



Cluster Analysis in market segmentation research

Đỗ Huệ Hương, PhD

Outline

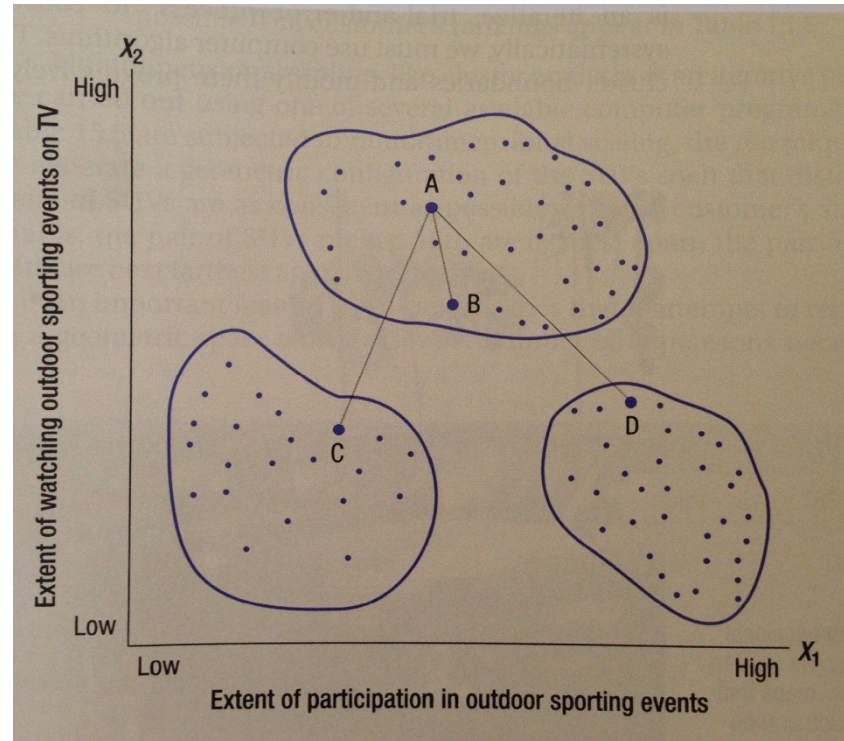
- Cluster Analysis: principles
- Cluster Analysis: practices
- Example: profiling visitors to Cat Tien National Park

Cluster Analysis

- Exploratory multivariate data analysis technique
- Grouping customers
- Different classes of data: psychographics, behaviour, product ratings, usage information, perceived needs or benefits
- CA always forms clusters whether 'natural' clusters exist or not
- Solid conceptual support from the literature
(Hair et al., 2010)

Market segmentation

"a process of dividing customers whose valuations of a product or service vary greatly into groups or segments containing customers whose valuations vary very little within the group but vary greatly among groups"



(Lilien&Rangaswamy, 1998, p.56)

Market segmentation

- Better understand customers → target marketing efforts to the right segments
- Homogeneity
- Parsimony
- Accessibility

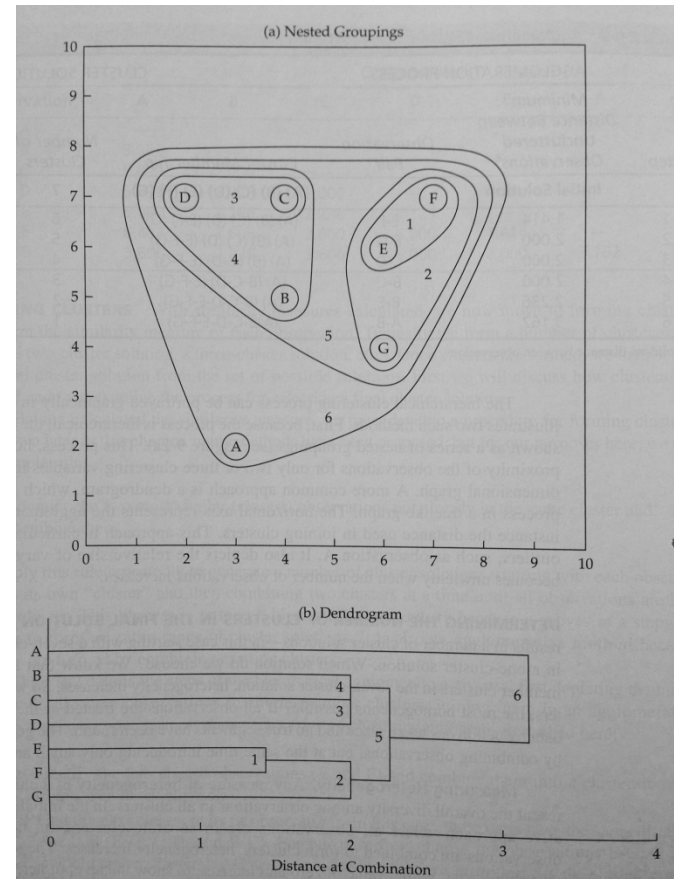
(Lilien & Rangaswamy, 1998)



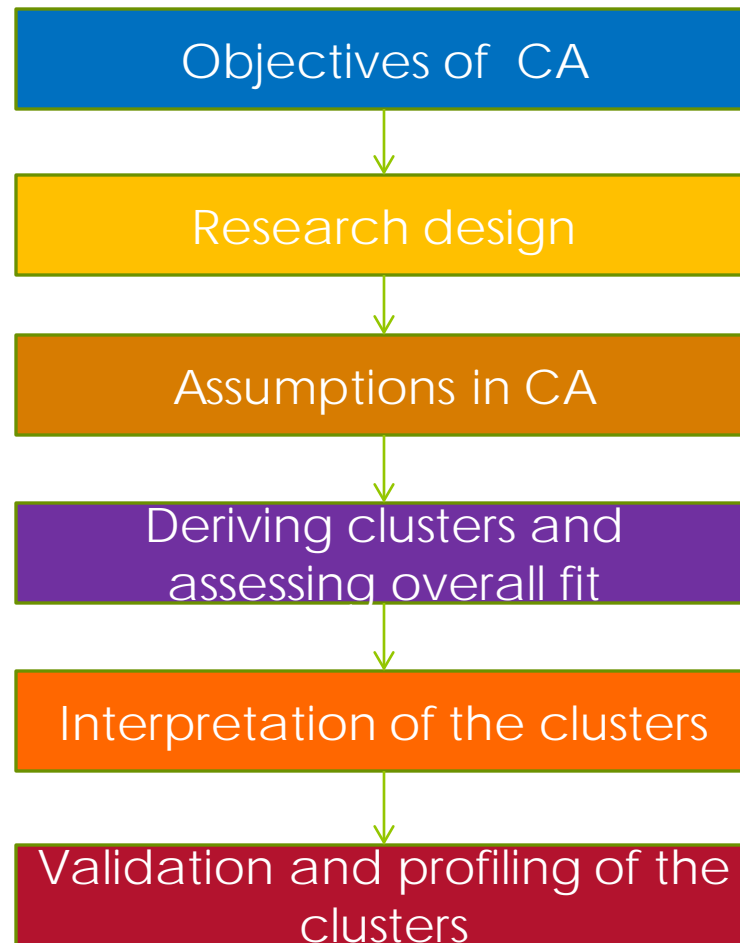
What to look at when clustering?

- Separation
- How many observations per cluster?
- Cluster profiles
- Validation

(SPSS, 2010)



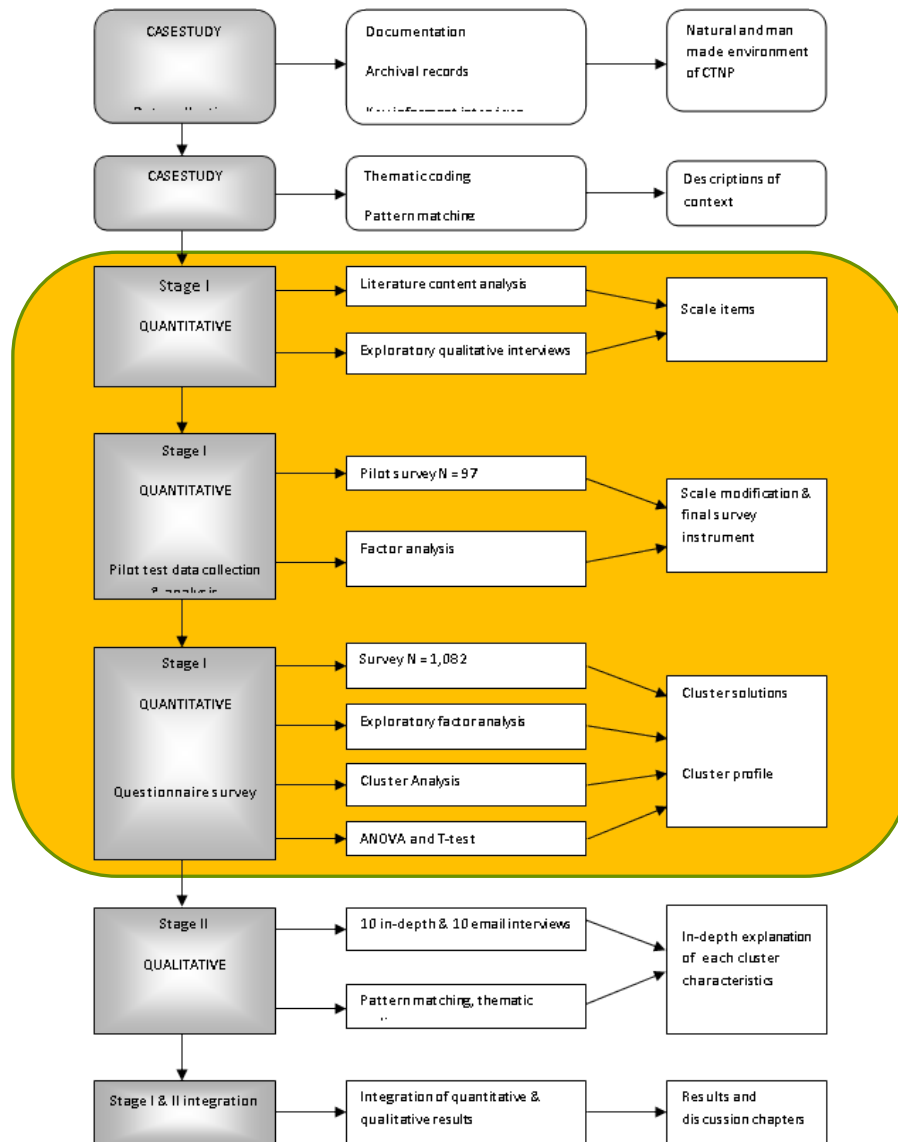
CA decision process





Example:
Profiling
visitors to Cat
Tien National
Park

Overall study



Stage 1 – Objectives of CA

- Research questions
- Selection of clustering variables

Research background

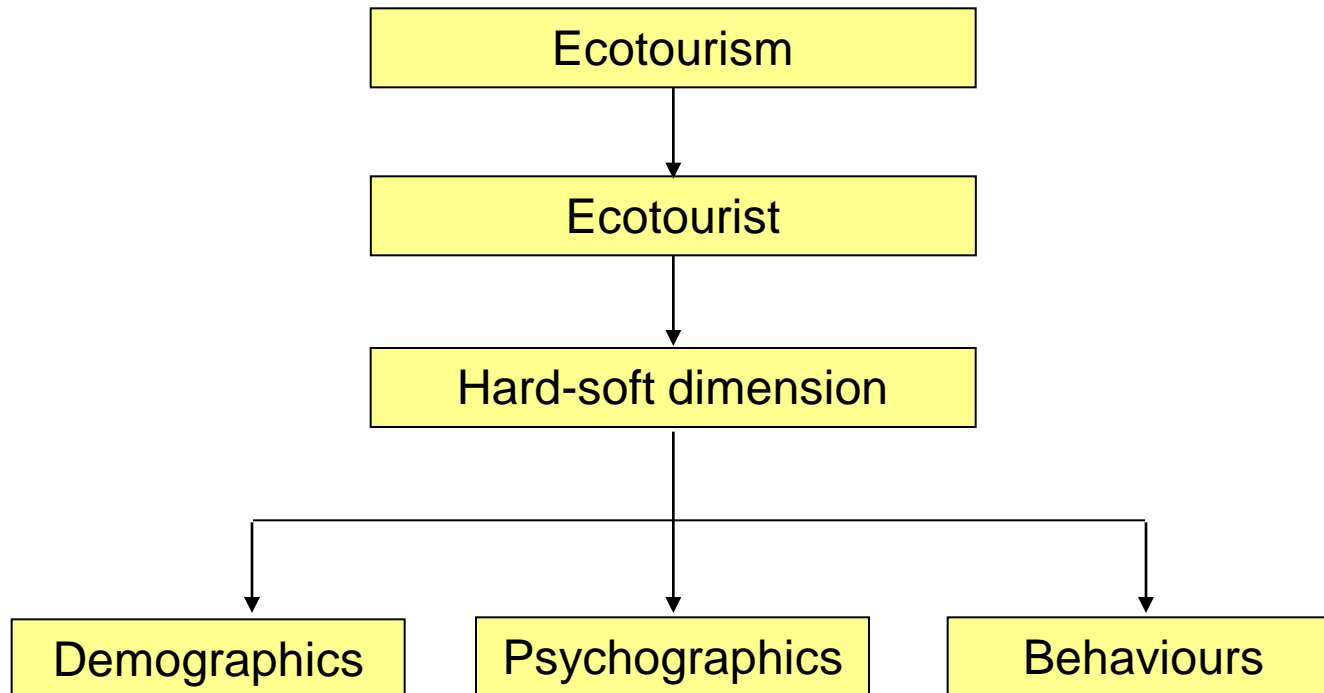
- Ecotourism as a major component of global tourism (Weaver, 2008) with an estimation of contribution up to 20 per cent by UNWTO (Wight, 2001)
- Dominance of the Western conventional market (Eagles & Higgins, 1998)
- Increasing evidence of potential sizable markets in Asia (Cochrane, 2006; Weaver, 2002)



Research background (cont.)

- Two mandates of protected areas: to protect nature and to accommodate visitors
 - New prosperous middle class in Viet Nam
 - Increasing trend of visitation in protected areas (Cochrane, 2007)
 - Potential incompatibility of the two mandates → ecotourism
- This requires a better understanding of ecotourist component of protected area visitation

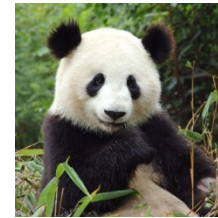
Literature review



Definition of ecotourism

Three core criteria of ecotourism (Blamey, 1997)

- Nature-based attractions
- Learning/education motives
- Environmental, socio/cultural and economic sustainability - Triple bottom line (TBL)



Ecotourists

Definition:

- **Supply side:** people who visit a relatively wild and undisturbed areas (Ceballos Lascuráin, 1986)
- **Demand side:** people have at least an ecotourism experience (Blamey, 1995)



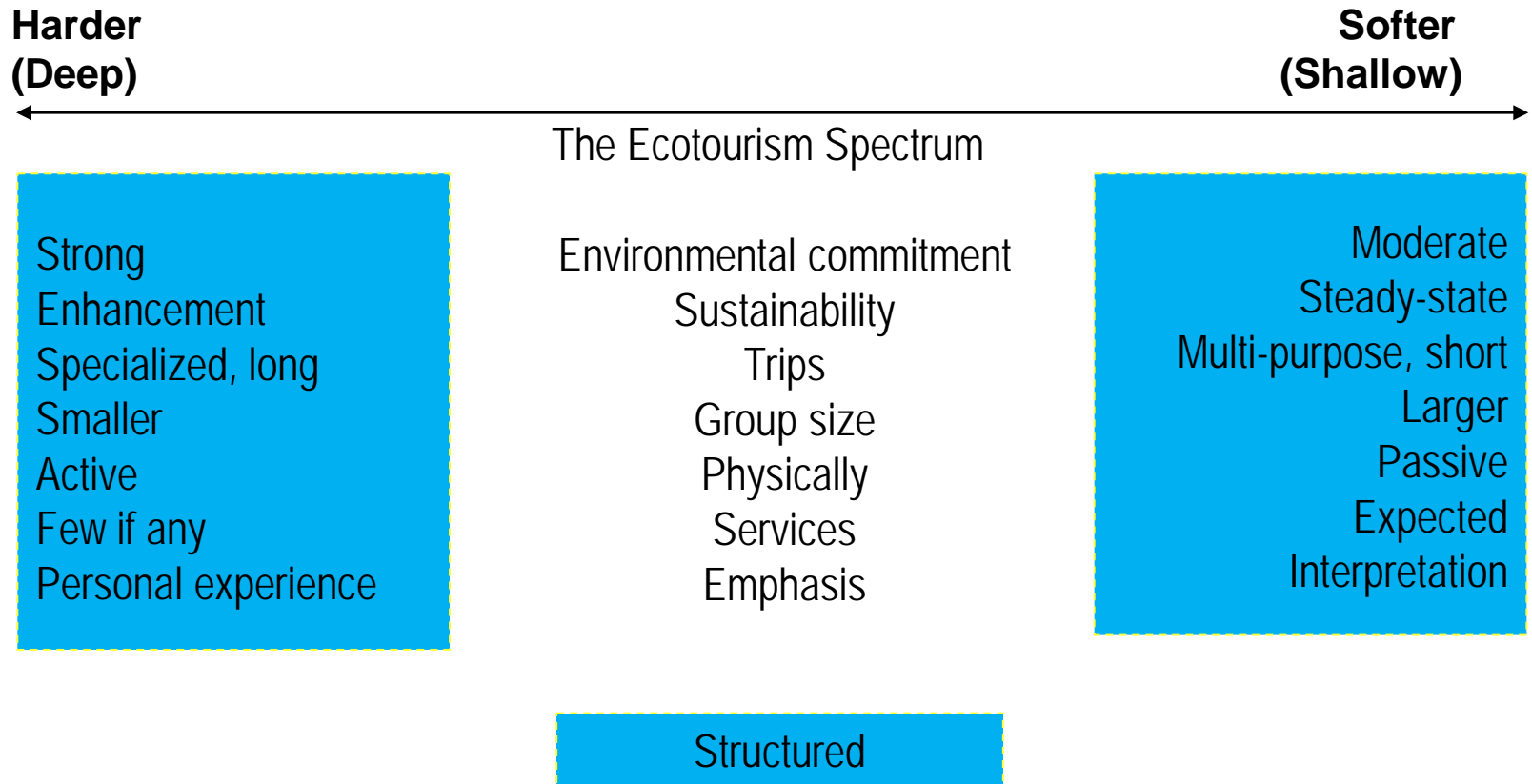
Ecotourist typologies

Source	Hard		Soft	
Conceptual Naess (1973)	Deep ecology		Shallow ecology	
Laarman & Durst (1987)	Hard		Soft	
Acott et. al. (1989)	Deep ecotourism		Shallow ecotourism	
Lindberg (1991)	Hard-core	Dedicated	Mainstream	Casual
Queensland (1997)	Self-reliant ecotourism	Small group ecotourism	Popular ecotourism	
Empirical Pearce & Moscardo (1994)	Nature experience & appreciation 24%			
	Get away, relax with nature 66%		Novelty sun-seekers* 10%	
	[based on a sample of 545 general travellers passing through Cardwell, northern Qld]			
Chapman (1995)	Nature activity 24%	Personal involvement 19%	Laid-back 19%	Social development 8%
	[based on a sample of 507 users of NSW state forests during time of participation]			
Palacio & McCool (1997)	Ecotourists 18%	Nature escapists 22%	Comfortable naturalists 33%	Passive players * 27%
	[based on a sample of 207 travellers through Belize's international airport]			
Meric & Hunt (1998)	Hard-core 1.3%	Dedicated 45%	Mainstream 6.1%	Casual 47.6%
	[based on a sample of 245 visitors who have recent nature-based travel experience in North Carolina and donate money or belong to nature/environmental organisations]			
Diamantis (1999)	Frequent 60%		Occasional 40%	
	[based on a sample of 1760 UK residents from databases of ecotourism-related tour operators and organisations]			
Weaver & Lawton (2002)	Harder 34%	Structured 40%	Softer 27%	
	[based on a sample of 1180 overnight patrons of two ecolodges in Lamington National Park, southeastern Qld]			
	Adapted from Weaver (2000)			

Adapted from Weaver (2008)

Hard-Soft ecotourists

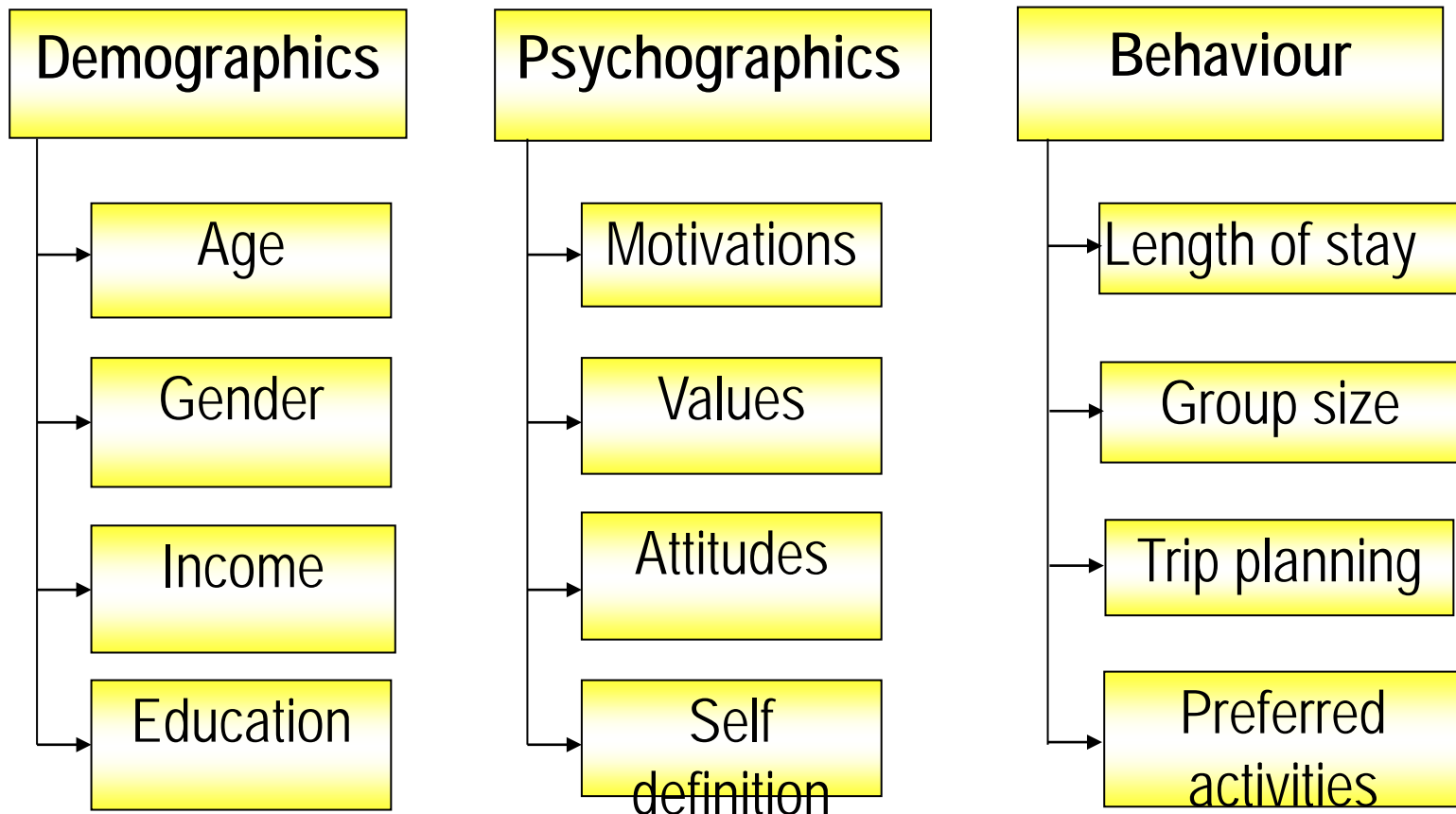
Motivational/behavioural segmentation



A research in Lamington N.P., Australia. Adapted from Weaver and Lawton (2002)

Hard-Soft ecotourists (cont.)

Variables for segmentation & comparison



Research problem

- Lack of research on core criteria and hard-soft typology of Asian and Vietnamese ecotourists
 - Lack of research on comparison between Western and Vietnamese ecotourists in the same site
- To what degree do Vietnamese & Western ecotourists in Viet Nam differ with regard to their ecotourism affiliation?

CA variables

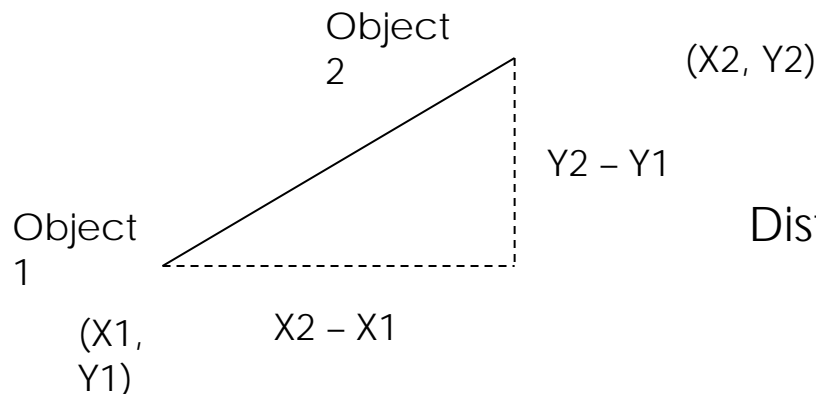
Categories	Number of items
Nature-based, Harder	3
Nature-based, Softer	3
Learning, Harder	3
Learning, Softer	3
Sustainability, Harder	3
Sustainability, Softer	3
Asian distinctiveness	6
Total	<u>24</u>

Stage 2 – Research design in CA

- Sample size
- Detecting outliers
- Measuring similarity/ distance
- Standardizing data

Research design

- Sample size: 500 VN – 500 Western visitors
- Stratified sampling, exit survey
- 1082 usable questionnaires (500 VN – 582 Western) after excluding outliers
- Cronbach alpha: 0.662
- Squared Euclidean distance measure
- No need for standardization because all clustering variables are measured on the same scale (1-5)



$$\text{Distance} = \sqrt{(X2 - X1)^2 + (Y2 - Y1)^2}$$

Stage 3 – Assumptions in CA

- CA is not a statistical inference technique but a method for quantifying the structural characteristics of a set of observations
- Strong mathematical properties but not statistical foundations
- No requirement on normality, linearity, homoscedasticity
- Two critical issues:
 - Representative of the sample
 - Impact of multicollinearity

Multicollinearity

- Impact of multicollinearity to clustering
- Is there any underlying factors among 22 variables?
- Factor analysis as a frequent used pre-processing technique prior to clustering (Dolnicar, 2002)
- Principal Component Analysis

Principal Component Analysis

- Each item corelates at least .30 with at least one other item
 - Barlett's test of sphericity is significant ($<.05$)
 - Kaiser-Meyer-Olkin (KMO) $\geq .60$
 - Measure of sampling adequacy in anti-image correlation matrix $\geq .50$
 - Toral variance explained $\geq .60\%$
- (Hair et al., 2010)

Principal Component Analysis

- All Eigenvalues ≥ 1
- All communalities for each item $\geq .40$
- Factor loading for each item $> .50$

(Hair et al., 2010)

- Cronbach's alpha at least .60 in exploratory research

(Nunnally, 1978)

Stage 4 – Deriving clusters and assessing overall fit

- Select the partitioning procedure used for forming clusters
- Make the decision on the number of clusters to be formed

Types of CA

- **Hierarchical clustering**

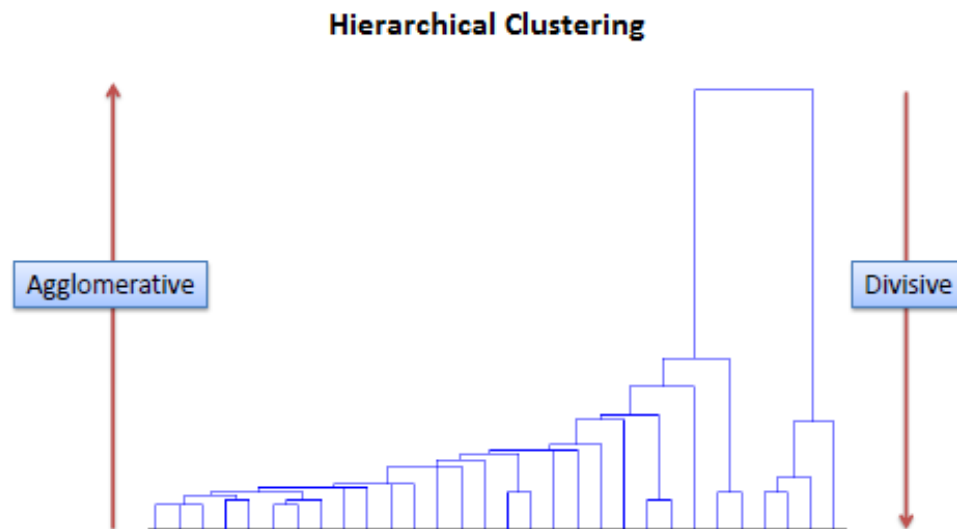
- No particular number of clusters identified in advance

- **Non-hierarchical (K-mean) clustering**

- Number of clusters is specified
- Assign cases into clusters

Hierarchical clustering

- Agglomerative methods
- Divisive methods



(Sayad, 2012)

Agglomerative methods

- Single-linkage
- Complete-linkage
- Average linkage
- Centroid method
- Ward's method

(Hair et al., 2010)

Deriving clusters

- Run the C.A in SPSS
- Examine each solution

Stage 5 – Interpretation of the clusters

- Run ANOVA to compare differences among groups
- Finalise cluster solution
- Name each cluster

6 cluster solution

- Sociable wildlife engagers (cluster 1, n=208 or 19%)
- Unenthusiastic visitors (cluster 2, n=239, or 22.2%)
- Typical visitors (cluster 3, n=283, or 26%)
- Service shunners (cluster 4, n=89, or 8%)
- Service seekers (cluster 5, n=92, or 8%)
- Classic Western ecotourists (cluster 6, n=171, or 16%)

Stage 6 – Validation and Profiling of the clusters

● Validation

- Run C.A in a separate sample
- Divide the sample in two half
- Cross-tabulation
- Examine motivation, attitude items that have relationship with behaviour items, but not included in C.A

● Profiling of clusters

- Discriminant Analysis
- Qualitative interviews

Conclusions

- C.A is the “**art**” of finding groups in data
- Theoretical background is important
- Random sampling
- Factorial Cluster Analysis
- Validation

References

- Blamey, R. (1997). Ecotourism: the search for an operational definition. *Journal of Sustainable tourism*, 5(2), 109-130.
- Ceballos-Lascuráin, H. (1996). *Tourism, Ecotourism and Protected Areas: The State of Naturebased Tourism and around the World and Guidelines for its Development*. Gland, Switzerland, and Cambridge, UK: IUCN.
- Cochrane, J. (2006). Indonesian national parks: Understanding leisure users. *Annals of Tourism Research*, 33(4), 979-997.
- Cochrane, J. (2007). Ecotourism and biodiversity conservation in Asia: institutional challenges and opportunities. In J. Higham (Ed.), *Critical Issue in Ecotourism: Understanding a Complex Tourism Phenomenon* (pp. 439). Oxford: Butterworth-Heinemann.
- Eagles, P., & Higgins, B. (1998). Ecotourism market and industry structure. In K. Lindberg & D. E. Hawkins (Eds.), *Ecotourism: a Guide for Planners and Managers* (pp. 11-43). North Bennington: The Ecotourism Society.
- SPSS (2010). *Segmentation and Profiling: Cluster and Factor Analysis*, IBM Inc.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Upper Saddle River: Pearson.
- Lilien, G. L., & Rangaswamy, A. (1998). *Marketing Engineering: Computer-Assisted Marketing Analysis and Planning*. New York: Addison-Wesley.
- Nunnally, J. C. (1978). *Psychometric theory* (2 ed.). New York: McGraw-Hill.
- Weaver, D. (2008). *Ecotourism* (2nd ed.). Milton: John Wiley & Sons.
- Weaver, D., & Lawton, L. (2002). Overnight ecotourist market segmentation in the Gold Coast hinterland of Australia. *Journal of Travel Research*, 40, 270-280.
- Weaver, D. (2002a). Asian ecotourism: patterns and themes. *Tourism Geographies*, 4(2), 153-172.
- Wight, P. A. (2001). Ecotourists: not a homogeneous market segment. In D. Weaver (Ed.), *The Encyclopedia of Ecotourism* (pp. 37-72). Wallingford: CABI.