MULTIVARIATE TECHNIQUES : FEW FUNDAMENTALS

RAKESH SRIVASTAVA
Deptt. of Statistics
Faculty of Science
The M.S.University of Baroda
Vadodara-390 002
E-mail: rakeshsrivastava30@yahoo.co.in

‘MULTIVARIATE METHODS IMPACT OUR LIVES EVERY DAY’

WHY MULTIVARIATE?

Most of the real life situations involve the information (data) collection on several variables of interest (usually more than two)

These techniques are being applied in many fields such as : Economics, commerce, sociology, Psychology, agriculture, Anthropology, biology and medicine.

Used to analyse huge data arising from volumionous trade points such as: super bazaars, big-bazaars, malls & mega-malls etc. (particularly data mining techniques)

To make accurate business decisions in today’s increasingly complex environment, we must analyse intricate relationships with many intervening variables. Multivariate methods are powerful analytical techniques mode for addressing such issues.

- **WHAT IS MULTIVARIATE ANALYSIS?**

- **'Analysis of multivariate data is called multivariate analysis'.**
- **<u>CLASSIFICATION :</u>**
- **Generally, multivariate data are classified into two groups- a group of dependent variables and a group of independent variables. Accordingly, one needs to study the 'extent' of relationship of dependent and independent sets.**
- **Thus, multivariate analysis can be classified into two types: (A) interdependence analysis, (B) Dependent analysis. The main components of interdependence analysis are : (i) Principal component Analysis (PCA) (ii) Factor Analysis (FA) (iii) cluster Analysis (CA) (Particularly dealing with multidimensional scaling (MDS).**

The dependent models deal with : (i) Discriminant analysis (ii) Cannonical correlation analysis (iii) Multivariate Analysis of Variance (MANOVA) (iv) Multivariate Regression Analysis (MRA)

FACTOR ANALYSIS :

(is used to summarize the information contained in a large number of variables into smaller subsets called factors)

As mentioned in many real life problems, the number of independent variables used in predicting a response variable will be too many. The difficulties in having too many independent variables are as follows:

- Increased time in data collection
- Too much expenditure in data collection
- Difficulty in making inferences
- Presence of redundant independent variables

These can be avoided using factor analysis. Factor analysis aims at 'grouping' the original input variables into 'factors' that underlie the input variables. Theoretically, the total no. of factors is equal to the total number of input variables. But after performing factor analysis, the total number of factors in the study can be 'reduced' by dropping the insignificant factors based on certain criterion.

BASIC IDEA:

Factor analysis could be used to verify your conceptualization of a construct of interest. For example in many studies, the construct of 'sales' has been observed to be composed of 'product' and expenditure on advt. lets say that for some reason you are developing a new questionnaire about 'sales' and you creat (say) 20 items. You think 10 will reflect 'product' elements and 10 'amount on advertisement' but since your items are new, you want to test your conceptualization Before you use the questionnaire on your sample, you decide to pretest it (always wise) on a group of people who are like those who will be completing your survey.

- **TYPES OF FACTOR ANALYSIS:**
- **(Two main types)**
- **Principal component analysis :**

  This method provideds 'unique solution' , so that the original data can be 'reconstructed' from the results. It looks at the 'total' variance among the variables, so that the solution generated will include as many factors as there are variables although it is unlikely that they will all meet the criteria for retention

- **COMMON FACTOR ANALYSIS:**

  This is what people generally mean when they say ` factor analysis' this family of techniques uses an estimate of 'common' variance among original variables to generate a factor solution. Because of this the number of 'Factors' will always be less than the number of original variables. So, choosing the number of factors to keep for 'further' analysis is more problematic using 'factor analysis' than in principal components.

- **FACTOR ANALYSIS BRIEF GLOSSARY :**

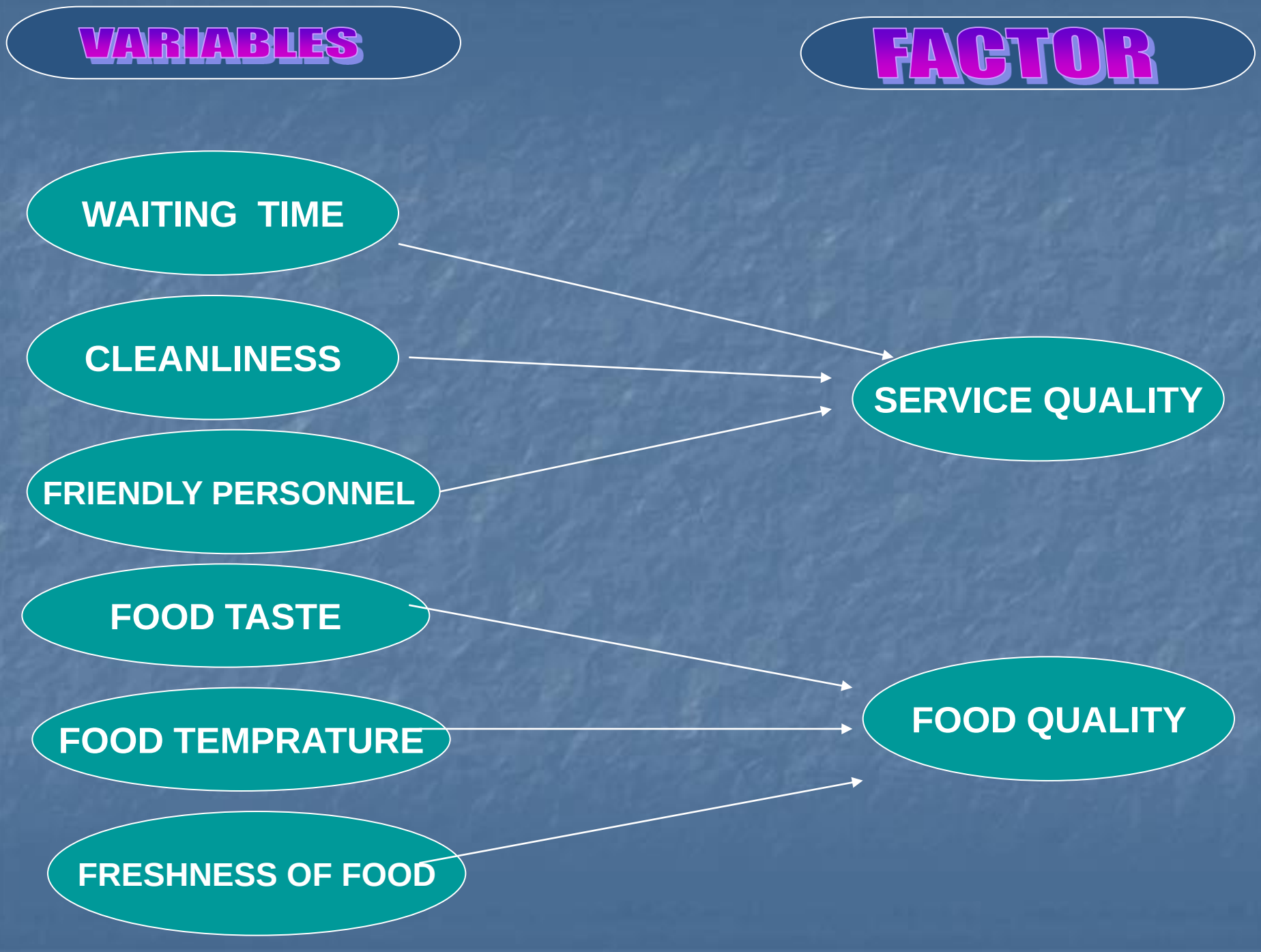
- **Factor :** An unobservable random quantify on which two or more variables load.
- **Common variance:** Variance in a variable shared with common factors. Factor analysis assumes that a variables variance is composed of three components: common, specific and error.
- **Communality:** The proportion of a variable's variance explained by a factor structure. A variables, Communality must be estimated PRIOR to per forming factor analysis. A Communality does not have to be estimated prior to performing a principal component Analysis. A communality is denoted by h2
- **Communality Estimates:** Final communality estimates are the sum of squared loadings for a variable in an orthogonal factor matrix.
- **Factor Loading:** A term used to refer to factor pattern coefficients or structure coefficients. (loading : The correlation between a variable and a facor)

- **STEPS IN CONDUCTING A FACTOR ANALYSIS :**
- Data collection and generation of the correlation matrix
- Extraction of initial factor solution
- Construction of scales or factors to use in further analyses.

AN EXAMPLE OF FACTOR ANALYSIS APPLICATION TO A FAST FOOD RESTAURANT:

VARIABLES

X_1: Try new and different things X_2: Party person X_3: People come to me X_4: Avoid tried food X_5: Likes to go out socially X_6: Friends come to come X_7: Self – confident X_8: Eat balanced, nutritious meals X_9: Buy new products X_{10}: Careful about what I eat X_{11}: Try new brands X_{12}:Friednly employees X_{13}: Fun place to eat X_{14}: Large size portions X_{15}: Fresh food	X_{16}: Reasonable price X_{17}: Attractive interior X_{18}: Excellent food taste X_{19}: Knowledgeable employees X_{20}: Proper food temperature X_{21}: Speed of service X_{22}: satisfaction X_{23}: likely to return X_{24}: likely to recommend X_{25}: Frequency of patronizing(say MC Donald's) X_{26}: Price X_{27}: Food quality X_{28}: Atmosphere


VARIABLES
FACTOR
WAITING TIME
CLEANLINESS
FRIENDLY PERSONNEL
FOOD TASTE
FOOD TEMPRATURE
FRESHNESS OF FOOD
SERVICE QUALITY
FOOD QUALITY

Customers were asked to rate a fast food restaurant on six characterstics (may be more !) on the basis of the pattern of their responses, these six measures were combined into two 'summary' measures (or FACTORS) : 'service quality' and

'Food Quality'. Marketing researchers use FACTOR ANALYSIS to summarise the information contained in a large number of variables into smaller in a large number of variables into smaller number of 'factors'. The result is that mangers can simplify their decision making as they have to consider only two broad areas, instead of six. The starting point in interpreting factor analysis is factor loadings. (factor loading refers to the correlation between each of the original variables and newly developed factors). Each factor loading is a measure of importance of variable in measuring each factor. Loadings like correlations can varry from -1.0 to 1.0 . The statistical analysis associated with factor analysis would produce factor loading between each factor and each of the original variable.

The next step in FA is to name resulting factors. The researcher examines the variables that have 'high' loadings on each factors.

FACTOR LOADINGS FOR THE TWO FACTORS

	CORRELATION WITH	
VARIABLE	FACTOR-1	FACTOR-2
A_1 (waiting time)	.79	.07
A_2 (cleanliness)	.72	.10
A_3(Friendly personnel)	.72	.05
A_4 (Food taste)	.09	.85
A_5 (Food temperature)	.11	.70
A_6 (Freshness of food)	.04	.74

We have chosen to name this factor 'service quality' because these three variables deal with some aspect of customer's service experience with restaurant. Variable A4 , A5 and A6 all load heavily to factor-2 which we have named 'food quality'. Naming factors is often a 'subjective' process of combining intuition with an inspection of the variables that have high loadings on each factor.

How many factors ?

An important measure to consider in deciding how may factors to retain is the percentage of the variation in the original data that is explained by each factor. (SCREE-PLOT is also used to decide about the same in SPSS out put).

PERCENTAGE OF VARIATION IN ORIGINAL DATA EXPLAINED BY EACH FACTOR

FACTOR	% OF VARIATION EXPLAINED
1	50.3
2	46.5
3	1.8
4	0.8
5	0.6

In this example we would definitely keep the first two factors, because they explain 96.8% of the variability in 5 measures.

Factor analysis (Marketing Research)

- **Advertising:** factor analysis can be used to better understand media habits of various customers
- **Pricing:** FA can help identify the characteristics of price- sensitive and prestige- sensitive customers.
- **Product:** FA can be used to identify brand attributes that influence customer choice.
- **Distribution:** FA can be employed to better understand channel selection crierion among distribution channel numbers.

- **SPSS APPLICATION:**

The SPSS click-through sequence is : ANALYZE - DATA REDUCTION – FACTOR, which leads to a dialog box where you select variables X12 to X21 (say) which are the customers perception of a restaurant.

Click-continue, click – variamax as rotational choice and continue.

- **The SPSS OUTPUT INTERPRETATION:**

  The SPSS OUTPUT for a FACTOR ANALYSIS is shown in the following tables: (the first table is the Rotated component matrix table) lablels for the eleven variables analysed (X12 –X21) are shown in the left column. To the right are four columns of numbers containing the factor loading for the four factors that resulted from the FA of restaurant perceptions. By suppressing loadings under. 30 we observe that only three numbers under column one (component 1 or factor 1), three numbers under column-2 (component 2 or factor 2) and two numbers under column-3 and 4 (components 3 & 4)

ROTATED COMPONENT MATRIX
COMPONENT

Variables	1	2	3	4
X_{18} -Excellent food	.910			
X_{15} – fresh food	.891			
X_{20} – Proper food temp	.862			
X_{12}-Friendly employees		.949		
X_{21}-Speed of service		.941		
X_{19} – knowledgeable employees		.749		
X_{16} – Reasonable prices			.978	
X_{14} – large size portions			.978	
X_{17}-attractive interior				.897
X_{13} – fun place to eat				.834

Extraction method: principal component Analysis Rotation method: varimax with Kaiser Normalisation

TOTAL VARIANCE EXPLAINED
ROTATION SUM OF SQUARED LOADINGS

Component	Total	% of variance	Cummulative %
1	2.493	24.927	24.927
2	2.418	24.184	49.112
3	1.994	19.936	69.048
4	1.663	16.628	85.876

Extraction method: principal component analysis

INTERPRETATION:

For example X18 – excellent food taste has a loading of . 910 on factor -1 and X12 – Friendly employees has a loading of . 949 on factor -2 we prefer a factor solution in which each original variable loads only on one factor (as in this example) but in many cases this does not happen.

Before trying to name the factors we must decide if four factors are enough or if we need more. Our objective is to have as few factors as possible yet account for a reasonable amount of information the number of factors, variables. To determine the number of factors, we look at information in total variance explained table. It shows that 4 factors have accounted for 85.76% of the variance in the original eleven variables. This is a substantial amount of information to account for, and we have Reduced the number of variables by two- thirds (from 11 to 4)

To determine, if the 4 factors are logical , look at the information in the Rotated component matrix. First, examine which original variables come to make a 'new' factor. Factor-1 is made up of X18 – excellent food taste, X15 –fresh food and X20 – proper food temperature. Factor -2 is made up of X12 friendly employees X21 = speed of service and X19 – knowledgeable employees. Factor-3 is made up of X16 – Reasonable prices and X14 – large size portions. Factor- 4 is made up of X17 – attractive interior and X13 – Fun place to eat.

Logic: To analyze the logic of combinations, we look at the variables with highest loadings (largest absolute size). That is why we suppressed loadings less than .30 factor-1 seems to be related to FOOD , factor-2 related to SERVICE , factor-3 realted to VALUE whereas FACTOR-4 related to ATMOSPHERE.

CONCLUSION : (FOR FA)

Thus, we have developed a four-factor solution that accounts for a substantial amount of variance and shows logic in the combinations of original eleven variables. So, with this solution the managers can think about only four variables- food , service, value and atmosphere- while developing their marketing strategies.

STEPS FURTHER:

Using factor analysis with multiple Regression if the dependent variable is 'Customer satisfaction' X22 (say). We can study the significance of Factor-1 to factor-4 , by combining factor analysis with Regression – ' to better' under stand you data.

It is also possible, however, to use other multivariate techniques in combination. For example if the dependent variable is nonmetric, such as gender, then you could use 'Discriminant analysis' . Also, one can use 'CLUSTER ANALYSIS' in combination with Regression or Discriminant analysis.

References:

1. D.F.Morrison: Multivariate statistical methods.

2. Hair, Bush, ortinau: marketing research

3. John T. Pohlmann: Factor Analysis Glossary

4. WWW.Statsoftinc.com / textbook / stfacan.html.

5. Phillip ingram: Factor Analysis a statistics page.

6. Factor Analysis and Related Techniques edited:
   M.S. Lewis -Beck

Thank you !