Conrad

School of Entrepreneurship and Business

Agenda

01

Recap

02

Factor Analysis

03

Hands-on Exercise

https://perceptions.uwaterloo.ca/

Logistic Regression

Form for Predicted Probabilities

$$\ln\left(\frac{\mathrm{P(Y)}}{1-\mathrm{P(Y)}}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_\mathrm{K} X_\mathrm{K}$$

$$\Updownarrow$$

$$\mathrm{P(Y)} = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_\mathrm{K} X_\mathrm{K})}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_\mathrm{K} X_\mathrm{K})}$$

In this latter form, the logistic regression model directly relates the probability of Y to the predictor variables.

Multicollinearity in Multiple Regression

Multicollinearity

- The term multicollinearity refers to the correlation among the independent variables. Causation is not implied by multicollinearity.

- When the independent variables are highly correlated (say, $|r| > .8$), it is not possible to determine the separate effect of any independent variable on the dependent variable.

- Since multicollinearity is almost always present, it is a problem of degree, not merely existence.

- In extreme case of multicollinearity, signs of coefficients may be contrary to those expected.

Multicollinearity: Remedies

- Do nothing – live with what you have!
- Increase sample size
- Omit Variables
- Factor Analysis

Factor Analysis
(Principal Component Analysis)

Factor Analysis

- **Factor analysis** is a class of procedures used for data reduction and summarization.
- It is an **interdependence technique:** no distinction between dependent and independent variables.
- Factor analysis is used:
  - 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

Factor Analysis Model

The common factors themselves can be expressed as linear combinations of the observed variables.

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \ldots + W_{ik}X_k$$

Where:

F_i = estimate of *i* th factor

W_i= weight or factor score coefficient

k = number of variables

- The first set of weights (factor score coefficients) are chosen so that the first factor explains the largest portion of the total variance.
- Then a second set of weights can be selected, so that the second factor explains most of the residual variance, subject to being uncorrelated with the first factor.
- This same principle applies for selecting additional weights for the additional factors.

Statistics Associated with Factor Analysis

Correlation matrix. A correlation matrix is a matrix showing the simple correlations, *r*, between all possible pairs of variables included in the analysis.

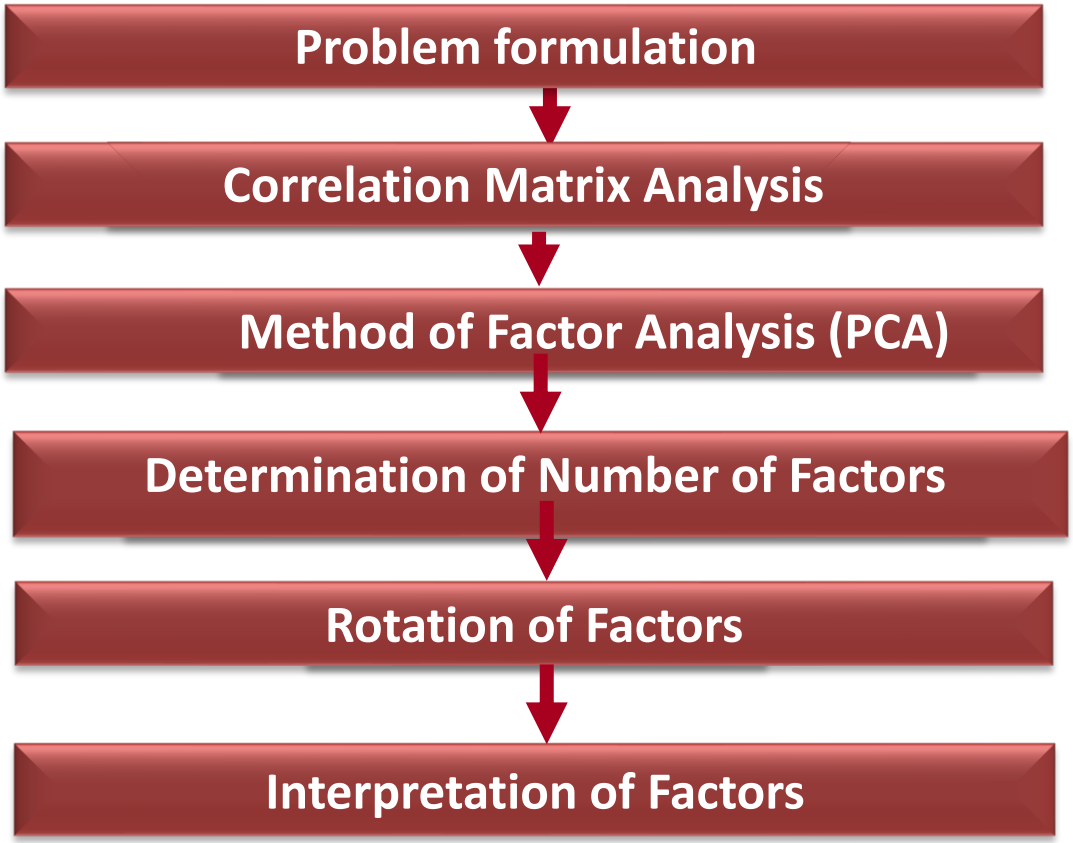
- **Bartlett's test of sphericity**. Bartlett's test of sphericity is used to test the hypothesis that the variables are uncorrelated in the population. If the Bartlett's test of sphericity is rejected ($p<0.01$), then factor analysis is appropriate.

- **Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy**. Used to examine the appropriateness of factor analysis. High values (between 0.5 and 1.0) indicate appropriateness. Values below 0.5 imply not.

Statistics Associated with Factor Analysis

- **Eigenvalue.** Represents the total variance explained by each factor
- **Factor loadings.** Correlations between the variables and the factors.
- **Factor matrix.** A factor matrix contains the factor loadings of all the variables on all the factors
- **Factor scores.** Factor scores are composite scores estimated for each respondent on the derived factors.
- **Percentage of variance.** The percentage of the total variance attributed to each factor.
- **Scree plot.** A scree plot is a plot of the Eigenvalues against the number of factors in order of extraction.

Conducting Factor Analysis



Formulate the Problem

- The objectives of factor analysis should be identified.
- The variables to be included in the factor analysis should be specified. The variables should be measured on an interval or ratio scale.
- An appropriate sample size should be used. As a rough guideline, there should be at least four or five times as many observations (sample size) as there are variables.

Correlation Matrix Analysis

- The analytical process is based on a matrix of correlations between the variables.

- If the **Bartlett's test of sphericity** is rejected, then factor analysis is appropriate.

- If the **Kaiser-Meyer-Olkin (KMO)** measure of sampling adequacy is small, then the correlations between pairs of variables cannot be explained by other variables and factor analysis may not be appropriate.

Determine the Number of Factors

- **A Priori Determination.** Use prior knowledge.

- **Determination Based on Eigenvalues.** Only factors with Eigenvalues greater than 1.0 are retained.

- **Determination Based on Scree Plot.** A scree plot is a plot of the Eigenvalues against the number of factors in order of extraction. The point at which the scree begins denotes the "true" number of factors.

- **Determination Based on Percentage of Variance** – benchmark is minimum 60%.

Rotation of Factors

- Through rotation the factor matrix is transformed into a simpler one that is easier to interpret.
- After rotation each factor should have non-zero, or significant, loadings for only some of the variables. Each variable should have non-zero or significant loadings with only a few factors, if possible, with only one.

- **Varimax procedure.** Axes maintained at right angles

  -Most common method for rotation.
  -An orthogonal method of rotation that minimizes the number of variables with high loadings on a factor.
  -Orthogonal rotation results in uncorrelated factors.

- **Oblique rotation.** Axes not maintained at right angles

  -Factors are correlated.
  -Oblique rotation should be used when factors in the population are likely to be strongly correlated.

Interpret Factors

- A factor can be interpreted in terms of the variables that load high on it
- Look for factor loadings of greater than 0.5
- Look at the highest factor loadings to give interpretation to the factors
- If a variable loads on two factors (and loadings are greater than 0.5), factor with the higher loading beats the tie

Performing the Analysis Using SPSS

Let's revert to the brand tracking questionnaire we have used earlier in the class.

Problem formulation

*Is there a simpler set of factors that explains the data captured by **brand imagery statements for Amazon***

Brand Imagery – Competitive Brand #1

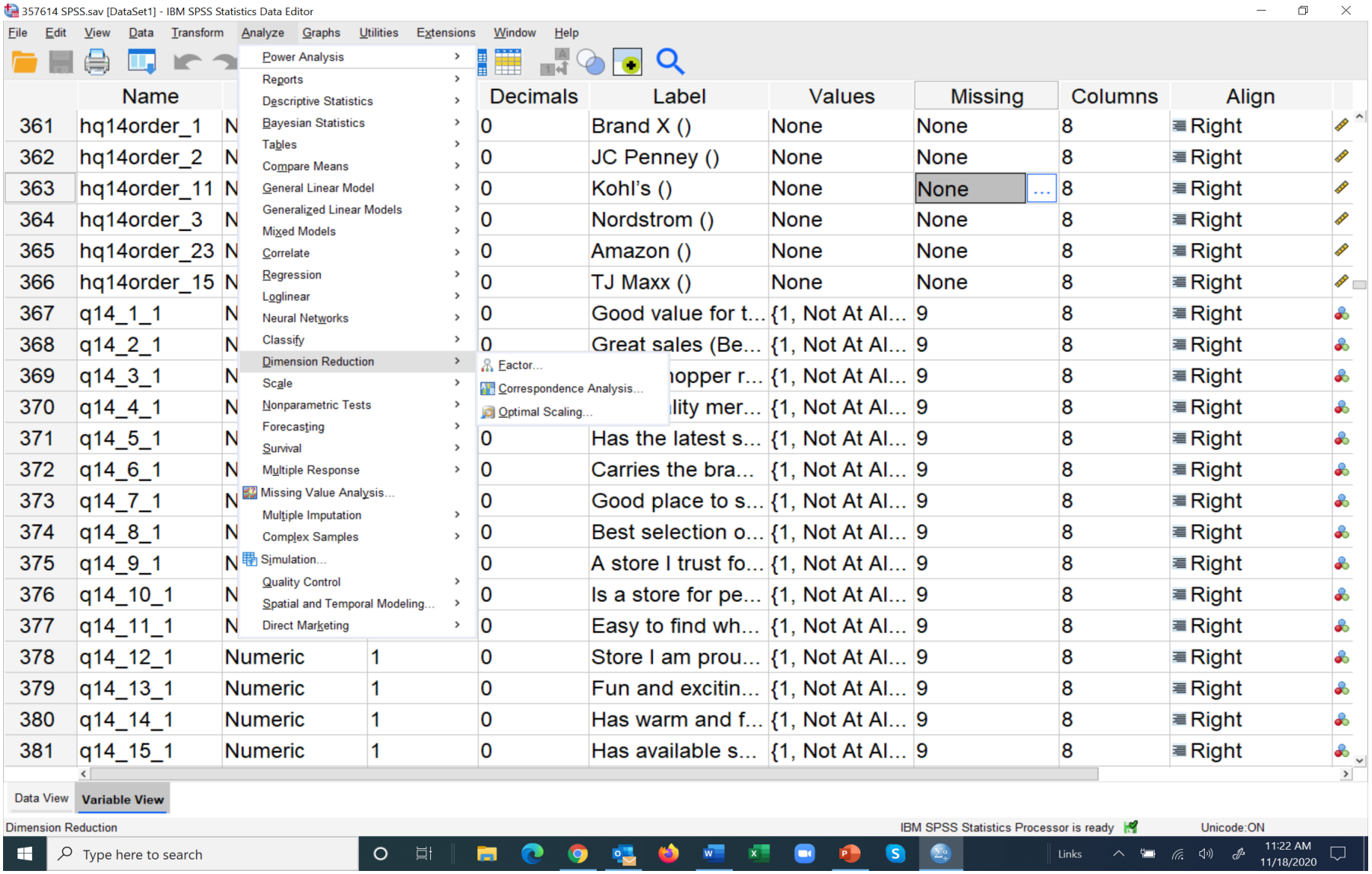
Q15 Below is a list of different statements that people have made about retail stores. Please read each statement and rate (PIPE BRAND) on a scale of 1 to 5, where 5 means "Describes Extremely Well," and 1 means "Does Not Describe At All."

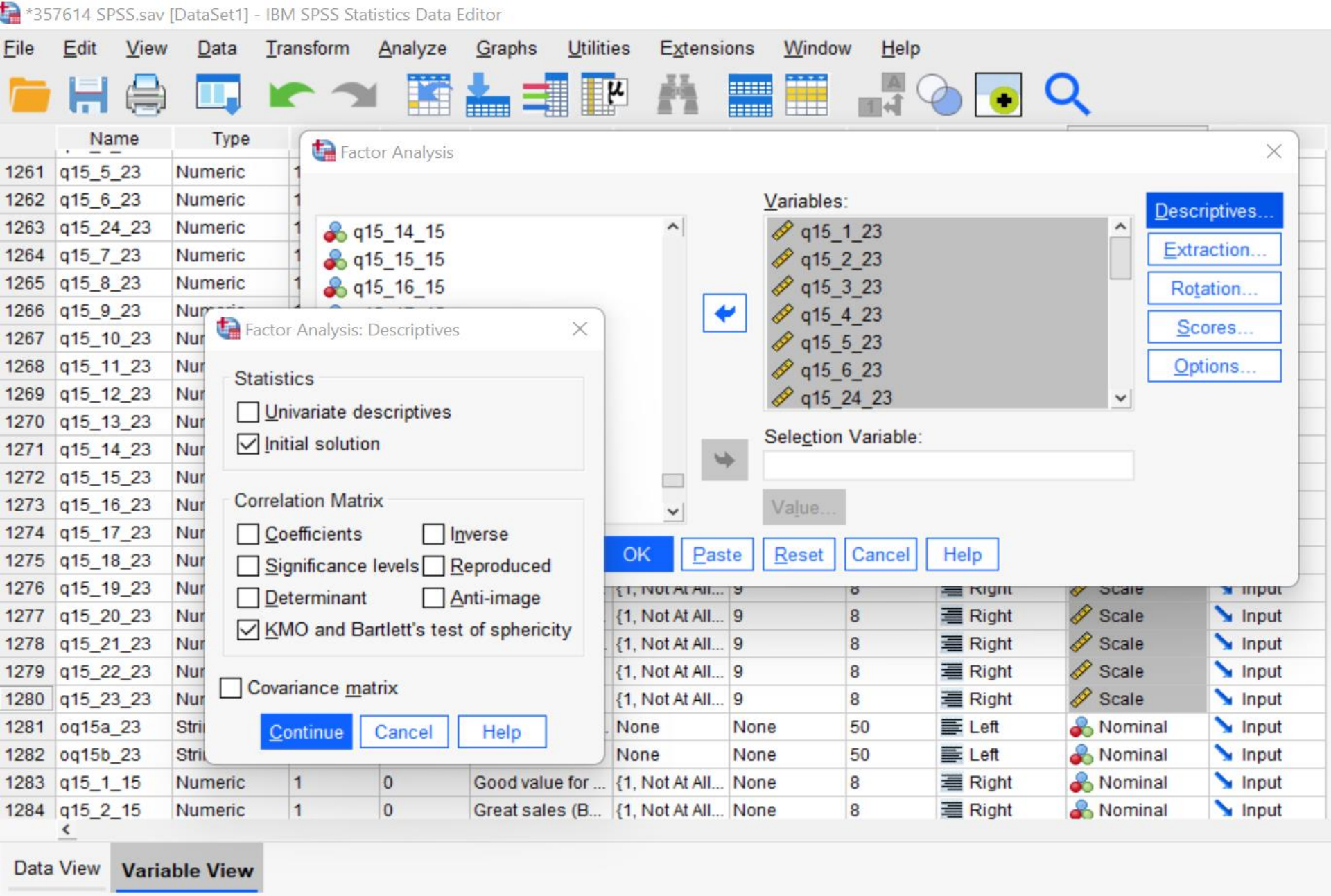
Add N/A Column ← **Add this to the missing data**

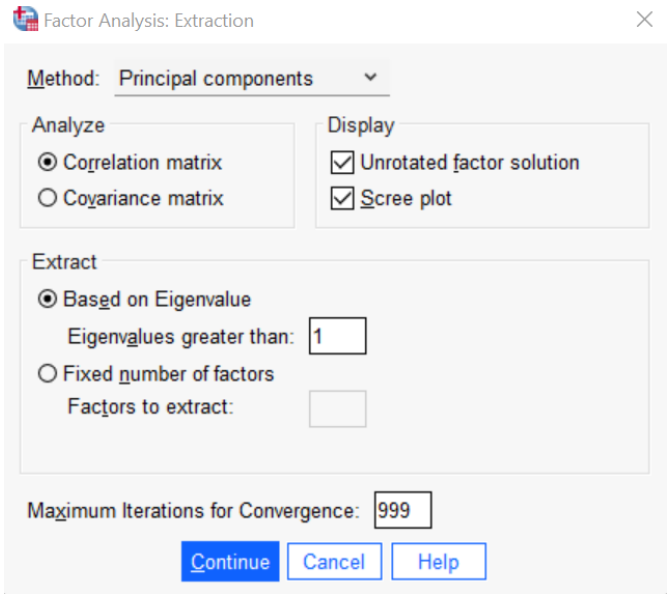
	not at all 1	not very well 2	somewhat 3	very well 4	Extremely Well 5
Value Attributes					
Good value for the money					
Great sales					
Good shopper rewards program					
Product Attributes					
Top quality merchandise					
Has the latest styles & fashions					
Carries the brands I want					
Has brands/products I want, but can't find anywhere else					
Good place to shop for gifts					
Best selection of brands and designers					
A store I trust for important occasions					
Store Experience Attributes					
Is a store for people like me					
Easy to find what I'm looking for					
Store I am proud to shop at					
Fun and exciting shopping experience					

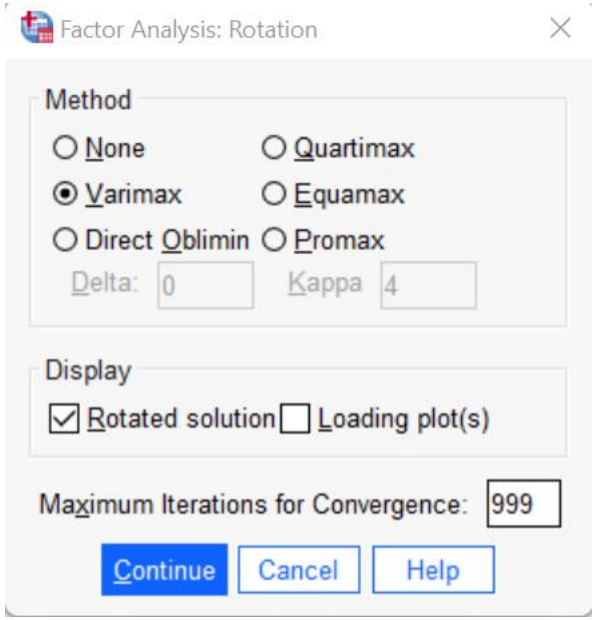
Service Attributes					
Has warm and friendly salespeople					
Has available salespeople					
Makes it easy to return items					
Brand Attributes					
I often discover new things to buy					
Helps express my personal style					
Inspires me to try new things					
Has merchandise worth paying more for					
Omnichannel					
Offers a consistent experience online and in store					
Offers convenient ways to shop in stores and online					
Innovative in making shopping easier					

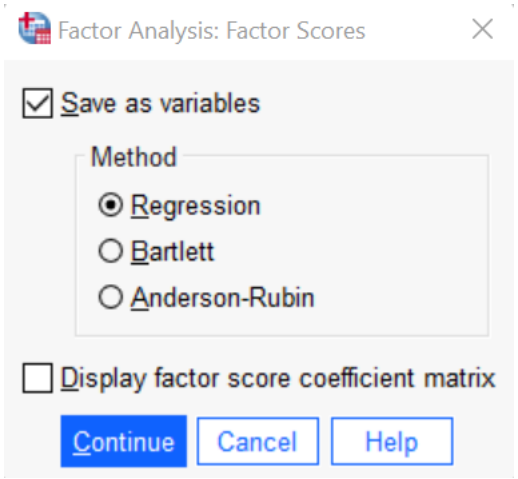
SPSS *Analyze | Dimension Reduction| Factor*

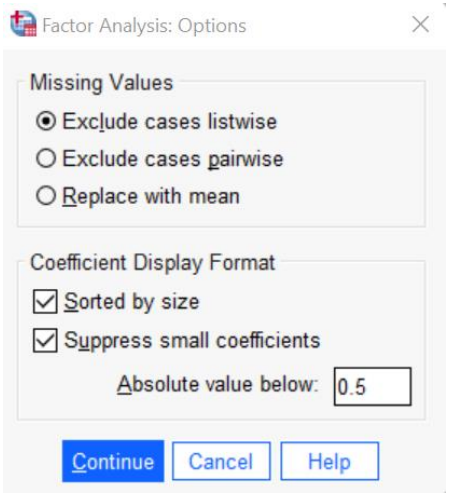



*357614 SPSS.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Factor Analysis
Variables:
q15_1_23
q15_2_23
q15_3_23
q15_4_23
q15_5_23
q15_6_23
q15_24_23
q15_14_15
q15_15_15
q15_16_15
Descriptives...
Extraction...
Rotation...
Scores...
Options...
Selection Variable:
Value...
OK Paste Reset Cancel Help
Factor Analysis: Descriptives
Statistics
Univariate descriptives
Initial solution
Correlation Matrix
Coefficients
Inverse
Significance levels
Reproduced
Determinant
Anti-image
KMO and Bartlett's test of sphericity
Covariance matrix
Continue Cancel Help
Data View
Variable View


Factor Analysis: Extraction
Method: Principal components
Analyze
Correlation matrix
Covariance matrix
Display
Unrotated factor solution
Scree plot
Extract
Based on Eigenvalue
Eigenvalues greater than: 1
Fixed number of factors
Factors to extract:
Maximum Iterations for Convergence: 999
Continue
Cancel
Help


Factor Analysis: Rotation
Method
None
Quartimax
Varimax
Equamax
Direct Oblimin
Promax
Delta: 0
Kappa 4
Display
Rotated solution
Loading plot(s)
Maximum Iterations for Convergence: 999
Continue
Cancel
Help


Factor Analysis: Factor Scores
Save as variables
Method
Regression
Bartlett
Anderson-Rubin
Display factor score coefficient matrix
Continue
Cancel
Help


Factor Analysis: Options
Missing Values
Exclude cases listwise
Exclude cases pairwise
Replace with mean
Coefficient Display Format
Sorted by size
Suppress small coefficients
Absolute value below: 0.5
Continue
Cancel
Help

Correlation Matrix Analysis-
Is it appropriate to proceed with FA?

Problem formulation

↓

Correlation Matrix Analysis

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.947
Bartlett's Test of Sphericity	Approx. Chi-Square	5005.806
	df	276
	Sig.	.000

If the Bartlett's test of sphericity is rejected, then factor analysis is appropriate.

KMO- High values (between 0.5 and 1.0) indicate appropriateness. Values below 0.5 imply not.

Problem formulation

Correlation Matrix Analysis

Method of Factor Analysis

Determination of Number of Factors

Number of Factors

Everything points to 3 factors – but worth exploring 4 factors as well

Eigenvalues > 1

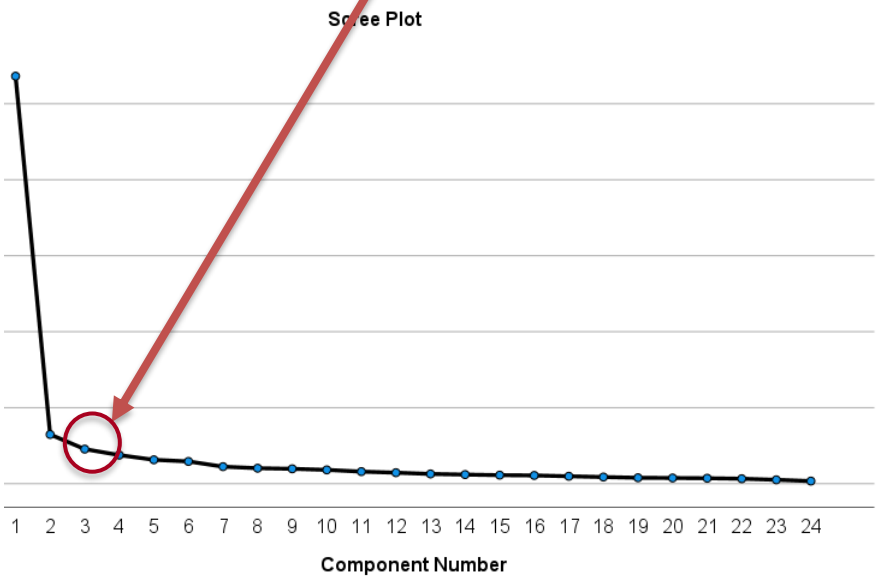
Percentage of Variance Explained

Total Variance Explained

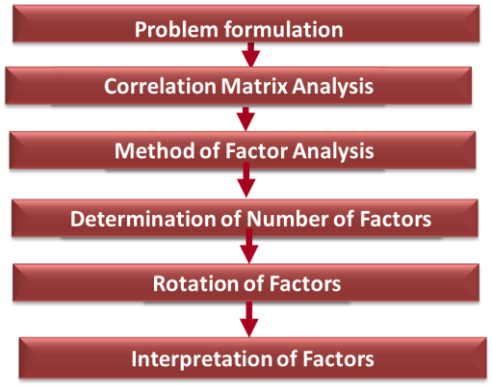
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	13.409	55.871	55.871	13.409	55.871	55.871	6.290	26.209	26.209
2	1.621	6.753	62.623	1.621	6.753	62.623	5.471	22.798	49.006
3	1.139	4.744	67.367	1.139	4.744	67.367	4.407	18.361	67.367
4	.942	3.925	71.292						
5	.787	3.280	74.571						
6	.733	3.055	77.627						
7	.562	2.340	79.967						
8	.511	2.127	82.094						
9	.491	2.044	84.138						
10	.456	1.898	86.036						
11	.399	1.661	87.696						
12	.363	1.513	89.210						
13	.321	1.339	90.549						
14	.301	1.253	91.802						
15	.283	1.177	92.979						
16	.272	1.135	94.114						
17	.245	1.023	95.137						
18	.220	.918	96.055						
19	.196	.818	96.873						
20	.189	.788	97.660						
21	.181	.752	98.412						
22	.165	.687	99.099						
23	.129	.537	99.636						
24	.087	.364	100.000						

Extraction Method: Principal Component Analysis.

Based on Scree Plot



Factor Rotation & Interpretation



Look for factor loadings of greater than 0.5

Look at the highest factor ladings to give interpretation to the factors

If a variable loads on two factors (and loadings are greater than 0.5), factor with the higher loading beats the tie

Rotated Component Matrix[a]

	Component 1	2	3
Carries the brands I want	.778		
Best selection of brands and designers	.773		
Has the latest styles & fashions	.725		
Top quality merchandise	.660		
Has brands/products I want, but can't find anywhere else	.658		
A store I trust for important occasions	.635		
Good value for the money	.567		
Great sales	.565		
Offers a consistent experience online and in store		.734	
Innovative in making shopping easier		.688	
Is a store for people like me	.504	.664	
Offers convenient ways to shop in stores and online		.645	
Good place to shop for gifts		.624	
Fun and exciting shopping experience		.620	
I often discover new things to buy		.614	
Easy to find what I'm looking for	.567	.602	
Store I am proud to shop at		.588	
Makes it easy to return items		.530	
Inspires me to try new things		.502	
Has warm and friendly salespeople			.870
Has available salespeople			.865
Good shopper rewards program			.695
Has merchandise worth paying more for			.529
Helps express my personal style			

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 7 iterations.

Use Factor Scores for your Analyses

What is impact of the various **brand imagery perceptions** on **Brand Commitment** for Amazon?

Brand Imagery – Competitive Brand #1

Q15 Below is a list of different statements that people have made about retail stores. Please read each statement and rate (PIPE BRAND) on a scale of 1 to 5, where 5 means "Describes Extremely Well," and 1 means "Does Not Describe At All."

Add N/A Column

	not at all 1	not very well 2	somewhat 3	very well 4	Extremely Well 5
Value Attributes					
Good value for the money					
Great sales					
Good shopper rewards program					
Product Attributes					
Top quality merchandise					
Has the latest styles & fashions					
Carries the brands I want					
Has brands/products I want, but can't find anywhere else					
Good place to shop for gifts					
Best selection of brands and designers					
A store I trust for important occasions					
Store Experience Attributes					
Is a store for people like me					
Easy to find what I'm looking for					
Store I am proud to shop at					
Fun and exciting shopping experience					
Service Attributes					
Has warm and friendly salespeople					
Has available salespeople					
Makes it easy to return items					
Brand Attributes					
I often discover new things to buy					
Helps express my personal style					
Inspires me to try new things					
Has merchandise worth paying more for					
Omnichannel					
Offers a consistent experience online and in store					
Offers convenient ways to shop in stores and online					
Innovative in making shopping easier					

Brand Commitment

Q11 **Imagine you had to shop at a retail store, which of these statements best describes how much you would consider shopping at each of these stores? (Please select one answer for each brand)**

1 PROGRAMMER: ALLOW ONLY ONE ANSWER PER BRAND.

ACCEPT ONLY ONE ANSWER IN FIRST ROW "ONLY STORE WOULD CONSIDER"

	Brand X	~~JC~~ ~~Penney~~Target	Kohl's	Nordstrom	Amazon	TJ Maxx
Favorite store; only one I consider	○	○	○	○	○	○
Store I prefer and consider highly	○	○	○	○	○	○
Store I consider equally with others	○	○	○	○	○	○
Store I might consider, less so than others	○	○	○	○	○	○
Not a store I usually consider	○	○	○	○	○	○
Store I would never consider	○	○	○	○	○	○

Use Factor Scores for your Analyses

What is impact of the various **brand imagery perceptions** on **Brand Commitment** for Amazon?

FA Model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.428[a]	.183	.164	.815

a. Predictors: (Constant), Service, Experience, Product

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	19.059	3	6.353	9.566	.000[b]
	Residual	85.002	128	.664		
	Total	104.061	131			

a. Dependent Variable: Amazon

b. Predictors: (Constant), Service, Experience, Product

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2.230	.071		31.262	.000
	Product	-.292	.073	-.321	-3.991	.000
	Experience	-.288	.073	-.318	-3.952	.000
	Service	-.028	.074	-.031	-.383	.702

a. Dependent Variable: Amazon

?

All Variables

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.575[a]	.330	.180	.807

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34.390	24	1.433	2.201	.003[b]
	Residual	69.671	107	.651		
	Total	104.061	131			

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	4.555	.495		9.211	.000
	Has brands/products I want, but can't find anywhere else	-.285	.137	-.291	-2.082	.040

Hands-on Analysis

- **All questions below refer to the brand tracking questionnaire and data made available on LEARN**
  1. What is impact of Brand Commitment, Likelihood to Purchase and Brand Salience on being a promoter for TJ Maxx?
  2. What is the probability of someone being a promoter for TJ Maxx who is likely to make a purchase within the next 1 month, thinks that TJ Maxx as a brand is on its way up and considers TJ Maxx their favourite store?
  3. Is there a simpler set of factors that explains the data captured by brand imagery perception statements for Nordstrom ?
  4. Should Nordstrom be focusing on the brand imagery factors above or Brand Commitment and Likelihood to Purchase in order to increase the Likelihood of Recommendation?

TILL NEXT TIME