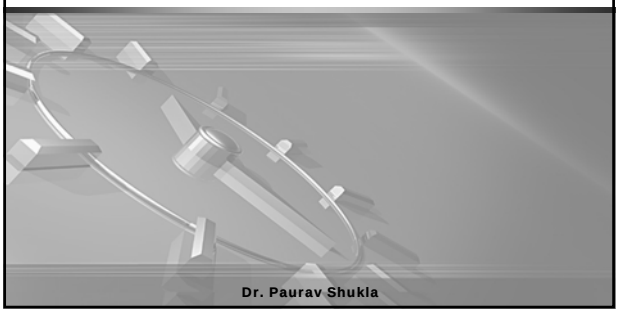
T-test & factor analysis



Dr. Paurav Shukla

Parametric tests

- Better than non parametric tests
  - Stringent assumptions
  - More strings attached
- Assumes population distribution of sample is normal
  - Major problem
  - Alternatives
    - Continue using parametric test if the sample is large or enough evidence available to prove the usage
    - Transforming and manipulating variables
    - Use non parametric test

- Tests used for significant differences between groups
  - Independent sample t-test
  - Paired sample t-test
  - One-way ANOVA (between groups)
  - One-way ANOVA (repeated groups)
  - Two-way ANOVA (between groups)
  - MANOVA

t-test

- Used when they are two groups only
- Compares the mean score on some continuous variable
- Largely used in measuring before – after effect
  - Also called intervention effect
  - Option 1: Same sample over a period of time
  - Option 2: Two different groups at a single point of time

Assumptions associated with t-test

- Population distribution of sample is normal
  - Probability (random) sampling
- Level of measurement
  - Dependent variable is continuous
- Observed or measured data must be independent
  - If interaction is unavoidable use a stringent alpha value ($p<.01$)
- Homogeneity of variance
  - Assumes that samples are obtained from populations of equal variance
  - ANOVA is reasonably robust against this

Why t-test is important?

- Highly used technique for hypothesis testing
- Can lead to wrong conclusions
  - Type 1 error
    - Reject null hypothesis instead of accepting
      - When we assume there is a difference between groups, when it is not
  - Type 2 error
    - Accept null hypothesis instead of rejecting
      - When we assume there is no difference between groups, when there is
  - Solution
    - Interdependency between both errors
    - Selection of alpha level

Factors influencing power of t-test

- Sample size
- Strength of interdependency between dependent and independent variable (Strength of Association or Effect Size)
- Alpha level

Procedure for independent sample t-test

- Analyze
  - Compare means
    - Independent sample t-test

Interpreting independent samples t-test

- Group statistics
  - Look for N (missing values)
- Independent samples test
  - Levene's test
    - If sig value is higher than 0.05 use equal variance assumed
    - If sig value is lower than 0.05 use equal variances not assumed
- Assessing difference between groups
  - If the sig (2-tailed) is equal or less than 0.05
    - There is a significant difference in the mean scores
  - If the sig (2-tailed) is great than 0.05
    - There is no significant difference between the two groups

Calculating the effect size

- SPSS does not calculate Eta squared to measure effect size for t-test
- Calculation

$$\text{Eta squared} = \frac{t^2}{t^2 + (N1 + N2 - 2)}$$

- Interpretation values
  - 0.01 = Small effect
  - 0.06 = Moderate effect
  - 0.14 = Large effect

Paired sample t-test

- Analyze
  - Compare means
    - Paired samples t-test

Interpreting paired sample t-test

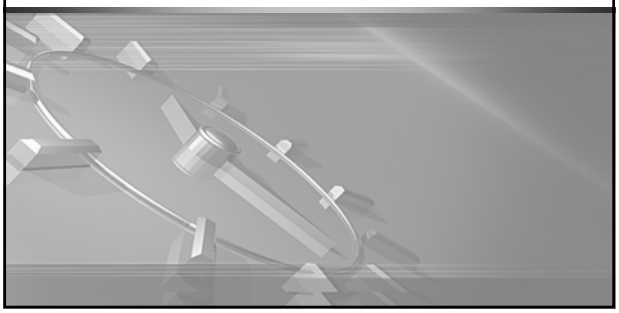
- Determining overall significance
  - If the sig (2-tailed) is less than 0.05
    - Significant difference between two scores
  - If the sig (2-tailed) is higher than 0.05
    - No significant difference between two scores
- Comparing mean values
  - Mean values
- Effect size

$$\text{Eta squared} = \frac{t^2}{t^2 + N - 1}$$

Non-parametric alternatives

- Independent sample t-test
  - Mann-Whitney U test
    - Instead of comparing mean it compares medians
  - Procedure
    - Analyze
    - Non-parametric tests
    - 2 Independent samples
- Paired sample t-test
  - Wilcoxon signed rank test
  - Procedure
    - Analyze
    - Non-parametric tests
    - 2 Related samples

Factor Analysis



Session overview

- Basic Concept
- Factor Analysis Model
- Types of factor analysis
- Statistics Associated with Factor Analysis
- Conducting factor analysis
- Applications of factor analysis

Basic concept

- A data reduction technique designed to represent a wide range of attributes on a smaller number of dimensions.
- Factor analysis is an interdependence technique
  - in that an entire set of interdependent relationships is examined without making the distinction between dependent and independent variables.
- Factor analysis is used in the following circumstances:
  - To identify underlying dimensions, or factors, that explain the correlations among a set of variables.
  - To identify a new, smaller, set of uncorrelated variables to replace the original set of correlated variables in subsequent multivariate analysis (regression or discriminant analysis).
  - To identify a smaller set of salient variables from a larger set for use in subsequent multivariate analysis.

Standard Model

- Mathematically, each variable is expressed as a linear combination of underlying factors. The covariation among the variables is described in terms of a small number of common factors plus a unique factor for each variable. If the variables are standardized, the factor model may be represented as:

$$Xi = Ai\,1F1 + Ai\,2F2 + Ai\,3F3 + \ldots + AimFm + ViUi$$

where

Xi	=	i th standardized variable
Aij	=	standardized multiple regression coefficient of variable i on common factor j
F	=	common factor
Vi	=	standardized regression coefficient of variable i on unique factor i
Ui	=	the unique factor for variable i
m	=	number of common factors

Types of factor analysis

- Exploratory factor analysis
  - Largely known as Principal Components Analysis (PCA)
- Confirmatory factor analysis
  - Structural Equation Modelling

Exploratory Factor Analysis

- A company producing single malt scotch whisky has asked you to study the status consumption behaviour within a specific group of people. Through literature you have developed three constructs which influence consumer choice of such status consumption brand.
- The three constructs are:
  - Psychological association scale
  - Brand association scale
  - Situations/events scale

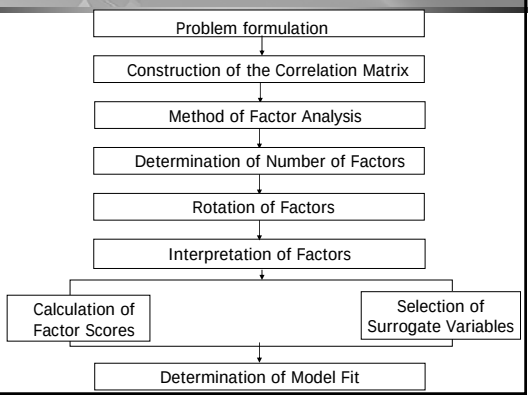
Exploratory Factor Analysis

- Psychological associations scale
  - I buy status alcoholic beverages:
    - for gaining respect from others
    - to be a popular person in groups.
    - to show who I am.
    - to be a symbol of success.
    - to be a symbol of prestige.
    - to indicate my wealth.
    - to be noticed by others.
    - to indicate my achievement.
    - to be appreciated by others.
    - because status interests me.
    - because status is important to me.
    - because status enhances my image.
    - because I am interested in new alcoholic beverages with status.
    - because I would pay more for an alcoholic beverage if it had status.
    - because the status of an alcoholic beverage is irrelevant to me.
    - to increase my own value from others' point of view.
    - to be more attractive than others.

The question

- How to know what factors are important to consumers?
- What variables are grouped together in consumer mind?
- There being no dependent variable how can I measure the impact?

Conducting Factor Analysis



Statistics Associated with Factor Analysis

- **Bartlett's test of sphericity**
  - Bartlett's test of sphericity is a test statistic used to examine the hypothesis that the variables are uncorrelated in the population. In other words, the population correlation matrix is an identity matrix; each variable correlates perfectly with itself ($r = 1$) but has no correlation with the other variables ($r = 0$).
- **Correlation matrix**
  - A correlation matrix is a lower triangle matrix showing the simple correlations, *r*, between all possible pairs of variables included in the analysis. The diagonal elements, which are all 1, are usually omitted.
- **Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy**
  - The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is an index used to examine the appropriateness of factor analysis. High values (between 0.5 and 1.0) indicate factor analysis is appropriate. Values below 0.5 imply that factor analysis may not be appropriate.
- **Percentage of variance**
  - The percentage of the total variance attributed to each factor.
- **Residuals**
  - are the differences between the observed correlations, as given in the input correlation matrix, and the reproduced correlations, as estimated from the factor matrix.

Statistics Associated with Factor Analysis

- **Communality**
  - Communality is the amount of variance a variable shares with all the other variables being considered. This is also the proportion of variance explained by the common factors.
- **Eigenvalue**
  - The eigenvalue represents the total variance explained by each factor.
- **Factor loadings**
  - Factor loadings are simple correlations between the variables and the factors.
- **Factor loading plot**
  - A factor loading plot is a plot of the original variables using the factor loadings as coordinates.
- **Factor matrix**
  - A factor matrix contains the factor loadings of all the variables on all the factors extracted.
- **Scree plot**
  - A scree plot is a plot of the Eigenvalues against the number of factors in order of extraction.
- **Factor scores**
  - Factor scores are composite scores estimated for each respondent on the derived factors.

Assumptions for factor analysis

- Sample size
  - 1 to 10 ratio
- Variable correlation higher than 0.3 in most cases
- Linearity
- Outliers

Procedure

- Analyze
  - Data reduction
    - Factor
  - Descriptives
    - Correlation matrix
      - Coefficients
      - KMO and Bartlett's test of sphericity
    - Statistics
      - Initial solution
  - Options
    - Missing values
      - Exclude cases pairwise
    - Coefficients display format
      - Sorted by size
      - Suppress absolute value less than 0.3
- Extraction
  - Method
    - Principal components
  - Analyse
    - Correlations matrix
  - Display
    - Screeplot
    - Unrotated factor solution
  - Extract
    - Eigenvalues over 1
- Rotation
  - Method
    - Varimax

Interpretations

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.888
Bartlett's Test of Sphericity	Approx. Chi-Square	2750.261
	df	136
	Sig.	.000

KMO value 0.6 and above

Sig. Value less than 0.05

		Gain respect	Popularity	Show who I am	Symbol of success	Symbol of prestige	No o
Correlation	Gain respect	1.000	.379	.464	.524	.465	
	Popularity	.379	1.000	.598	.102	.009	
	Show who I am	.464	.598	1.000	.255	.102	
	Symbol of success	.524	.102	.255	1.000	.749	
	Symbol of prestige	.465	.009	.102	.749	1.000	
	Notice by others	.421	.543	.647	.247	.209	
	Indicate achievement	.598	.294	.411	.582	.503	
	Presence of others	.440	.538	.650	.165	.143	
	Interest in status	.371	.206	.228	.442	.499	
	Status enhance my image	.394	.310	.369	.376	.368	
	Willing to try new drink	.200	.338	.369	.173	.124	
	Pay more for status	.374	.397	.410	.293	.240	
	More value	.456	.604	.594	.149	.044	
	Attractiveness	.325	.631	.506	.105	.098	

Most correlations of 0.3 and above

Interpretations

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	6.815	40.087	40.087	6.815	40.087	40.087	4.741	27.887	27.887
2	2.915	17.145	57.231	2.915	17.145	57.231	3.440	20.237	48.124
3	1.328	7.812	65.043	1.328	7.812	65.043	2.876	16.920	65.043
4	.965	5.678	70.722						
5	.732	4.308	75.030						
6	.568	3.343	78.372						
7	.555	3.264	81.637						
8	.485	2.851	84.487						
9	.421	2.476	86.964						
10	.391	2.298	89.261						
11	.373	2.196	91.457						
12	.345	2.029	93.487						
13	.296	1.742	95.229						
14	.239	1.407	96.636						
15	.209	1.230	97.866						
16	.192	1.130	98.996						
17	.171	1.004	100.000						

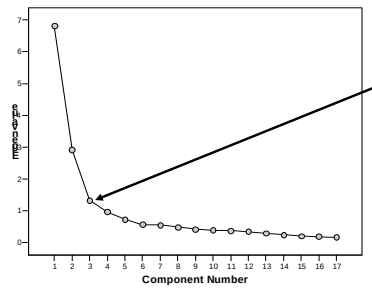
Extraction Method: Principal Component Analysis.

Cumulative percent of variance explained.

We are looking for an eigenvalue above 1.0.

Interpretations

Scree Plot



Clear bend in Scree plot

Interpretations

Component Matrix[a]

	Component		
	1	2	3
Indicate achievement	.733		
Status enhance my image	.723		-.428
Notice by others	.715	-.326	
Presence of others	.701	-.379	
More value	.699	-.473	
Gain respect	.688		.399
Pay more for status	.688		
Show who I am	.677	-.425	
Status is important to me	.646	.450	-.412
Interest in status	.643	.393	-.376
Attractiveness	.630	-.442	
Popularity	.609	-.493	
Willing to try new drink	.587		-.467
Status not related to me	-.407		
Indicate wealth	.388	.718	
Symbol of prestige	.518	.695	
Symbol of success	.573	.577	

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Most components loading on factor 1
Less components loading on factor 2
Least components loading on the last factor

Interpretations

Rotated Component Matrix[a]

	Component		
	1	2	3
Show who I am	.820		
More value	.795		
Presence of others	.794		
Notice by others	.779		
Popularity	.774		
Attractiveness	.724		
Status not related to me	-.353		
Symbol of prestige		.871	
Symbol of success		.843	
Indicate wealth		.792	
Indicate achievement	.457	.643	
Gain respect	.515	.618	
Status is important to me		.441	.771
Status enhance my image			.764
Interest in status		.415	.727
Willing to try new drink	.372		.664
Pay more for status	.462		.538

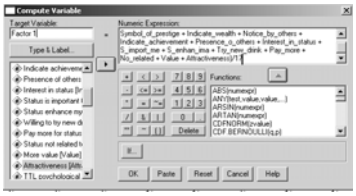
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 6 iterations.

When a variable is loaded on two or more factors include it with the factor where it has highest value

What shall the component be called?

Show who I am More value Presence of others Notice by others Popularity Attractiveness Status not related to me	Symbol of prestige Symbol of success Indicate wealth Indicate achievement Gain respect	Status is important to me Status enhance my image Interest in status Willing to try new drink Pay more for status

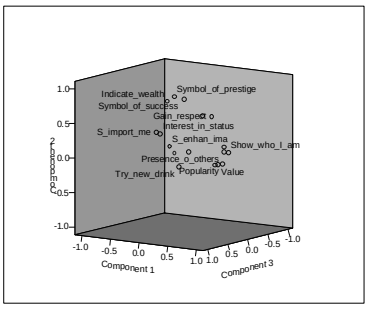
Calculating component scores



Helps in brand level comparisons

Interpretations

Component Plot in Rotated Space



Because it's a 3 factor solution we are getting a 3 dimensional plot.

What if we get a 2 factor solution?

How to get this represented in 2 dimension plot so it is easier to understand?

Tricks of the trade

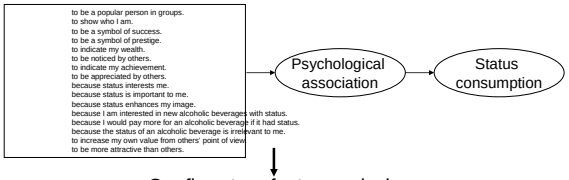
- Self-determining number of factors
- Combining reliability analysis (Cronbach's Alpha)

Applications of factor analysis

- A more concise representation of the marketing situation and hence communication may be enhanced
- Fewer questions may be required on future surveys
- Market segmentation and identifying the underlying variables on which to group customers
- Identifying brand attributes that influence customer choice
- Identifying media consumption habits of consumers
- Perceptual maps become feasible

Further questions

- What if we want to measure a hypothesis or theories concerning the structure underlying a set of variables?
- Such as, in our case do all these variables affect the psychological association of a brand and which in turn would affect status consumption?



Confirmatory factor analysis
(Structural Equation Modelling)

Confirmatory factor analysis

- Robust but complex
- Used by a lot of advance level researchers
- The 'in' phenomenon in research
- Software: Lisrel or Amos (SPSS) and several others