Slide-4

Factor Analysis is a statistical method used to analyze the relationships among a set of observed variables by explaining the correlations or covariance between them in terms of a smaller number of unobserved variables called factors.

→ used for data reduction and summarization.

Factor analysis can have significant impact on correlation by identifying underlying factors that account for correlations among a set of observed variables.

i) Factor analysis reduces the complexity of datasets by grouping correlated variables under a smaller set of unobserved variable (factors). This helps in understanding the structure of relationships without having to look at individual correlations between every variable.

ii) Beyond reducing complexity, factor analysis helps identify hidden relationships between variables that may not be immediately obvious. It reveals patterns that explains why certain variables tend to move together.

iii) Factor analysis can help mitigate multicolinearity by combining correlated variables into factors.

Q: Determine the suitability of factor analysis.

[i] Bartlett's Test: বাদ

[ii] Kaiser-Meyer-Olkin (KMO) Measure: is a statistical method to determine how suited the data is for factor analysis. The statistic measures the proportion of variance among variables that might be common variance. The higher the proportion, the higher the KMO value, and the more suited the data is for factor analysis.

Range: 0 to 1

value >0.6 is acceptable

Extraction method - 1 (Centroid method)

Conditions:

① If the correlation matrix is positive manifold, meaning the correlation values are positive, the all the values will stay same as before.

② If not positive, it needs to be reflected before the calculation of first centroid, means all the negative signs will be converted to positive.

variables

variables		1	2	3	4	5	6	7	8
	1	1.00							
	2	.709	1.00						
	3	.204	.051	1.00					
	4	.081	.089	0.671	1.00				
variables	5	.626	.581	0.123	.022	1.00			
	6	.113	.098	.689	.798	.047	1.00		
	7	.155	.083	.582	.613	.201	.801	1.00	
	8	.774	.652	.072	.111	.724	.120	.152	1.00

Step-1: Sum of the co-efficient (including the diagonal unity) in each column of the correlation matrix. Sum of columns (T). The Sum of each column divided by $\sqrt{T}$

R=

	1	2	3	4	5	6	7	8
1	1·00	·709	·204	·081	·626	·113	·155	·774
2	·709	1·00	·051	·089	·581	·098	·083	·652
3	·204	·051	1·00	·671	·123	·689	·582	·072
4	·081	·089	·671	1·00	·022	·798	·613	·111
5	·626	·581	·123	·022	1·00	·047	·201	·724
6	·113	·098	·689	·798	·047	1·00	·801	·120
7	·155	·083	·582	·613	·201	·801	1·00	·152
8	·774	·652	·072	·111	·724	·120	·152	1·00
Column Sums:	3·662	3·263	3·392	3·385	3·324	3·666	3·587	3·605

Sum of the Columns Sums: 27·884

$\sqrt{T} = 5{\cdot}281$

• Divide the Column sum by $\sqrt{T}$

First Centroid Factor A:

$$\frac{3{\cdot}662}{5{\cdot}281}, \frac{3{\cdot}263}{5{\cdot}281}, \frac{3{\cdot}392}{5{\cdot}281}, \frac{3{\cdot}385}{5{\cdot}281}, \frac{3{\cdot}324}{5{\cdot}281}, \frac{3{\cdot}666}{5{\cdot}281}, \frac{3{\cdot}587}{5{\cdot}281}, \frac{3{\cdot}605}{5{\cdot}281}$$

$= 0{\cdot}693, 0{\cdot}618, 0{\cdot}642, 0{\cdot}641, 0{\cdot}629, 0{\cdot}694, 0{\cdot}679, 0{\cdot}683$

Step-2: Find the factor cross product (Q_1)

$Q_1(1,1) = (.693 \times .693) = 0.480$

$Q_1 =$

	.693	.618	.642	.641	.629	.694	.679	.683
.693	.480	.428	.445	.444	.436	.481	.471	.473
.618	.428	.382	.397	.396	.389	.429	.420	.422
.642	.445	.397	.412	.412	.404	.446	.436	.438
.641	.444	.396	.412	.411	.403	.445	.435	.438
.629	.436	.389	.404	.403	.396	.437	.427	.430
.694	.481	.429	.446	.445	.437	.482	.471	.474
.679	.471	.420	.436	.435	.427	.471	.461	.464
.683	.473	.422	.438	.438	.430	.474	.464	.466

Step-3:

$R_1 = R - Q_1$

	1	2	3	4	5	6	7	8
1	.520	.281	-.241	-.363	.190	-.368	-.316	.301
2	.281	.618	-.346	-.307	.192	-.331	-.337	.230
3	-.241	-.346	.588	.259	-.281	.243	.146	-.366
4	-.363	-.307	.259	.589	-.381	.353	.178	-.327
5	.190	.192	-.281	-.381	.604	-.390	-.217	.294
6	-.368	-.331	.243	.353	-.390	.518	.330	-.354
7	-.316	-.337	.146	.178	-.226	.330	.539	-.312
8	.301	.230	-.366	-.327	.294	-.354	-.312	.534

Step-4: Reflected matrix, negative signs should be converted to positive.

	1	2	3*	4*	5	6*	7*	8
1	·520	·281	·241	·363	·190	·368	·316	·301
2	·281	·618	·346	·307	·192	·331	·337	·230
3*	·241	·346	·588	·259	·281	·243	·146	·366
4*	·363	·307	·259	·589	·381	·353	·178	·327
5	·190	·192	·281	·381	·604	·390	·217	·294
6*	·368	·331	·243	·353	·390	·518	·330	·354
7*	·316	·337	·146	·178	·226	·330	·539	·312
8	·301	·230	·366	·327	·294	·354	·312	·534

Column Sums:

2·580 2·642 2·470 2·757 2·558 2·887 2·375 2·718

Sum of Column Sums: (T) = 20·987 $\therefore \sqrt{T} = 4·581$

Second $ Centroid factor B:

$\frac{2·580}{4·581}, \frac{2·642}{4·581}, \frac{2·470}{4·581}, \frac{2·757}{4·581}, \frac{2·558}{4·581}, \frac{2·887}{4·581}, \frac{2·375}{4·581}, \frac{2·718}{4·581}$

·563, ·577, -·539, -·602, ·558, -·630, -·518, ·593

Step-5: Factor Extraction : To identify initial factors

Eigen Value = Sum of squared of variables

Variables	Factor loading		Communality (h^2)
	Centroid factor A	Centroid factor B	
1	.693	.563	$(.693)^2 + (.563)^2 = .797$
2	.618	.577	.715
3	.642	-.539	.703
4	.641	-.602	.773
5	.629	-.558	.707
6	.694	-.630	.879
7	.679	-.518	.729
8	.683	-.593	.818
Eigen Value	3.490	2.631	6.121
Proportion of Total variance	$\frac{3.490}{8} = .44 = (44\%)$	$\frac{2.631}{8} = .33 = (33\%)$	$.77 = (77\%)$
Proportion of Common Variance	$\frac{3.490}{6.121} = .57\ (57\%)$	$\frac{2.631}{6.121} = .43\ (43\%)$	100%

Extraction method-2 (PCA) : is a procedure to convert a set of observation of possible correlated variables into a set of values of linearly uncorrelated variables.

R matrix:

. . . .

Column Sums: U_{a_1}

3·662 3·263 3·392 3·385 3·324 3·666 3·587 3·605

Normalizing Factor:

$$\sqrt{(3.662)^2+(3.263)^2+(3.392)^2+(3.385)^2+(3.324)^2+(3.666)^2+(3.587)^2+(3.605)^2}$$

$= 9.868$

* $V_{a_1} = \frac{U_{a_1}}{\text{Normalizing factor}} = \left[\frac{3.662}{9.868}, \dots\right.$

$= [.371, .331, .344, .343, .337, .372, .363, .365]$

Now, multiply each Row of R with V_{a_1}, we get U_{a_2}

$U_{a_2} = [1.296, 1.143, 1.201, 1.201, 1.165, 1.308, 1.280, 1.275]$

Normalizing factor = 3·493

* $V_{a_2} = \frac{U_{a_2}}{\text{Normalizing factor}} = [.371, .327, .344, .344, .334, .374, .366, .365]$

If we observe, the values of Va_1 and Va_2 are nearly identifical So, the convergence has occured. Va_1 will be the characteristics vector we will multiply it by the normalizing factor of $u_{a_2} = \sqrt{3.493}$
$= 1.868$

we will get PC1;

Example: $.371 \times 1.868 = .69$
$.331 \times 1.868 = .62$

⋮

PC1 = [.69, .62, .64, .64, .63, .70, .68, .68]

Now if we want to find PC2, using the PC1, follow the steps of previous math from step-2

Variables	PC1	PC2	h^2
1	.69	.57	.801
2	.62	.59	.733
3.	.64	−.52	.680
4	.64	−.59	.758
5	.63	+.57	.722
6	.70	−.61	.862
7	.68	−.49	.703
8	.68	−.61	.835
Eigenvalue	3.4914	2.6007	6.0921
Proportion of total variance	3.4914/8 = .436 (43.6%)	2.6007/8 = .325 (32.5%)	.761 = (76%)
Proportion of common variance	.573 (57%) 3.4914/6.0921	.427 (43)% 2.6007/6.0921	1.00 (100%)