

Joerg Niessing
Affiliate Professor of Marketing
Theos Evgeniou
Professor of Decision Sciences



[Big]-Data Analytics for Businesses

Understand the world. Expand your world.

Group Project: Deliverable #1

SESSION 4: Please prepare a 1-page report. You will need to email this report **before 9am of the day of sessions 4-5**. *Delayed reports will lose 10% of the grade*. For this report you need to answer the following questions:

Name of the company you are studying (if it is for a new venture, please describe the business model, market, and key processes)

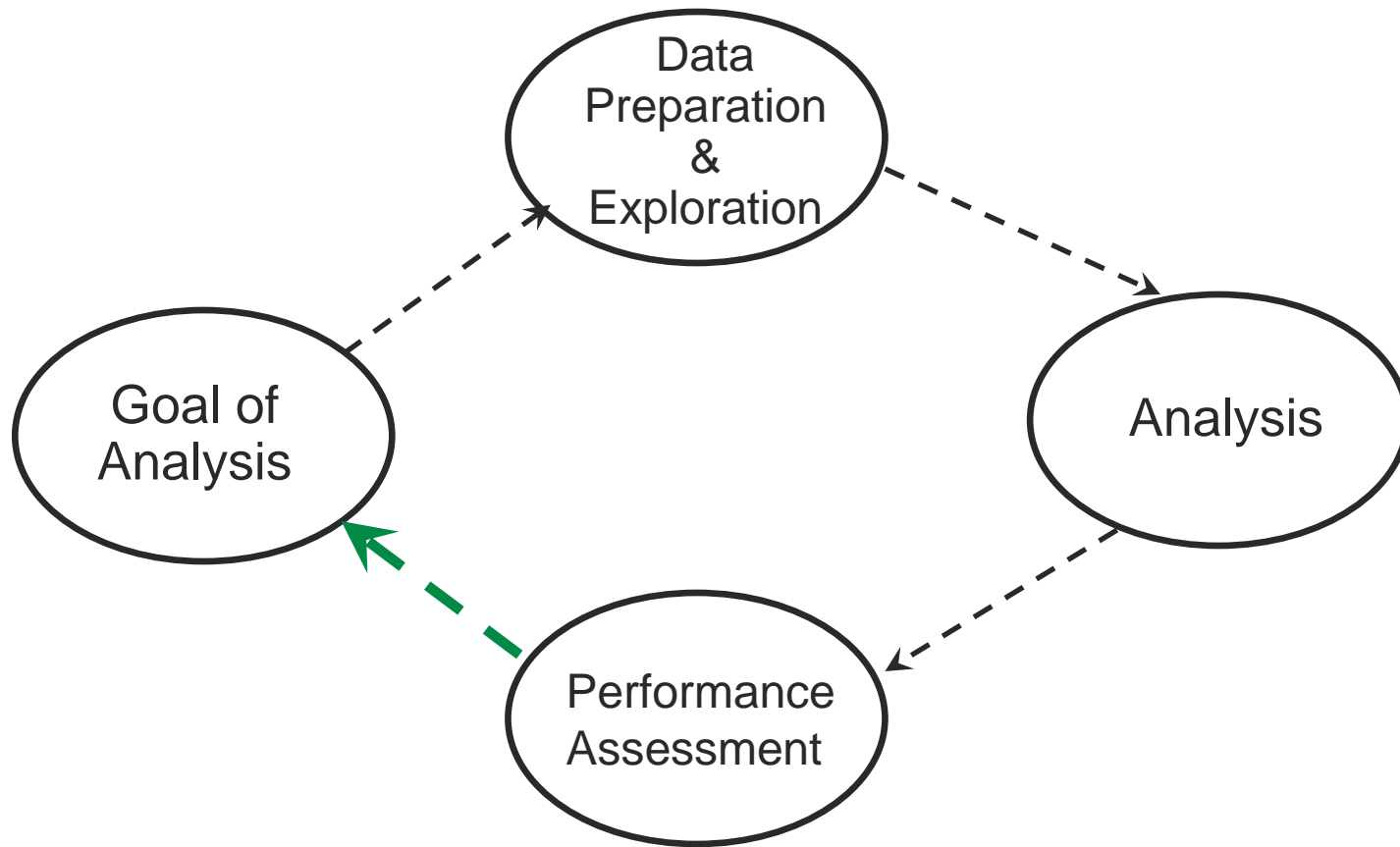
1. Which business processes and management decisions are affected from your data analytics proposal that you are sending to that company?
2. What changes (e.g. organizational, processes, systems) do you expect the company will need to make order to effectively use the results of data analytics? What types of resistance do you expect, and why?
3. How will you measure the value of your proposed data analytics project (e.g. process efficiency, predictive improvement with specific business impact, etc)?

Also be prepared to briefly describe your project in class (no presentation needed, only brief answers to these questions as well as any other information you may find necessary).

Why this class? Three goals

1. “Big Picture”: Develop your intuition about identifying data analytics opportunities and their implementation challenges
2. “Medium Picture”: learn how to approach data analytics projects
3. “Dirty Hands”: Learn how to perform, read, and use key data analytics methods.

The *Iterative Process Cycle*



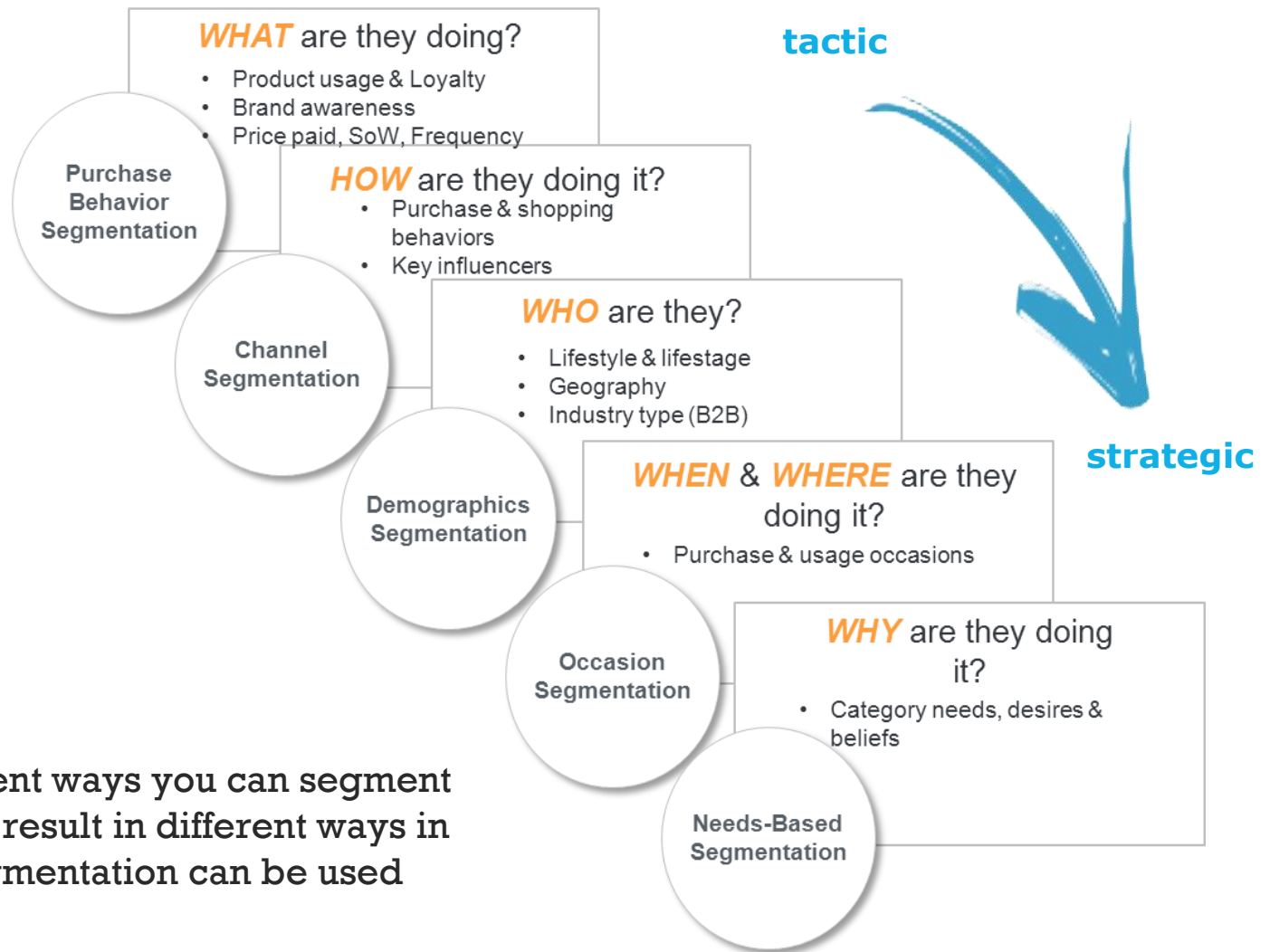
What is Segmentation?

- Process of categorizing observations into groups for the purpose of focusing on where value is being created and could be captured (**homogeneous within and heterogeneous across groups**)
- Primary **OBJECTIVE**: Driving business efficiency, effectiveness, and profitability by aligning the business with the needs of strategically defined customer groups
- Helps **ANSWER** the **who, what, when, where, why**, and **how** of current and potential customers.

Why Segmentation?

- **SEGMENTATION** is a critical enabler to achieve business objectives and realize benefits (concept of WINWIN and OVP)
- **SEGMENTATION** enables organizations to become more customer-centric (better aligned products to customer needs)
- **SEGMENTATION** is critical to identify white spaces for new products/offerings
- **SEGMENTATION** helps organizations to optimize their retention and acquisition strategy
- **SEGMENTATION** is often used to optimize pricing across different products and segments
- **SIMPLE TERMS: Focus!**

What data should we gather?



There are different ways you can segment a market, which result in different ways in which the segmentation can be used

Often different data sources will be combined in one segmentation

Primary Data (Qual and Quant)

Sample variables:

- Behavioral
- Satisfaction
- Preferred communication channels
- Preferred engagement level
- Attitudes about the industry

Customer Data

Sample variables:

- Product/Service Usage
- Subscription
- Features usage
- Social Network integration
- Demographics
- Acquisition channel

3rd Party Data

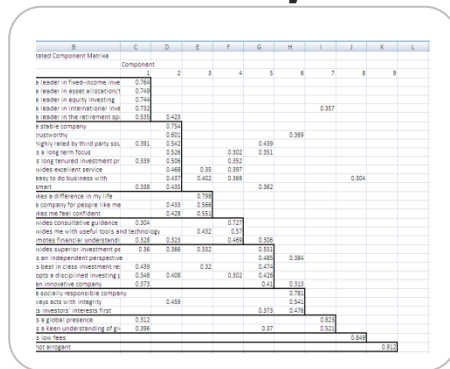
Sample variables:

- Lifestyle
- Behavioral
- Demographics



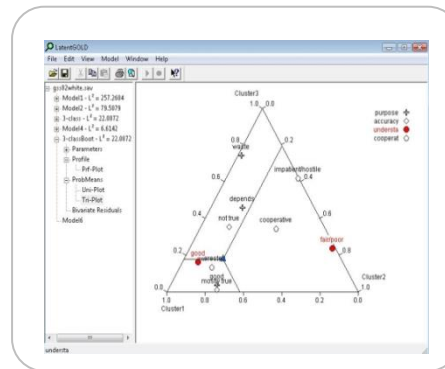
What we will cover today and in the next session?

Data Processing/ Factor Analysis



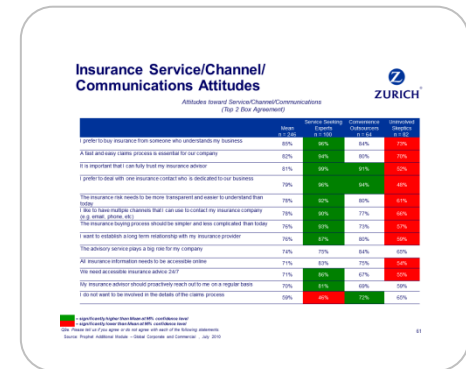
Reduces number of variables by assessing and eliminating those which are saying the same thing

Cluster analyses



First iterations: we look for the best solutions, that minimize differences within segments, but maximize differences between segments

Review and refine



Arriving at the final segmentation solution involves a process of reviewing and refining latent class outputs and different solution profiles

Class Outline: three tools you will learn

- 1. Finding important factors that summarize your data, and visualizing your data:**

Factor Analysis (Sessions 2 and 3)

- 2. Finding a few clusters of similar data:**

Cluster Analysis (Sessions 4 and 5)

- 3. Discriminating among and predicting successes vs failures:**

Logistic Regression and Tree Analyses (Sessions 6 and 7)

Today's Plan

- Factor Analysis for Data Preparation
- Judgment in Data Analytics

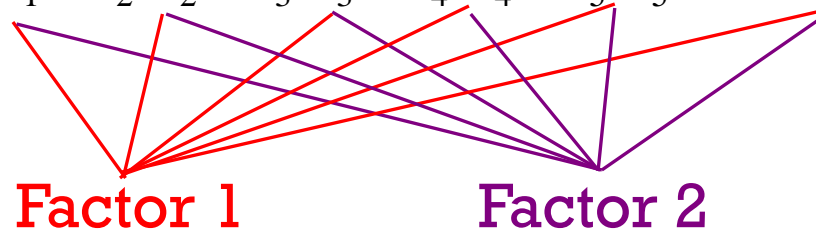
What is Factor Analysis?

- A Variables Reduction Technique (“group the columns”)
 - the goal is to summarize the information in a larger number of *correlated* variables into a smaller number of factors that are not correlated with each other
 - a factor is a weighted sum of the original variables
- Factor analysis can be used as a first step for further analysis later (i.e. cluster analysis, discriminant analysis, regression, etc). Can also be used to solve multicollinearity in regression.

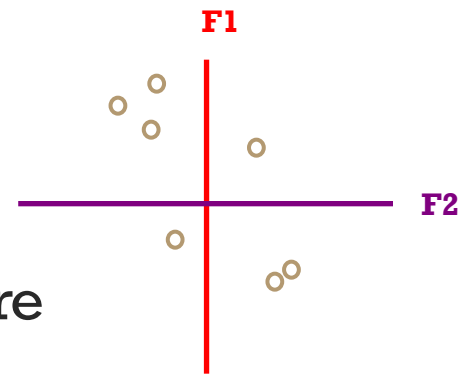
Factor Analysis and Multicollinearity

1. Transform the original variables (ONLY the x's) into a smaller set of factors

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + \dots + b_nX_n$$

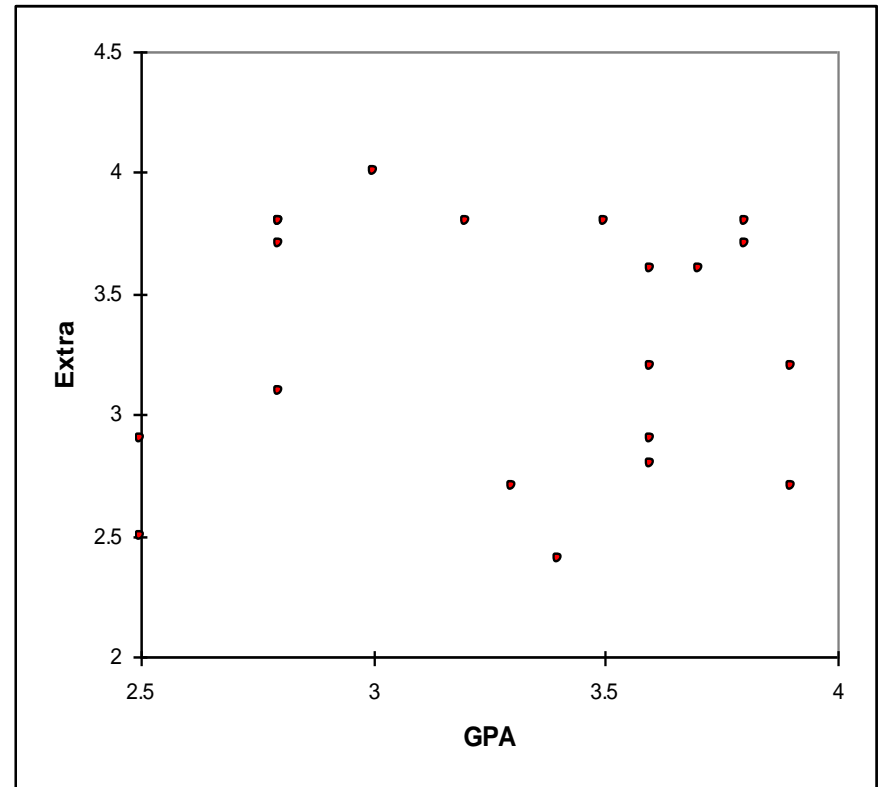
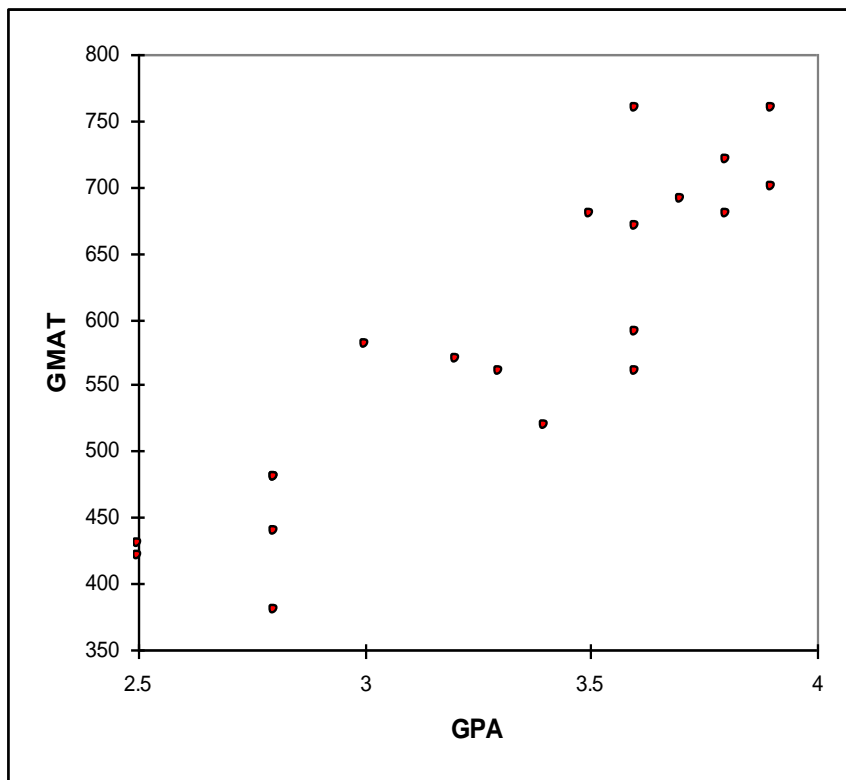


$$Y = d_0 + d_1F_1 + d_2F_2$$



2. Understand the underlying structure
3. Use the factors for subsequent analysis

Graphical Intuition: Factor Analysis works when data are correlated



Key Questions

1. Can we really simplify the data by grouping the columns?
2. How many factors should we use?
3. How good are the factors we found?
4. How interpretable are the factors we found?

Factor Analysis and Perceptual Maps: 8 (Easy) Steps

1. Confirm the data are metric (interval scale)
2. Decide whether to scale standardize the data
3. Check correlation matrix to see if Factor Analysis makes sense
4. Decide on the number of factors to be derived
5. Estimate factors and get the factor loadings (“components”) and loading plots (“component plots”)
6. Consider rotation of factors (technical but useful)
7. Interpret the factors
8. Save factor scores for subsequent analyses

Factor Analysis: Example

Applying Factor Analysis: Evaluating MBA Applications

Variables available:

- GPA
- GMAT score
- Scholarships, fellowships won
- Evidence of Communications skills
- Prior Job Experience
- Organizational Experience
- Other extra curricular achievements

*Which variables do you believe correlate with each other?
What do these variables capture?*

Step 3: Quick and dirty sense of the data – Looking at the correlation matrix

		Correlations						
		GPA	GMAT	Fellow	Comm	Job Ex	Organze	Extra
GPA	Pearson Correlation	1	.898**	.924**	.555*	.154	-.031	.011
	Sig. (2-tailed)		.000	.000	.011	.518	.898	.964
	N	20	20	20	20	20	20	20
GMAT	Pearson Correlation	.898**	1	.858**	.782**	.329	.188	.160
	Sig. (2-tailed)	.000		.000	.000	.157	.428	.500
	N	20	20	20	20	20	20	20
Fellow	Pearson Correlation	.924**	.858**	1	.585**	.182	.011	.021
	Sig. (2-tailed)	.000	.000		.007	.443	.963	.928
	N	20	20	20	20	20	20	20
Comm	Pearson Correlation	.555*	.782**	.585**	1	.597**	.474*	.391
	Sig. (2-tailed)	.011	.000	.007		.005	.035	.089
	N	20	20	20	20	20	20	20
Job Ex	Pearson Correlation	.154	.329	.182	.597**	1	.802**	.768**
	Sig. (2-tailed)	.518	.157	.443	.005		.000	.000
	N	20	20	20	20	20	20	20
Organze	Pearson Correlation	-.031	.188	.011	.474*	.802**	1	.610**
	Sig. (2-tailed)	.898	.428	.963	.035	.000		.004
	N	20	20	20	20	20	20	20
Extra	Pearson Correlation	.011	.160	.021	.391	.768**	.610**	1
	Sig. (2-tailed)	.964	.500	.928	.089	.000	.004	
	N	20	20	20	20	20	20	20

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

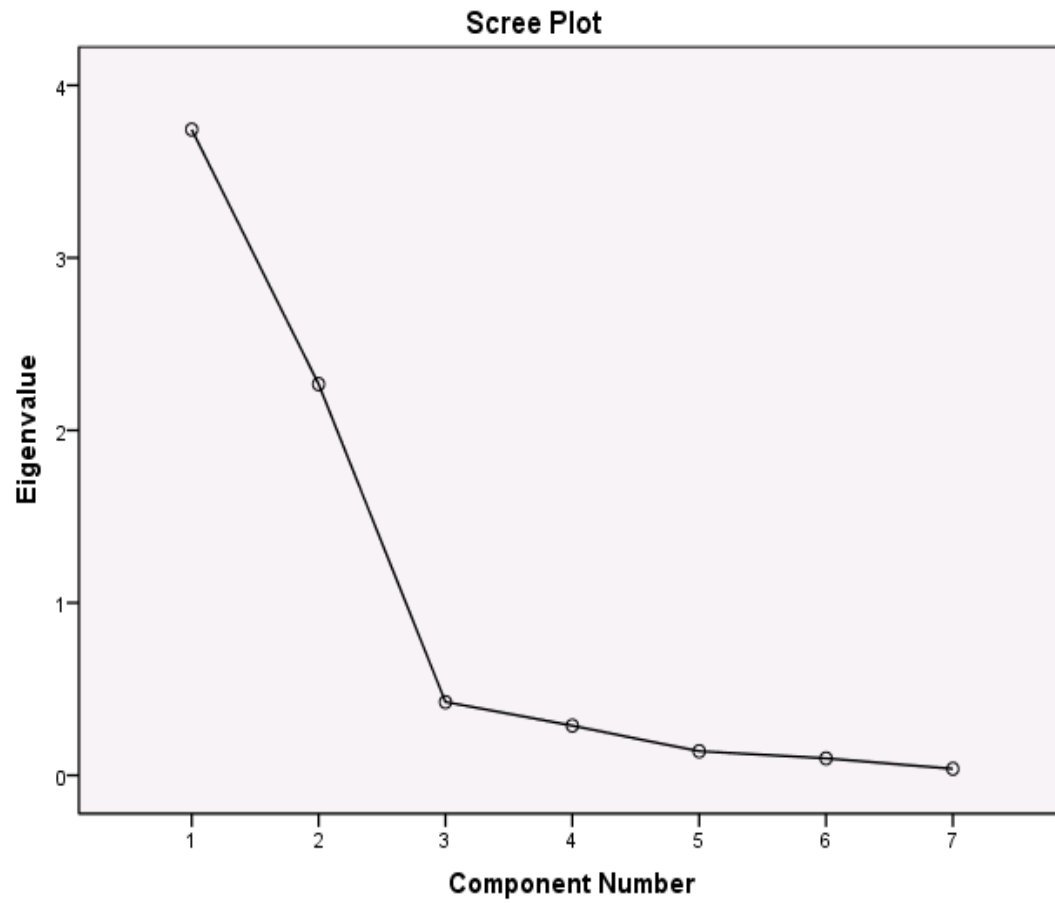
**Even if data is not as neatly correlated as here...
Factor analysis will be helpful**

Step 4: Decide number of factors

Use PCA (Principal Component Analysis)

- Use the correlation matrix of the data – unless scaling is clearly important (pay attention!), then use the covariance
- Factors
 - If there are n variables we will have n factors in total
 - First factor will explain most variance, second next and so on...
- Variance Explained by Factors
 - with standardized variables each variable has a variance of 1, so the total variance in n variables is n
 - each factor will have an associated eigenvalue which is the amount of variance explained by that factor

Scree Plot

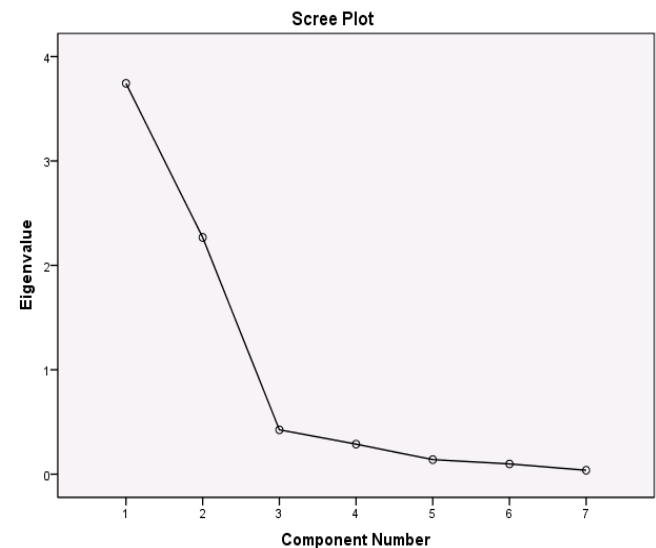


How many factors?

- The maximal number of factors is the number of original variables
- Eigenvalue > 1
- “Elbow” in the Scree plot

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3.744	53.480	53.480
2	2.268	32.398	85.878
3	.425	6.069	91.948
4	.288	4.113	96.060
5	.140	1.994	98.054
6	.098	1.406	99.460
7	.038	.540	100.000

Extraction Method: Principal Component Analysis.



Steps 5-8: Do Factor Analysis with # of factors selected from Step 3

Step 7: Interpret resulting factors

- Use **factor loadings** and **loading plot** to interpret factors
- If it is not interpretable use **rotation** options until we get something that can be interpreted

Step 8: The **factor scores** describe what the factors are – and will be useful for visualization...

(Rotated – Step 6) Factor Loadings: Variables' Correlations with the Factors

Rotated Component Matrix^a

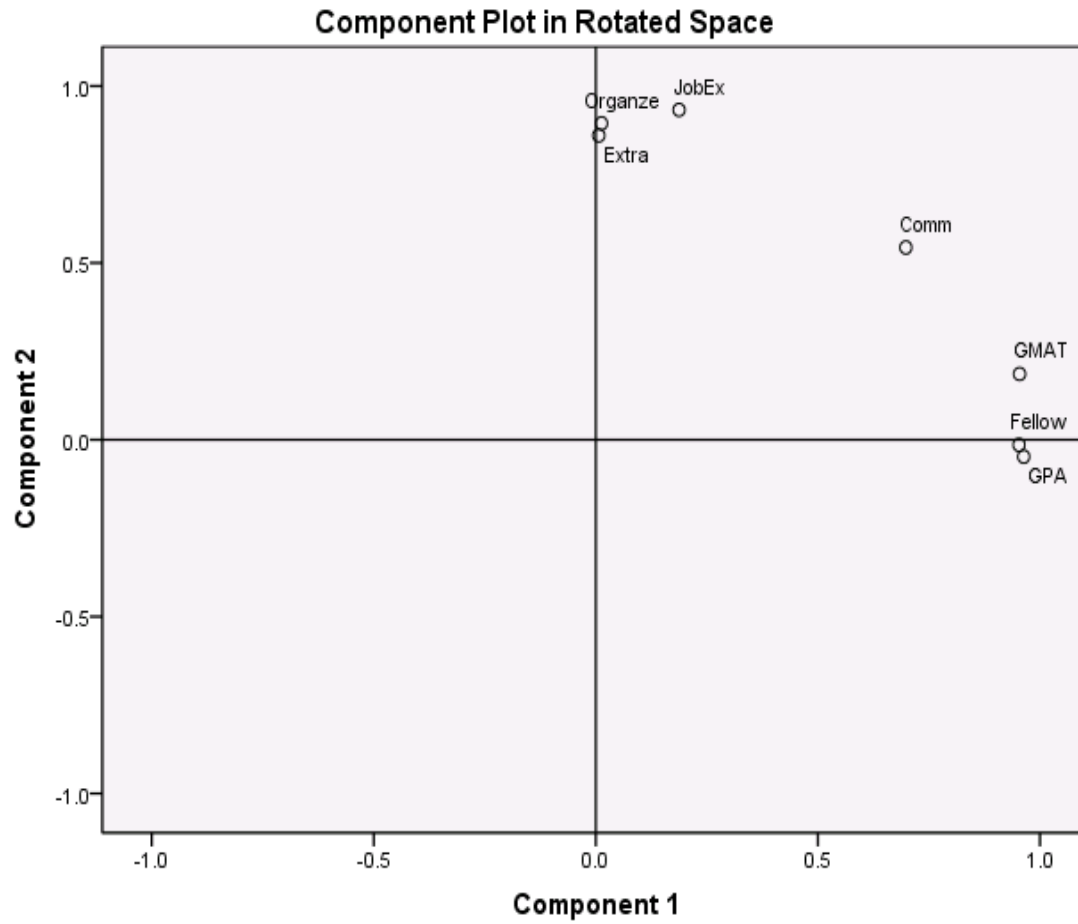
	Component	
	1	2
GPA	.963	
GMAT	.954	
Fellow	.953	
Comm	.698	.543
Job Ex		.933
Organze		.895
Extra		.860

Extraction Method: Principal
Component Analysis.

Rotation Method: Varimax
with Kaiser Normalization.

a. Rotation converged in 3
iterations.

Interpreting Factors: Looking at Loading Plot (with Rotation)



What Factor Loads “Look Good”?

Three Quality Criteria

1. For each factor (column) only a few elements are large (in absolute value)
2. For each initial variable (row) only a few elements are large (in absolute value)
3. Any pair of factors (columns) should have different “patterns” of loading

Factor Analysis: Key (Technical) Notions

1. Correlation
2. Variance explained (eigenvalues)
3. Scree plot
4. Varimax rotation
5. Factor Loadings (“components”)
6. Loading Plot (“component plot”)
7. Factor scores

Example in SPSS:

The Coffee Project

1. Confirm the data are metric (interval scale)
2. Decide whether to scale standardize the data
3. Check correlation matrix to see if Factor Analysis makes sense
4. Decide on the number of factors to be derived
5. Interpret the factors (Consider rotation of factors)
6. Save factor scores for subsequent analyses

Boating Case: Part 1

Group work...

1. How many factors and why (stats and interpretation)?
2. How would you name the factors? What do they mean?
3. What are some issues to consider?

Boating Results

Factor Analysis: A Solution

Session 2+3 Factor Results Boating case - Microsoft Excel

Rotated Component Matrix ^a												
	1	2	3	4	5	6	7	8	9	10	11	12
Q1_27_Boating is the number one thing I do with my spare time	.791											
Q1_28_Boating is my true passion in life	.755											
Q1_20_When not on my boat, I often find myself doing boating related activities	.662											
Q1_16_People tend to come to me for advice about boating	.598	.358				.392						
Q1_17_I consider myself more knowledgeable about boating than the rest of my boating peers	.528	.395				.384						
Q1_13_When buying a boat, I tend to buy the latest and greatest		.764										
Q1_14_When buying accessories for my boat, I tend to buy the latest and greatest		.762										
Q1_15_I am serious about the technology on my boat		.612										
Q1_26_Having a powerful boat is what is most important to me	.421	.511										
Q1_5_I am willing to pay a premium for a brand with a reputation for high quality		.478					.406					
Q1_19_Boating allows me to experience nature			.846									
Q1_18_Boating gives me a feeling of adventure			.727									
Q1_21_Boating helps me escape from everyday life and relax			.585									
Q1_22_Boating helps me stay active			.471									.383
Q1_25_Being in charge of my boat is very important to me	.348		.440	.324								
Q1_7_Owning a boat gives me a sense of achievement				.740								
Q1_6_Owning a boat is a way of rewarding myself for my hard work			.382	.627								
Q1_9_I see my boat as a status symbol		.431		.621								
Q1_3_The brand of boat I buy says a lot about who I am		.440		.543								
Q1_2_When buying a boat, getting the lowest price is more important than the boat brand					.796		-.308					
Q1_12_When it comes to boating, I tend to prefer a basic boat with little to no bells					.732							
Q1_11_I tend to perform minor boat repairs and maintenance on my own						.886						
Q1_4_I only consider buying a boat from a reputable brand							.847					
Q1_8_When buying a boat, functionality is more important than style								.954				

Q1_30_When buying a boat, I rely more on expert opinion than my own

Variance explained: 5 factors - 10 factors - 11 factors - 12 factors - 13 factors - 14 factors

Ready

8:07 PM 5/3/2014

There is A LOT of
JUDGMENT in ANALYTICS:

Your involvement is
CRUCIAL

INSEAD

The Business School
for the World®