1 | 2 | 3 | 4 | 5 |
| Taste | 1 | 1.00 | 0.02 | 0.96 | 0.42 | 0.01 |
| Good buy for money | 2 | 0.02 | 1 | 0.13 | 0.71 | 0.85 |
| Flavor | 3 | 0.96 | 0.13 | 1 | 0.5 | 0.11 |
| Suitable for snack | 4 | 0.42 | 0.71 | 0.5 | 1 | 0.79 |
| Provides lot of energy | 5 | 0.01 | 0.85 | 0.11 | 0.79 | 1 |

Assuming m=2. Calculate

a) matrix of factor loading L.

b) Communalities

c) Specific variances.

d) Proportion of total population variance explained by the first common factor.

Q.2 Consider a sample correlation matrix.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 |  |  |  |  |  |
|  | 0.577 | 1 |  |  |  |  |
| Rs = | 0.509 | 0.599 | 1 |  |  |  |
|  | 0.387 | 0.389 | 0.436 | 1 |  |  |
|  | 0.462 | 0.322 | 0.426 | 0.523 | 1 |  |

Assuming m=2 factor model. Estimate factor loadings, communalities, specific variances and proportion of total (standardized) sample variance explained by each factor using 1) principal comp. method, 2) maximum likelihood method.

**ALGORITHM**

**FACTOR ANALYSIS BY PC METHOD:-**

* Enter correlation matrix .
* Find eigen values and eigen vector of correlation matrix.
* **L=**

**where, m= number of factors ,p=numbers of variables & L(loading matrix )**

**F1=√λ1\*e1,F2=√λ2\*e2,…., Fm=√λm\*em are the columns of matrix L.**

**λ1 ≥ λ2 ≥ … ≥λm  are eigen values & e1,e2,…em are eigen vectors .**

* **Communality hjj2**

**h112=l112+ l122 +…+l1m2**

**h222=l212+ l222 +…+l2m2**

**………**

**hpp2=lp12+ lp22 +…+lpm2**

* **Specific variance :-**

**Ѱi=1- hjj2**

* **Proportion of total variation explained by actors:-**

Proportion of total variation explained by 1st factor

Proportion of total variation explained by 1st two factor

**FACTOR ANALYSIS BY MAXIMUM LIKELIHOOD METHOD:-**

**BY USING MINTAB PROCESS:-**

Stat--🡪multivariate--🡪factor analysis---🡪maximum likelihood method---🡪ok

**SOLUTION**

Q1)

1. **matrix of factor loading L.**

>> A=[1 0.02 0.96 0.42 0.01;0.02 1 0.13 0.71 0.85; 0.96 0.13 1 0.5 0.11; 0.42 0.71 0.5 1 0.79; 0.01 0.85 0.11 0.79 1]

A =

1.000000 0.020000 0.960000 0.420000 0.010000

0.020000 1.000000 0.130000 0.710000 0.850000

0.960000 0.130000 1.000000 0.500000 0.110000

0.420000 0.710000 0.500000 1.000000 0.790000

0.010000 0.850000 0.110000 0.790000 1.000000

>> [v e]=eig(A)

v =

7.0178e-01 -1.3866e-01 -9.8485e-02 -6.0722e-01 3.3145e-01

7.1675e-02 2.8212e-01 -7.4256e-01 3.9003e-01 4.6016e-01

-7.0872e-01 -1.1700e-01 -1.6841e-01 -5.5651e-01 3.8206e-01

1.6564e-03 5.6824e-01 6.0158e-01 7.8065e-02 5.5598e-01

9.0126e-03 -7.5140e-01 2.2054e-01 4.0419e-01 4.7256e-01

e =

Diagonal Matrix

0.033677 0 0 0 0

0 0.102409 0 0 0

0 0 0.204490 0 0

0 0 0 1.806332 0

0 0 0 0 2.853090

>> f1=sqrt(e(5,5))\*v(:,5)

f1 =

0.5599

0.7773

0.6453

0.9391

0.7982

>> f2=sqrt(e(4,4))\*v(:,4)

f2 =

-0.8161

0.5242

-0.7479

0.1049

0.5432

>> L=[f1 f2]

L =

0.5599 -0.8161

0.7773 0.5242

0.6453 -0.7479

0.9391 0.1049

0.7982 0.5432

1. **Communalities**

**>> h11sq=((L(1,1))^2)+((L(1,2))^2)**

**h11sq = 0.9795**

**>> h22sq=((L(2,1))^2)+((L(2,2))^2)**

**h22sq = 0.8789**

**>> h33sq=((L(3,1))^2)+((L(3,2))^2)**

**h33sq = 0.9759**

**>> h44sq=((L(4,1))^2)+((L(4,2))^2)**

**h44sq = 0.8929**

**>> h55sq=((L(5,1))^2)+((L(5,2))^2)**

**h55sq = 0.9322**

**>> h=[h11sq;h22sq;h33sq;h44sq;h55sq]**

**h =**

**0.9795**

**0.8789**

**0.9759**

**0.8929**

**0.9322**

**c) Specific variances.**

**>> psi1=1-h11sq**

**psi1 = 0.020539**

**>> psi2=1-h22sq**

**psi2 = 0.1211**

**>> psi3=1-h33sq**

**psi3 = 0.024117**

**>> psi4=1-h44sq**

**psi4 = 0.1071**

**>> psi5=1-h55sq**

**psi5 = 0.067769**

**>> v1=(e(5,5)/5)\*100**

**v1 = 57.062**

Therefore, the percentage of variation explained by 1st factor is **57.062%**

**Q)2**

**1) FACTOR ANALYSIS BY USING PRINCIPAL COMPONANT METHOD :-**

>> A=[1 0.577 0.509 0.387 0.462;0.577 1 0.599 0.389 0.322; 0.509 0.599 1 0.436 0.426;0.387 0.389 0.436 1 0.523 ;0.462 0.322 0.

426 0.523 1]

A =

1.0000 0.5770 0.5090 0.3870 0.4620

0.5770 1.0000 0.5990 0.3890 0.3220

0.5090 0.5990 1.0000 0.4360 0.4260

0.3870 0.3890 0.4360 1.0000 0.5230

0.4620 0.3220 0.4260 0.5230 1.0000

>> [V E]=eig(A)

V =

0.4513 0.3866 -0.6117 -0.2403 0.4636

-0.6762 0.2065 0.1782 -0.5093 0.4571

0.4000 -0.6624 0.3351 -0.2604 0.4702

0.1756 0.4720 0.5408 0.5257 0.4215

-0.3850 -0.3824 -0.4352 0.5820 0.4212

E =

Diagonal Matrix

0.3429 0 0 0 0

0 0.4515 0 0 0

0 0 0.5397 0 0

0 0 0 0.8092 0

0 0 0 0 2.8567

>> F1=sqrt(E(5,5))\*V(:,5)

F1 =

0.7836

0.7726

0.7947

0.7123

0.7119

>> F2=sqrt(E(4,4))\*V(:,4)

F2 =

-0.2162

-0.4581

-0.2343

0.4729

0.5235

**>>loading matrix**

>> L1=[F1 F2 ]

L1 =

0.7836 -0.2162

0.7726 -0.4581

0.7947 -0.2343

0.7123 0.4729

0.7119 0.5235

**>> #communality**

>> H11sq=((L1(1,1))^2)+((L1(1,2))^2)

H11sq = 0.6607

>> H22sq=((L1(2,1))^2)+((L1(2,2))^2)

H22sq = 0.8068

>> H33sq=((L1(3,1))^2)+((L1(3,2))^2)

H33sq = 0.6864

>> H44sq=((L1(4,1))^2)+((L1(4,2))^2)

H44sq = 0.7310

>> H55sq=((L1(5,1))^2)+((L1(5,2))^2)

H55sq = 0.7809

>> H=[H11sq;H22sq;H33sq;H44sq;H55sq]

H =

0.6607

0.8068

0.6864

0.7310

0.7809

**>>#specific variance**

**>> PSi1=1-H11sq**

**PSi1 = 0.3393**

**>> PSi2=1-H22sq**

**PSi2 = 0.1932**

**>> PSi3=1-H33sq**

**PSi3 = 0.3136**

**>> PSi4=1-H44sq**

**PSi4 = 0.2690**

**>> PSi5=1-H55sq**

**PSi5 = 0.2191**

**>> V1=(E(5,5)/5)\*100**

**V1 = 57.134**

Therefore, the percentage of variation explained by 1st factor is **57.134** **%**

Therefore, the percentage of variation explained by 1st factor is **77.927%**

>> V1=((E(5,5)+E(4,4))/5)\*100

V1 = **73.317**

Therefore, the percentage of variation explained by 1st 2 factor is **73.317%**

**2) FACTOR ANALYSIS BY USING MAXIMUM LIKELIHOOD METHOD**

**Data Display**

Matrix Copy

1.000 0.577 0.509 0.387 0.462

0.577 1.000 0.599 0.389 0.322

0.509 0.599 1.000 0.436 0.426

0.387 0.389 0.436 1.000 0.523

0.462 0.322 0.426 0.523 1.000

**Factor Analysis: Copy**

Maximum Likelihood Factor Analysis of the Correlation Matrix

**Unrotated Factor Loadings and Communalities**

Variable Factor1 Factor2 Communality

Var 1 0.690 -0.165 0.503

Var 2 0.712 -0.496 0.753

Var 3 0.690 -0.223 0.525

Var 4 0.619 0.097 0.392

Var 5 0.773 0.461 0.811

Variance 2.4395 0.5452 2.9847

% Var 0.488 0.109 0.597

**Rotated Factor Loadings and Communalities**

**Varimax Rotation**

Variable Factor1 Factor2 Communality

Var 1 0.599 -0.380 0.503

Var 2 0.852 -0.165 0.753

Var 3 0.640 -0.340 0.525

Var 4 0.362 -0.511 0.392

Var 5 0.208 -0.876 0.811

Variance 1.6679 1.3168 2.9847

% Var 0.334 0.263 0.597

Factor Score Coefficients

Variable Factor1 Factor2

Var 1 0.182 -0.043

Var 2 0.666 0.190

Var 3 0.217 -0.019

Var 4 0.046 -0.117

Var 5 -0.207 -0.848

Therefore Percentage of variation explained by 1st 2 factors are **57.9%** 57.9%**57.9%.**