Required Sample Sizes for Data-Driven Market Segmentation Analyses in Tourism

Journal of Travel Research
2014, Vol. 53(3) 296–306 ©
2013 SAGE Publications
Reprints and permissions:
sagepub.com/journalsPer
missions.nav DOI:
10.1177/004728751349647
5 jtr.sagepub.com

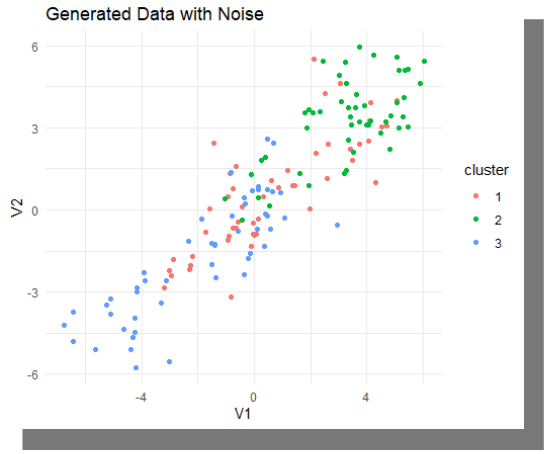
Sara Dolnicar1 , Bettina Grün2 ,
Friedrich Leisch3 , and Kathrin
Schmidt4

Keywords: market segmentation, cluster analysis, sample size, k-means, simulation study

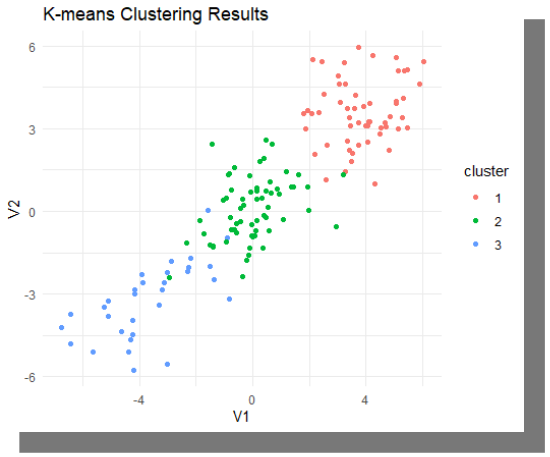
Exploring the issue of determining the appropriate sample size required for market segmentation using k-means clustering in the tourism sector:

- *Unsupervised learning and k-means clustering are widely used for market segmentation in tourism.*
- *The paper addresses a key issue: how large should a sample size be to ensure reliable clustering?*
- *The authors use simulation studies with artificial data to explore how different factors (number of variables, clusters, noise) affect clustering accuracy.*
- *The adjusted Rand index is used to measure the accuracy of the segmentation.*
- *The recommended sample size is 70 times the number of variables in the segmentation task.*
- *Larger sample sizes improve clustering results, especially for complex segmentation tasks with many variables or noise.*
- *This study provides practical guidelines for researchers and businesses in tourism to determine the required sample size for segmentation studies.*
- *The 70-times rule offers a conservative but practical approach to determine sample size.*
- *Future work could refine these guidelines for different clustering algorithms or data types.*

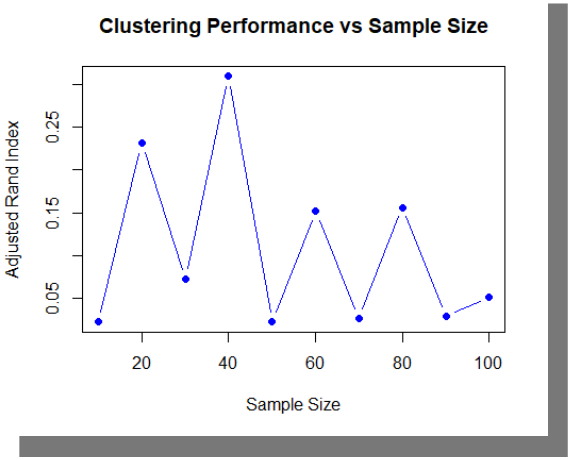
Stage 1: Data Generation



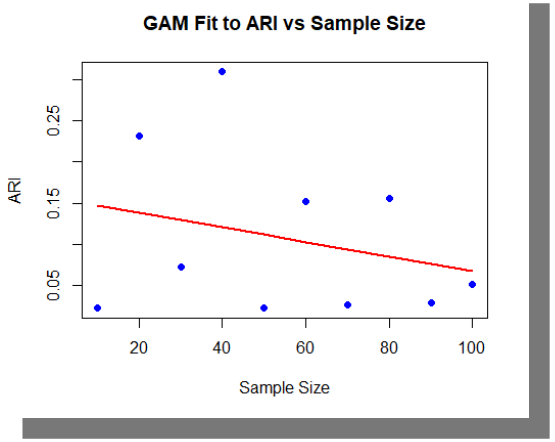
Stage 2: K-means Clustering



Stage 3: Performance Evaluation Using Adjusted Rand Index (ARI)



Stage 4: Generalized Additive Model (GAM) Analysis



Stage 5: Deriving Guidelines for Sample Size