FACTOR ANALYSIS

DR. HEMAL PANDYA

INTRODUCTION

Factor analysis is very useful method of reducing data complexity by reducing the number of variables being studied.

"What are the underlying significant drivers of customer's behavior?"

FA is good way of resolving this confusion and identify <u>LATENT or underlying factors</u> from an area of seemingly important variables.

Thus, it is also called DATA REDUCTION technique.

INTRODUCTION

Factor analysis is an INTERDEPENDENCE TECHNIQUE. Thus, there is no distinction between dependent and independent variables.

FA is used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimension i.e. Factor.

Thus, analysis can play a unique role in the application of other multivariate techniques.

EXAMPLE

- **Data Reduction technique**
- Customer's ratings of a fast food restaurant.

- Food taste
- Food temperature
- Freshness

Food quality

- Waiting time
- Cleanliness
- Friendliness of employees

Service quality

USES OF FA

- **SCALE CONSTRUCTION:**
- Example:-15-item scale to measure job satisfaction.
- At first step- Generate large number of statements, numbering say 100 or so as a part of exploratory research.
- Assume that we get 3 factors out of it.
- I want to construct 15- item scale to measure job satisfaction.
- Separate 5 items from each factors having highest factor loading.

USES OF FA

- **Establish antecedents: (Data summarization)**

▪ This method reduces multiple input variables into grouped factors.

• For example All the variables that measure safety clauses in mutual fund could be reduced to a factor called SAFETY CLAUSE**.**

USES OF FA

- **Psychographic profiling:**

- **Psychographics** can be defined as a quantitative methodology used to describe consumers on psychological attributes.

- When a relatively complete profile of a person or group's psychographic make-up is constructed, this is called a "psychographic profile".

- Some categories of psychographic factors used in market segmentation include:

• activity, interest, opinion (AIOs)

• attitudes

• values

• behavior

USES OF FA

❑ **Segmentation analysis:**

▪ Factor analysis could also be used for segmentation.

▪ Example: There could be different sets of two – wheelers customers owing two-wheelers because of different importance they give to factors like….

- Prestige,
- Economy consideration,
- Functional features,
- Traffic and time etc….

USES OF FA

❑ **Marketing studies:**

▪ This technique has extensive use in the field of marketing and can be successfully used for new product development; product acceptance research, developing advertising copy, pricing studies, branding studies and so on….

▪ For example it can be used to…

• identify the attributes of brands that influence consumer's choice.

• identify the characteristics of price sensitive customers etc..

CONDITIONS FOR A FA EXERCISE

- **FA exercise requires metric data. i.e.** the data should be either interval or ratio scale in nature.
- The variables for FA are identified through exploratory research. (By Literature Review or informal interviews, focused group discussion etc.)(Generally in survey research 5/7 point Likert scale may be used.)
- As the responses to different statements are obtained through **different scales, all the responses need to be standardized.**
- The size of sample respondents should be at least **four to five times more** than the number of variables (no. of statements)
- **The basic principle behind FA is that the initial set of variables should be highly correlated.**
- If the correlation coefficients between all the variables are small, FA may not be appropriate technique.
- **For application of FA the value of KMO (KAISER-MEYER OLKIN)- should be greater than 0.5.**
- The KMO statistic compares the magnitude of observed correlation coefficients with the magnitude of partial correlation coefficients. (RANGE OF KMO STATISTIC ➔ 0 to 1)
- **Bartlett statistic for sphericity:** Bartlett's test of sphericity tests the hypothesis that your correlation matrix is an identity matrix, which would indicate that your variables are unrelated and therefore unsuitable for structure detection.

STEPS IN FA EXERCISE

- **There are two stages in FA.**

STAGE 1: FACTOR EXTRACTION PROCESS

- Identify how many factors will be extracted from the data.
- This could be accomplished by various methods like **the centroid method, the principal component method and maximum likelihood method.**
- There is a rule of thumb based on the computation of EIGEN VALUES, to determine how many factors to extract.
- The Eigen value for a given factor measures the variance in all the variables which is accounted for by that factor.
- Eigen values measure the amount of variation in the total sample accounted for by each factor.
- If a factor has a low Eigen value, then it is contributing little to the explanation of variances in the variables and may be ignored as redundant with more important factors.
- In short, we can have as many factors as there are original variables But the objective is to reduce the variables to a fewer number of factors, RETAIN those WITH EIGEN VALUE of 1 or more.
- (Before extraction it is assumed that each of the original variable has an Eigen value = 1)

STEPS IN FA EXERCISE

STAGE : 2 ROTATION OF FACTORS / PRINCIPAL COMPONENT

This is actually optional, but highly recommended.

After the number of extracted factors is decided in stage 1, the next task of the researcher is to interpret and name the factor. The factor matrix is used for this purpose.

The original factor matrix is un-rotated and is a part of output from stage 1.

The rotated factor matrix comes about in stage 2.

Most of the computer software would give options for orthogonal rotation, Varimax rotation and Oblique rotation.

Generally, the Varimax rotation is used as this results in independent factors.

TERMS

❖**Factor:**
Linear combinations of original variables.

❖**Factor loadings:**
Correlations between original variable and the factors.

❖**Bartlett test of sphericity:-**
Statistical test for overall significance of all correlations within a correlation matrix.

❖**Factor rotation:-**
Process of manipulation or adjusting the factor axes.

❖**Common factor analysis:-**
The extracted factors are based only on the common variance with variable specific and error variance excluded.

TERMS

❖**Communality:-**
Total amount of variance an original variable shares with all other variables included in the analysis.

❖**Eigen values:-**
Column sum of squared loadings for a factor; also referred to as a latent root. It represents the amount of variance accounted for by a factor.

❖**Confirmatory factor analysis:** In confirmatory factor analysis (CFA), researchers can specify the number of factors required in the data and which measured variable is related to which latent variable.

❖**Exploratory factor analysis:**
In EFA data is simply explored and provides information about the numbers of factors required to represent the data.

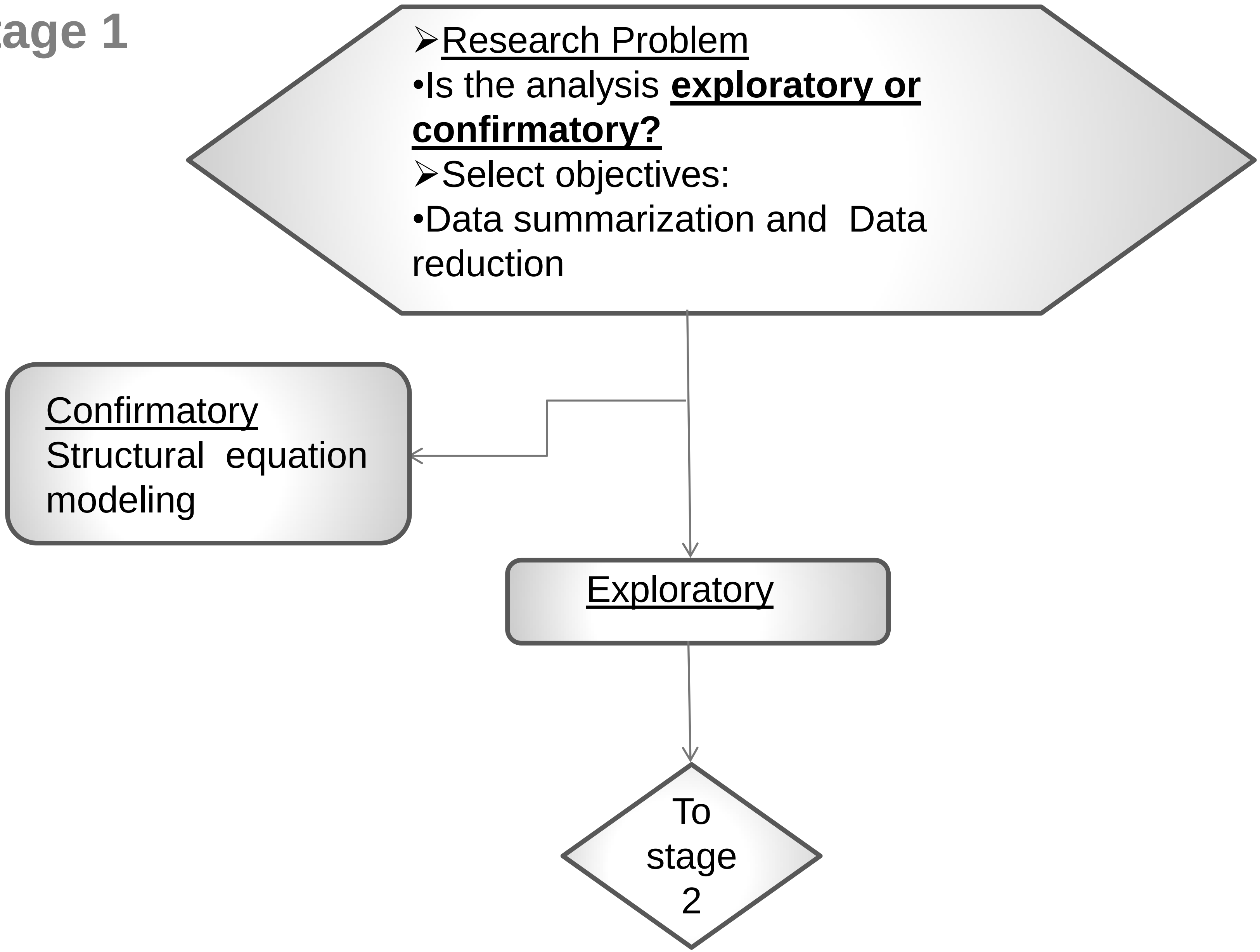
TERMS

❖**Orthogonal:-** Mathematical independence (no correlation) of factor axes to each other.(i.e., at right angle, or 90 degrees).

- ***Varimax Method*.** An orthogonal rotation method that minimizes the number of variables that have high loadings on each factor. This method simplifies the interpretation of the factors.
- ***Quartimax Method*.** A rotation method that minimizes the number of factors needed to explain each variable. This method simplifies the interpretation of the observed variables.
- ***Equamax Method*. .** A rotation method that is a combination of the varimax method, which simplifies the factors, and the quartimax method, which simplifies the variables. The number of variables that load highly on a factor and the number of factors needed to explain a variable are minimized
- ***Direct Oblimin Method*.** A method for oblique (non-orthogonal) rotation. When delta (Correlation coefficient) equals 0 (the default), solutions are most oblique. As delta becomes more negative, the factors become less oblique. To override the default delta of 0, enter a number less than or equal to 0.8
- ***Promax Rotation*** An oblique rotation, which allows factors to be correlated. This rotation can be calculated more quickly than a direct oblimin rotation, so it is useful for large datasets.

FLOW CHART

Stage 1



FLOW CHART

From stage 1

Stage 2

Research Design
- What variables are included?
- How are the variables measure?
- What is the desired sample size?

Stage 3

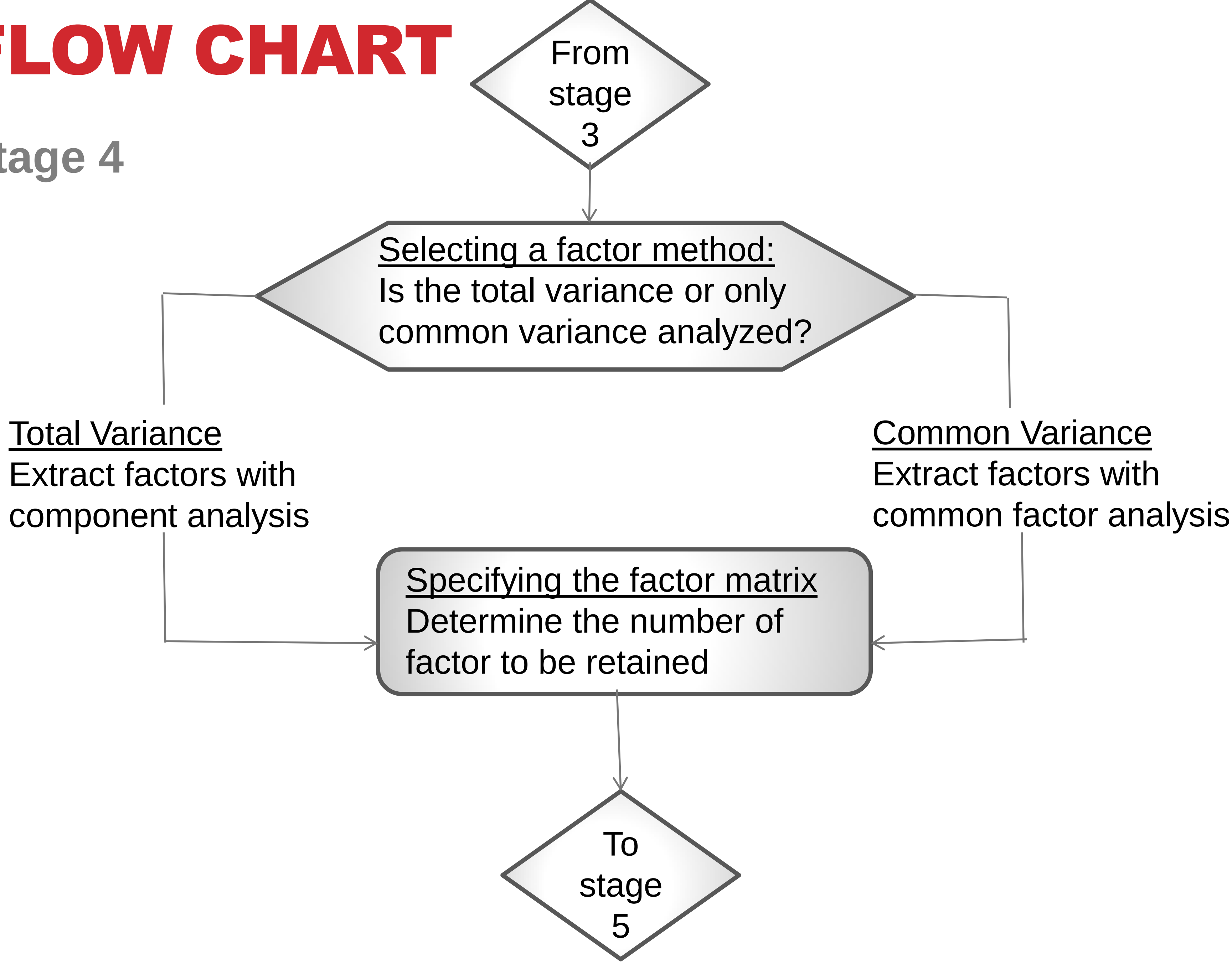
Assumptions

Statistical consideration of NORMALITY, LINEARITY and HOMOSCEDASTICITY.
HOMOGENETY OF SAMPLE

To stage 4

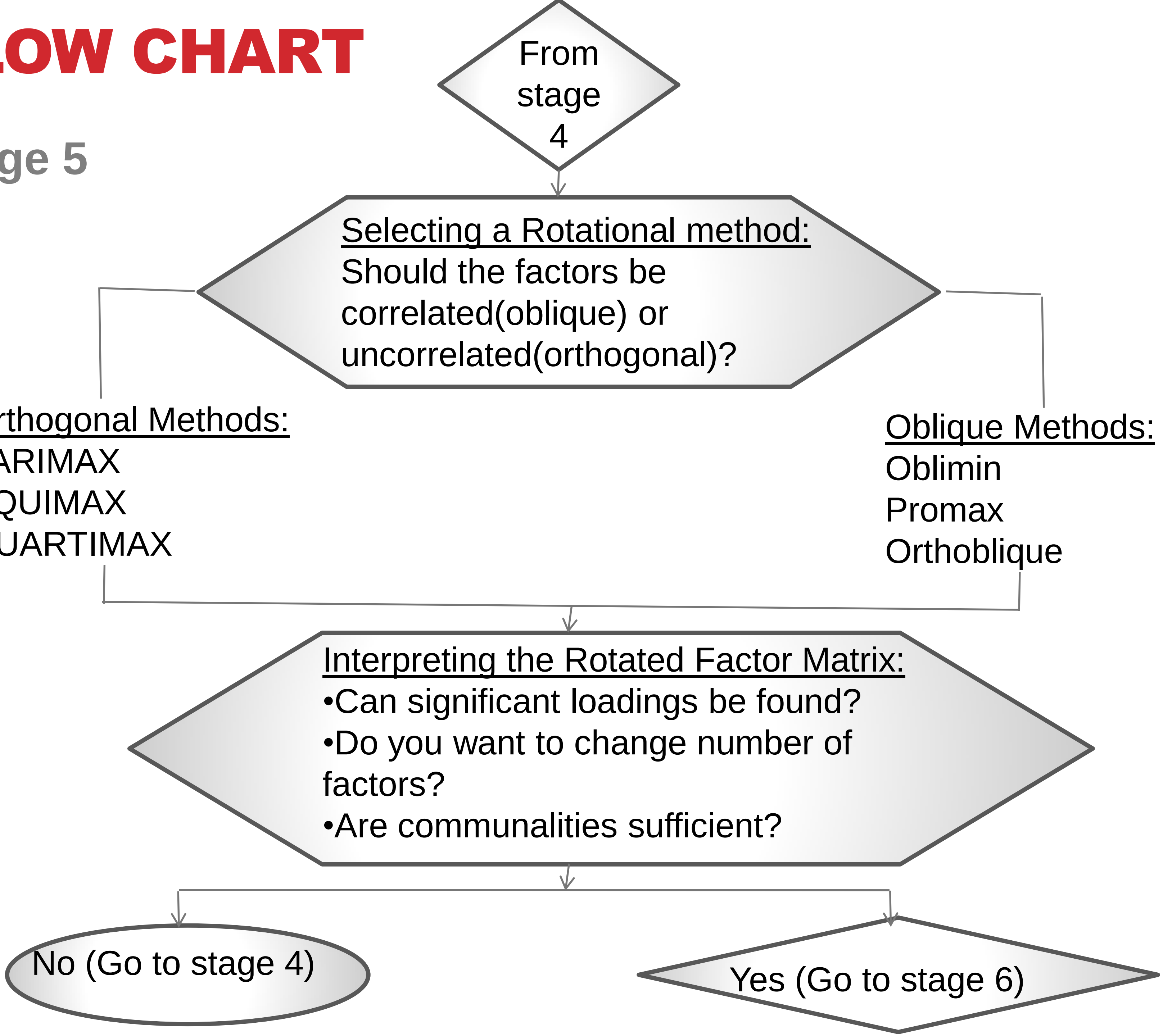
FLOW CHART

Stage 4

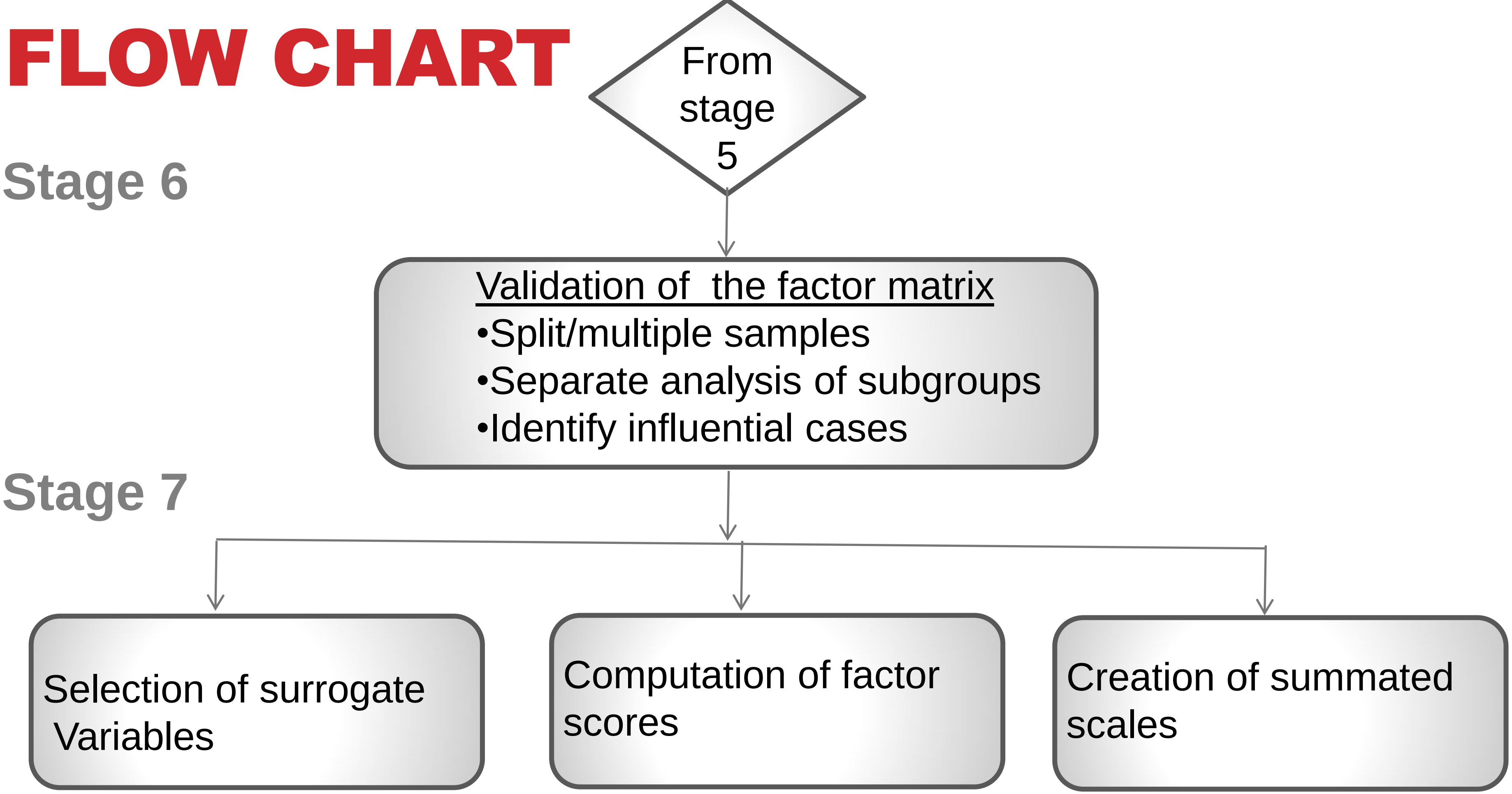


FLOW CHART

Stage 5



FLOW CHART


From stage 5
Stage 6
Validation of the factor matrix
•Split/multiple samples
•Separate analysis of subgroups
•Identify influential cases
Stage 7
Selection of surrogate Variables
Computation of factor scores
Creation of summated scales

WORKED EXAPLE

20 two-wheeler users were surveyed about their perception and image attributes of vehicles they owned.

Ten statements were as follows:

1. I use a two-wheeler because it is affordable.
2. It gives me sense of freedom to own a two-wheeler.
3. Low maintenance cost makes a two-wheeler very economical in the long run.
4. A two-wheeler is essentially a man’s vehicle.
5. I feel very powerful when I am on my two-wheeler.
6. Some of my friends who don’t have their own vehicle are jealous of me.
7. I feel good whenever I see the ad for my two-wheeler on TV, in a magazine or on a hording.
8. My vehicle gives me a comfortable ride.
9. I think two-wheelers are safe way to travel.
10. Three people should be legally allowed to travel on a two-wheeler.

INPUT DATA

S. NO.	QUESTION NO.									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1	4	1	6	5	6	5	2	3	2
2	2	3	2	4	3	3	3	5	5	2
3	2	2	2	1	2	1	1	7	6	2
4	5	1	4	2	2	2	2	3	2	3
5	1	2	2	5	4	4	4	1	1	2
6	3	2	3	3	3	3	3	6	5	3
7	2	2	5	1	2	1	2	4	4	5
8	4	4	3	4	4	5	3	2	3	3
9	2	3	2	6	5	6	5	1	4	1
10	1	4	2	2	1	2	1	4	4	1
11	1	5	1	3	2	3	2	2	2	1
12	1	6	1	1	1	1	1	1	2	2
13	3	1	4	4	4	3	3	6	5	3
14	2	2	2	2	2	2	2	1	3	2
15	2	5	1	3	2	3	2	2	1	6
16	5	6	3	2	1	3	2	5	5	4
17	1	4	2	2	1	2	1	1	1	3
18	2	3	1	1	2	2	2	3	2	2
19	3	3	2	3	4	3	4	3	3	3
20	4	3	2	7	6	6	6	2	3	6

RESULTS

Descriptive Statistics

	Mean	S. D.	N	C.V.
VAR00001	2.35	1.3089	20	55.7
VAR00002	3.25	1.4824	20	45.6
VAR00003	2.25	1.118	20	49.7
VAR00004	3.1	1.8035	20	58.2
VAR00005	2.8	1.5079	20	53.9
VAR00006	3.05	1.6051	20	52.6
VAR00007	2.7	1.4546	20	53.9
VAR00008	3.05	1.905	20	62.5
VAR00009	3.2	1.5079	20	47.1
VAR00010	2.8	1.4726	20	52.6

Lowest C.V. indicates that particular variable is most consistent. Variable 2 (sense of freedom) is having lowest c.v., thus it is the most consistent variable.

RESULTS

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.618
Bartlett's Test of Sphericity	Approx. Chi-Square	164.098
	df	45
	Sig.	.000

•The **Kaiser-Meyer- Olkin Measure of Sampling Adequacy** is a statistic that indicates the proportion of variance in your variables that might be caused by underlying factors.

•High values (close to 1.0) generally indicate that a factor analysis may be useful with your data.

•If the value is less than 0.50, the results of the factor analysis probably won't be very useful.

•**Bartlett's test of sphericity** tests the hypothesis that your correlation matrix is an identity matrix, which would indicate that your variables are unrelated and therefore unsuitable for structure detection.

•Small values (less than 0.05) of the significance level indicate that a factor analysis may be useful with your data.

RESULTS

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.883	38.828	38.828	3.883	38.828	38.828	3.841	38.409	38.409
2	2.777	27.770	66.598	2.777	27.770	66.598	2.429	24.294	62.703
3	1.375	13.747	80.346	1.375	13.747	80.346	1.764	17.643	80.346
4	.945	9.449	89.795						
5	.479	4.793	94.588						
6	.292	2.923	97.511						
7	.117	1.166	98.677						
8	.068	.680	99.356						
9	.037	.374	99.730						
10	.027	.270	100.000						

Extraction Method: Principal Component Analysis.

RESULTS

The first section of the table shows the **Initial Eigen values**.

The **Total** column gives the Eigen value, or amount of variance in the original variables accounted for by each component.

The **% of Variance** column gives the ratio, expressed as a percentage, of the variance accounted for by each component to the total variance in all of the variables.

The **Cumulative %** column gives the percentage of variance accounted for by the first *n* components.

For example, the cumulative percentage for the second component is the sum of the percentage of variance for the first and second components.

RESULTS

The second section of the table shows the extracted components.

They explain nearly 80% of the variability in the original ten variables, so you can considerably reduce the complexity of the data set by using these components, with only a 20% loss of information.

RESULTS

Component Matrix[a]

	Component		
	1	2	3
VAR00001	.176	.670	.493
VAR00002	-.136	-.608	.254
VAR00003	-.107	.820	.218
VAR00004	.966	-.036	-.097
VAR00005	.951	.166	-.136
VAR00006	.952	-.084	-.025
VAR00007	.971	.096	-.046
VAR00008	-.322	.775	-.308
VAR00009	-.069	.735	-.482
VAR00010	.161	.319	.814

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Rotated Component Matrix[a]

	Component		
	1	2	3
VAR00001	.126	.313	.780
VAR00002	-.181	-.639	-.107
VAR00003	-.116	.604	.594
VAR00004	.970	-.064	-.006
VAR00005	.964	.131	.063
VAR00006	.945	-.140	.030
VAR00007	.971	.024	.106
VAR00008	-.262	.848	.101
VAR00009	.010	.881	-.044
VAR00010	.063	-.149	.874

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

RESULTS

Communalities

	Initial	Extraction
VAR00001	1.000	.722
VAR00002	1.000	.452
VAR00003	1.000	.731
VAR00004	1.000	.945
VAR00005	1.000	.950
VAR00006	1.000	.914
VAR00007	1.000	.955
VAR00008	1.000	.799
VAR00009	1.000	.777
VAR00010	1.000	.789

Extraction Method: Principal

- Examine the communality values to assess how well each variable is explained by the factors.

- The closer the communality is to 1, the better the variable is explained by the factors.

- You can decide to add a factor if the factor contributes significantly to the fit of certain variables.

NAME OF THE FACTORS

FACTORS	NAME
FACTOR 1 ➔	‘MALE EGO’ or ‘PRIDE OF OWERNERSHIP’
FACTOR 2 ➔	‘LOW MAINTAINACE’ or ‘COMFORT’ or ‘SAFETY’
FACTOR 3 ➔	‘AFFFORDABILITY’

- ➢ Communality of variable 2 is 45.2%.
- It implies that the only 45.2% of variation in variable 2 is captured by our extracted factors.
- This may also partially explain why variable 2 is not appearing in our final interpretation of the table.

THANK YOU