Quantitative Business Research Methods – Day 4

Spring 2023

Sami Kajalo

Aalto University
School of Business

Data Analysis with SPSS Software: Multivariate Analysis Methods

Aalto University
School of Business

Multivariate analysis methods

Two general types of multivariate analysis methods:

- Analysis of dependence:
  - One (or more) variables are dependent variables, which are explained or predicted by others.
  - E.g. Multiple regression and structural equation modeling (SEM).
- Analysis of interdependence:
  - No variables thought of as “dependent”.
  - Focus on the relationships among variables, or objects (e.g. consumers).
  - E.g. factor analysis (focus on variables) and cluster analysis (focus on objects e.g. consumers).

Strategy for using these methods in a Master's thesis

1. Read the chapter about the method from: Malhotra, N.K., Birks, D.F. and Nunan, D. (2017). Marketing Research: An Applied Approach, 5. Edition.
2. Find one master's thesis where the method has been applied. Do not trust the master's thesis blindly.
3. Find one journal article where the method has been applied.
4. Read the chapter about the method and follow step-by-step advice from: Hair, Joseph F., William C. Black, Barry J. Babin, and Rolph E. Anderson (2014). Multivariate data analysis, Pearson New International Edition (7th ed.).
5. Explain the method in your master's thesis using these two textbooks as references. Report the results using master's thesis and journal article as examples.

Analysis of interdepence with SPSS

- Most common methods in Master’s thesis:

1. Factor Analysis and Cronbach’s Alpha statistic
2. Cluster Analysis

Example material for lecture

- Aino Kymäläinen. Exploring motivations to engage in collaborative consumption - Case: Facebook recycling groups.
- Frösén, J., Tikkanen, H., Jaakkola, M. and Vassinen, A. (2013), "Marketing performance assessment systems and the business context", *European Journal of Marketing*, Vol. 47 No. 5, pp. 715–737.

SPSS: Factor Analysis

Factor analysis

- Factor analysis is a general name denoting a class of procedures primarily used for data reduction and summarisation.
- 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.

Source: Nunan et al. (2020): Marketing Research - Applied Insight. Sixth Edition.

Factor analysis: Example

First column: factor loadings

Second column: communalities

Factor	Metrics	Factor loading	Communality	Cronbach's alpha
F1: Brand equity Actual and potential customer attitudes, perceptions, thoughts, and feelings	Awareness	0.70	0.58	0.90
	Salience	0.68	0.53	
	Perceived quality/esteem	0.68	0.61	
	Consumer satisfaction	0.63	0.52	
	Relevance to consumer	0.67	0.58	
	Image/personality/identity	0.69	0.58	
	(Perceived) differentiation	0.62	0.56	
	Commitment/purchase intent	0.64	0.54	
	Other attitudes, e.g. liking	0.65	0.55	
	Knowledge	0.65	0.55	
F2: Market position Position of the firm relative to competitors	Market share	0.59	0.56	0.84
	Relative price	0.61	0.56	
	Loyalty (share)	0.62	0.59	
	Penetration	0.58	0.57	
	Relative consumer satisfaction	0.59	0.63	
	Relative perceived quality	0.59	0.66	
	Share of voice	(0.50)[a]	0.55	
F3: Financial position The level of incoming cash flow and profitability as the difference between this cash flow and the investment required	Sales	0.82	0.76	0.84
	Gross margins	0.72	0.64	
	Profit/profitability	0.82	0.74	

Aalto University
School of Business

Source: Frösén, J., Tikkanen, H., Jaakkola, M. and Vassinen, A. (2013), "Marketing performance assessment systems and the business context", European Journal of Marketing, Vol. 47 No. 5, pp. 715–737.

Factor analysis: key statistics

- Factor loading: correlation between the variables and the factors
- Communality: the amount of variance a variable shares with all the other variables being considered. The symbol typically used for a communality coefficient is h^2.
- Eigenvalue: the total variance explained by each factor. This is commonly used for determination the number of factors. In this approach, only factors with eigenvalues greater than 1.00 are retained.

Factor analysis: key statistics

- Bartlett's test of sphericity can be used to test the null hypothesis that the variables are uncorrelated in the population; in other words, the population correlation matrix is an identity matrix. If this hypothesis cannot be rejected, then the appropriateness of factor analysis should be questioned.
- Another useful statistic is the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy. Small values of the KMO statistic indicate that the correlations between pairs of variables cannot be explained by other variables and that factor analysis may not be appropriate.

Source: Nunan et al. (2020): Marketing Research - Applied Insight. Sixth Edition.

Factor analysis: different methods of factor analysis

- In **principal components analysis**, the total variance in the data is considered. The diagonal of the correlation matrix consists of unities, and full variance is brought into the factor matrix. Principal components analysis is recommended when the primary concern is to determine the minimum number of factors that will account for maximum variance in the data for use in subsequent multivariate analysis. The factors are called principal components.
- In **common factor analysis**, the factors are estimated based only on the common variance. Communalities are inserted in the diagonal of the correlation matrix. This method is appropriate when the primary concern is to identify the underlying dimensions and the common variance is of interest. This method is also known as principal axis factoring.

Source: Nunan et al. (2020): Marketing Research - Applied Insight. Sixth Edition.

Factor analysis: number of factors

- **A priori determination.** Sometimes, because of prior knowledge, the researcher knows how many factors to expect and thus can specify the number of factors to be extracted beforehand.
- **Determination based on eigenvalues.** In this approach, only factors with eigenvalues greater than 1.0 are retained. An eigenvalue represents the amount of variance associated with the factor. Hence, only factors with a variance greater than 1.0 are included. Factors with variance less than 1.0 are no better than a single variable, since, due to standardisation, each variable has a variance of 1.0. If the number of variables is less than 20, this approach will result in a conservative number of factors.

Source: Nunan et al. (2020): Marketing Research - Applied Insight. Sixth Edition.

Factor analysis: number of factors

- **Determination based on scree plot.** A scree plot is a plot of the eigenvalues against the number of factors in order of extraction. Experimental evidence indicates that the point at which the scree begins denotes the true number of factors. Generally, the number of factors determined by a scree plot will be one or a few more than that determined by the eigenvalue criterion.
- **Determination based on percentage of variance.** In this approach, the number of factors extracted is determined so that the cumulative percentage of variance extracted by the factors reaches a satisfactory level. It is recommended that the factors extracted should account for at least 60% of the variance.

Source: Nunan et al. (2020): Marketing Research - Applied Insight. Sixth Edition.

Factor analysis: number of factors

- **Determination based on split-half reliability.** The sample is split in half and factor analysis is performed on each half. Only factors with high correspondence of factor loadings across the two subsamples are retained.
- **Determination based on significance tests.** It is possible to determine the statistical significance of the separate eigenvalues and retain only those factors that are statistically significant. A drawback is that with large samples (size greater than 200), many factors are likely to be statistically significant, although from a practical viewpoint many of these account for only a small proportion of the total variance.

Source: Nunan et al. (2020): Marketing Research - Applied Insight. Sixth Edition.

Factor analysis: rotation

- Although the initial or unrotated factor matrix indicates the relationship between the factors and individual variables, it seldom results in factors that can be interpreted, because the factors are correlated with many variables. Therefore, **through rotation, the factor matrix is transformed into a simpler one that is easier to interpret.**
- The most commonly used method for rotation is the **varimax procedure**. This is an orthogonal method of rotation that minimises the number of variables with high loadings on a factor, thereby enhancing the interpretability of the factors. Orthogonal rotation results in factors that are uncorrelated.
- The rotation is called **oblique rotation** when the axes are not maintained at right angles, and the factors are correlated. Sometimes, allowing for correlations among factors can simplify the factor pattern matrix. Oblique rotation should be used when factors in the population are likely to be strongly correlated.

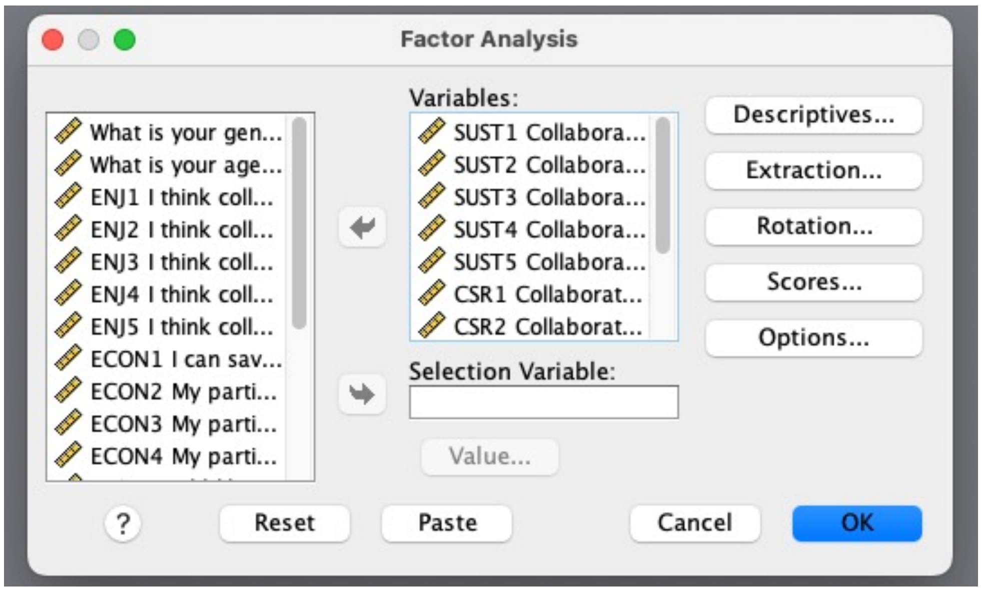
Source: Nunan et al. (2020): Marketing Research - Applied Insight. Sixth Edition.

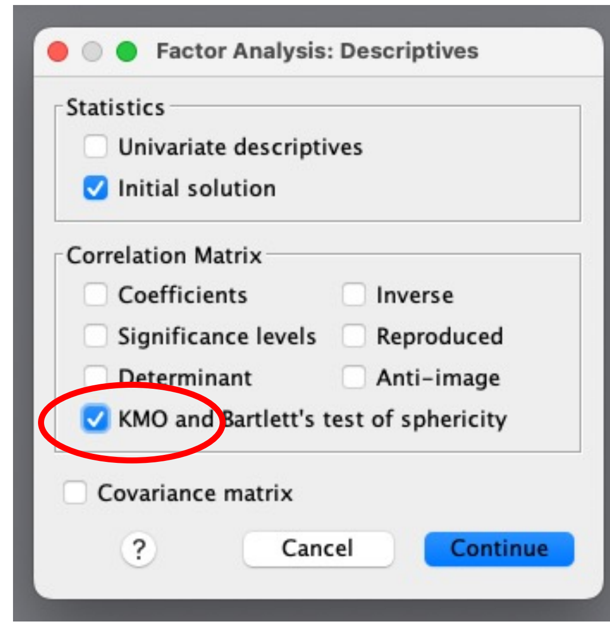
Factor analysis: summary

- KMO and Bartlett' test: If KMO>0.50 and Bartlett's test sig. <0.05 data is OK for factor analysis.
- Which choices are most common in master's thesis:
    - Principal components analysis
    - Determination of the number of factors: eigenvalues greater than 1.
    - Rotation: Varimax.

Performing factor analysis with SPSS

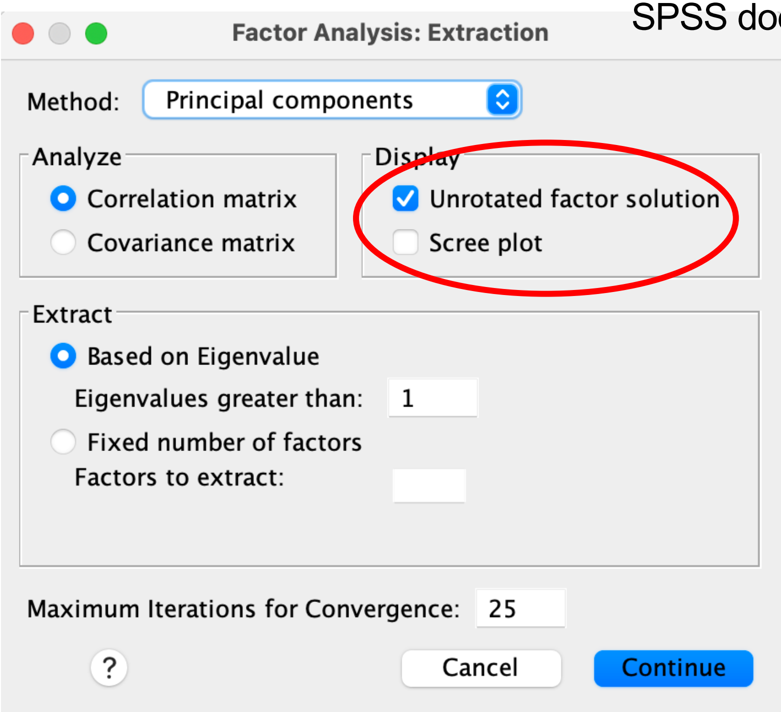
- Analyze – Dimension reduction - Factor

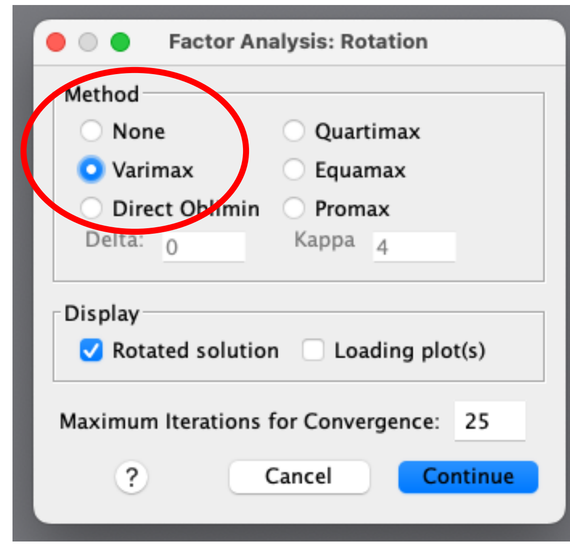




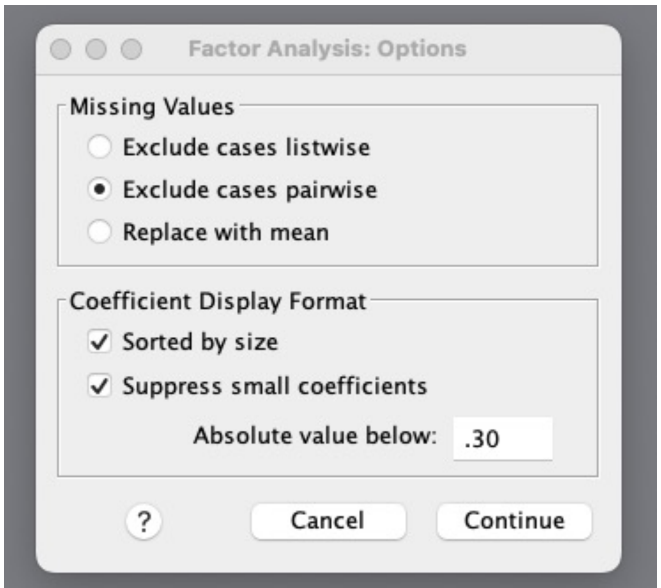
Performing factor analysis with SPSS

Keep the tick “Unrotated factor solution” there. You do not want the “unrotated factor solution” but without that SPSS does not print the communalities.





Performing factor analysis with SPSS



Listwise - if the respondent has any missing value then the respondent is omitted from the data analysis.

Pairwise - the respondent is dropped only on analyses involving variables that have missing values. <u>This is usually better choice.</u>

Coefficient Display Format affects only presentation of the results. These choices make the results easier to read.

Performing factor analysis with SPSS

Factor loadings and communalities from here. KMO & Bartlett's test below.

Rotated Component Matrix[a]

	Component 1	2	3
ENJ3 I think collaborative consumption is fun.	.863		
ENJ1 I think collaborative consumption is enjoyable.	.863		
ENJ5 I think collaborative consumption is pleasant.	.840	.304	
ENJ2 I think collaborative consumption is exciting.	.836		
ENJ4 I think collaborative consumption is interesting.	.736	.305	
ECON2 My participation in collaborative consumption benefits me financially.		.848	
ECON3 My participation in collaborative consumption can improve my economic situation.		.846	
ECON1 I can save money if I participate in collaborative consumption.		.813	
ECON4 My participation in collaborative consumption saves me time.	.349	.630	
GREEN1 I actively recycle items that I am able to.			.872
GREEN2 I try to repair or reuse items rather than throwing them away.			.772
GREEN3 I actively try to reduce my carbon footprint.			.729

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Communalities

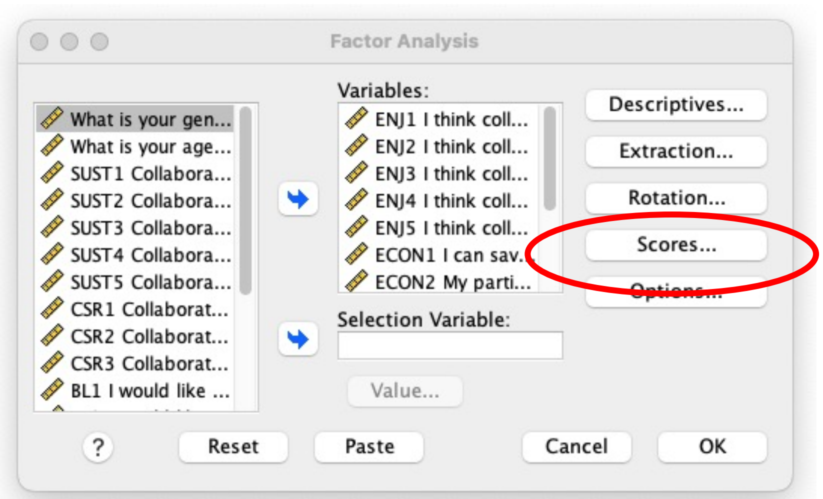
	Initial	Extraction
ENJ1 I think collaborative consumption is enjoyable.	1.000	.842
ENJ2 I think collaborative consumption is exciting.	1.000	.722
ENJ3 I think collaborative consumption is fun.	1.000	.811
ENJ4 I think collaborative consumption is interesting.	1.000	.648
ENJ5 I think collaborative consumption is pleasant.	1.000	.808
ECON1 I can save money if I participate in collaborative consumption.	1.000	.816
ECON2 My participation in collaborative consumption benefits me financially.	1.000	.851
ECON3 My participation in collaborative consumption can improve my economic situation.	1.000	.794
ECON4 My participation in collaborative consumption saves me time.	1.000	.519
GREEN1 I actively recycle items that I am able to.	1.000	.810
GREEN2 I try to repair or reuse items rather than throwing them away.	1.000	.645
GREEN3 I actively try to reduce my carbon footprint.	1.000	.544

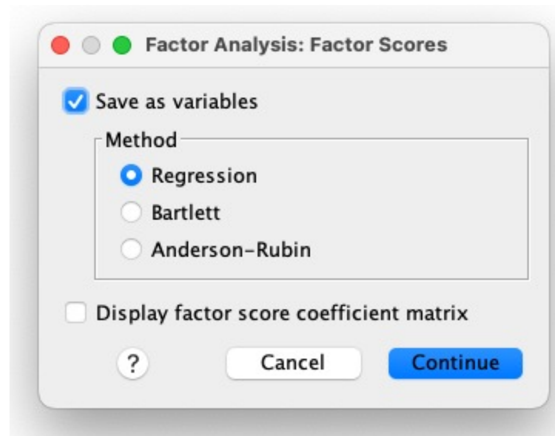
Extraction Method: Principal Component Analysis.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.860
Bartlett's Test of Sphericity	Approx. Chi-Square	1620.295
	df	66
	Sig.	<.001

Saving factor scores with SPSS





Factor scores are composite scores estimated for each respondent on the derived factors. You save them as new variables and use them in further analysis, e.g. cluster analysis.

SPSS: Cronbach's Alpha Statistic

Cronbach’s Alpha statistic

- The best way of estimating the reliability of a scale is to compute Cronbach’s alpha statistic.
- Reverse the scale if necessary (if two questions are: ‘I like.. ‘ and one ‘I dislike..’, you have to reverse the last scale)!
- Cronbach’s Alpha varies from 0 to 1, and often values greater than 0.60 are considered acceptable.
- ‘Alpha if item deleted’ statistic in SPSS is used to identify well-fitting and poorly fitting items.

Cronbach's Alpha statistic: Example

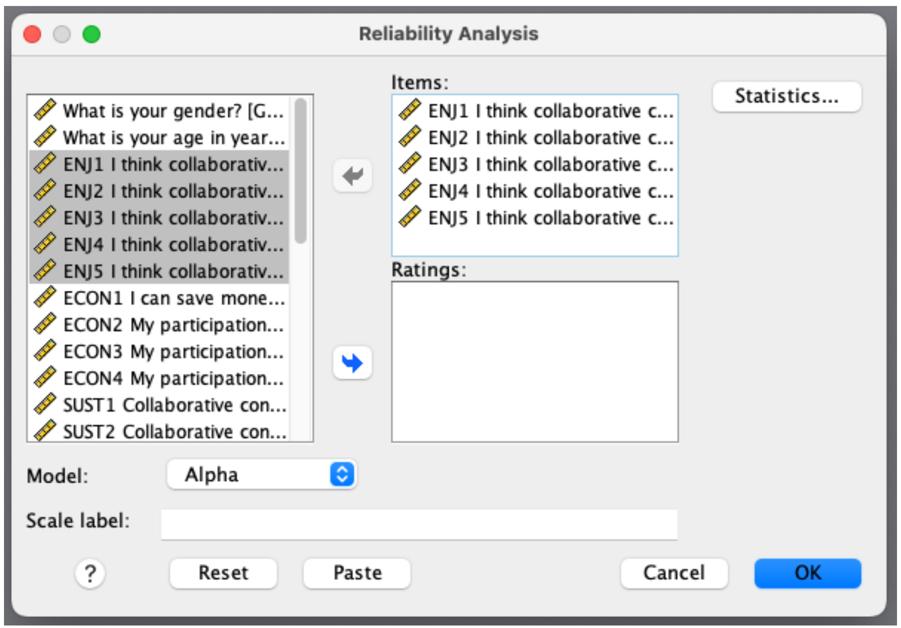
Factor	Metrics	Factor loading	Communality	Cronbach's alpha
F1: Brand equity Actual and potential customer attitudes, perceptions, thoughts, and feelings	Awareness	0.70	0.58	0.90
	Salience	0.68	0.53	
	Perceived quality/esteem	0.68	0.61	
	Consumer satisfaction	0.63	0.52	
	Relevance to consumer	0.67	0.58	
	Image/personality/identity	0.69	0.58	
	(Perceived) differentiation	0.62	0.56	
	Commitment/purchase intent	0.64	0.54	
	Other attitudes, e.g. liking	0.65	0.55	
	Knowledge	0.65	0.55	
F2: Market position Position of the firm relative to competitors	Market share	0.59	0.56	0.84
	Relative price	0.61	0.56	
	Loyalty (share)	0.62	0.59	
	Penetration	0.58	0.57	
	Relative consumer satisfaction	0.59	0.63	
	Relative perceived quality	0.59	0.66	
	Share of voice	(0.50)[a]	0.55	
F3: Financial position The level of incoming cash flow and profitability as the difference between this cash flow and the investment required	Sales	0.82	0.76	0.84
	Gross margins	0.72	0.64	
	Profit/profitability	0.82	0.74	

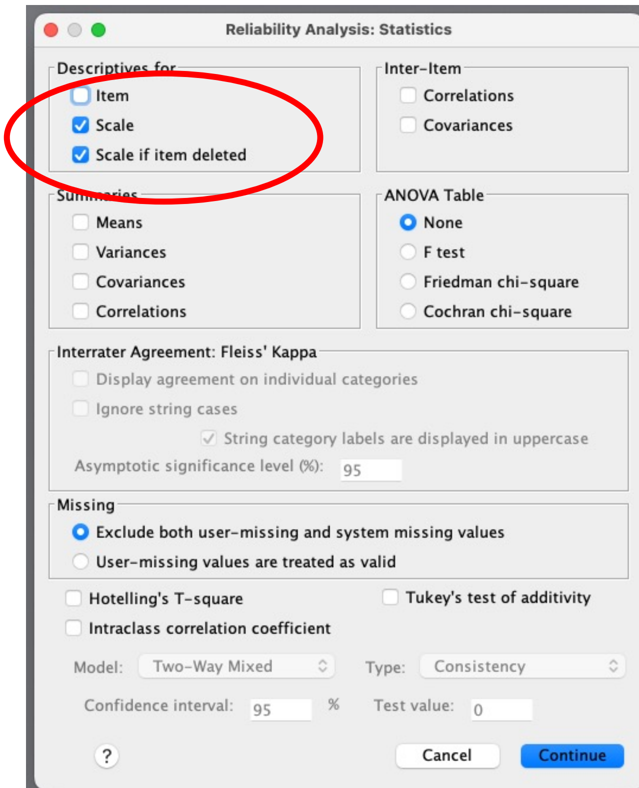
Aalto University
School of Business

Source: Frösén, J., Tikkanen, H., Jaakkola, M. and Vassinen, A. (2013), "Marketing performance assessment systems and the business context", European Journal of Marketing, Vol. 47 No. 5, pp. 715–737.

Performing reliability analysis with SPSS

- Analyze – Scale – Reliability Analysis





Performing reliability analysis with SPSS

Reliability Statistics

Cronbach's Alpha	N of Items
.920	5

Excellent (.80) Alpha value.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
ENJ1 I think collaborative consumption is enjoyable.	20.7350	20.809	.858	.890
ENJ2 I think collaborative consumption is exciting.	21.1050	20.667	.746	.913
ENJ3 I think collaborative consumption is fun.	20.9600	19.697	.834	.894
ENJ4 I think collaborative consumption is interesting.	20.3450	22.609	.709	.918
ENJ5 I think collaborative consumption is pleasant.	20.8750	20.512	.834	.894

Alpha does not get better if some item is deleted. But it would be almost the same without ENJ4

SPSS: Cluster Analysis

Cluster analysis

- The objective of cluster analysis is to group objects based on their characteristics so that there is a greater similarity among units within groups than there is among units in different groups.
- Thus, cluster analysis reduces the number of objects (e.g. customers) whereas factor analysis reduces number of variables.
- Most common method in master's thesis: k-means clustering (non-hierarchical clustering).
- Theoretical, conceptual or practical considerations may suggest a certain number of clusters.
- The relative sizes of the clusters should be meaningful.

Cluster analysis: Example

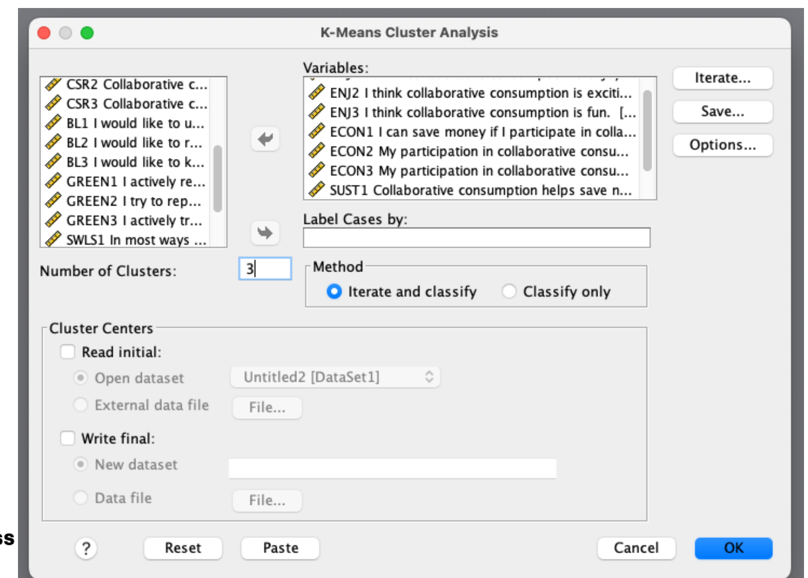
Table III.
Cluster centroids of the groupings of companies

Cluster	F1: Brand equity	F2: Market position	F3: Financial position	F4: Long-term firm value	F5: Innovation	F6: Customer feedback	F7: Customer equity	F8: Channel activity	F9: Sales process
C1: Parsimony seekers	−0.47	−0.52	0.59	0.36	−0.39	0.47	−0.86	0.18	−0.33
C2: Casual marketers	−0.40	−0.21	−2.36	−0.33	−0.17	−0.14	−0.35	−0.28	−0.28
C3: Data collectors	−0.28	0.79	0.07	0.68	0.47	0.02	0.32	0.17	0.43
C4: Future builders	0.46	−0.89	0.27	−0.04	0.49	−0.05	0.28	−0.48	0.32
C5: Conventional marketers	0.40	0.46	0.35	−0.75	−0.54	−0.24	0.22	0.24	−0.34

Source: Frösén, J., Tikkanen, H., Jaakkola, M. and Vassinen, A. (2013), “Marketing performance assessment systems and the business context”, European Journal of Marketing, Vol. 47 No. 5, pp. 715–737.

Performing cluster analysis with SPSS

- Analyze – Classify – K-means cluster
- You can create a new variable showing cluster membership. Select Save and Cluster membership.



Performing cluster analysis with SPSS

Final Cluster Centers

	Cluster 1	Cluster 2	Cluster 3
ENJ1 I think collaborative consumption is enjoyable.	4.86	6.06	3.72
ENJ2 I think collaborative consumption is exciting.	4.13	5.79	3.81
ENJ3 I think collaborative consumption is fun.	4.39	5.99	3.53
ECON1 I can save money if I participate in collaborative consumption.	5.54	6.41	3.50
ECON2 My participation in collaborative consumption benefits me financially.	5.45	6.35	3.50
ECON3 My participation in collaborative consumption can improve my economic situation.	4.94	6.09	3.53
SUST1 Collaborative consumption helps save natural resources.	5.00	5.77	3.59
SUST2 Collaborative consumption is a sustainable mode of consumption.	5.20	5.86	3.94
SUST3 Collaborative consumption is ecological.	5.10	5.85	3.66

Number of Cases in each Cluster

Cluster	1	69.000
	2	99.000
	3	32.000
Valid		200.000
Missing		.000

Thank you!