SEA 2015: Using Cluster Analysis as an Evaluation Tool for Program Improvement

Kristy Kenney, MPH
Aundrea Carter, MS
Presentation Overview

- Introduction to Cluster Analysis
  - Overview of Methods
- Example of Cluster Analysis: National Breast and Cervical Cancer Early Detection Program (NBCCEDP)
- Evaluation Application: Using Cluster Analysis Results
- Next Steps
INTRODUCTION TO CLUSTER ANALYSIS
Cluster Analysis

![Cluster Analysis Diagram](image-url)
Cluster Analysis

Cluster analysis is a process often used in exploratory analysis to search for patterns in a given data set. The goal is to identify groups or clusters that are naturally related and more similar within a cluster and more distinct between clusters (Anderberg, 1973).
Cluster Analysis

- # of episodes
- # of days

Methods:
- Average Linkage
- Ward’s Minimum Variance
- Centroid Method
- K Means
EXAMPLE OF CLUSTER ANALYSIS: NATIONAL BREAST AND CERVICAL CANCER EARLY DETECTION PROGRAM (NBCCEDP)
National Breast and Cervical Cancer Early Detection Program (NBCCEDP)

http://www.cdc.gov/cancer/nbccedp/
NBCCEDP Strategic Direction
Incremental* transition to a program model using evidence-based strategies aimed at systems and policy change intended to reduce morbidity and mortality of breast and cervical cancers among all population subgroups with emphasis on disparate populations.

**INPUTS**
- Federal law 101-354
- Funding
- NBCCEDP policy
- Technical assistance, training, and consultation
- Evidence-based interventions
- Longitudinal data (registry, MDE, census, economic)
- National partnerships
- Support for program integration

**GRANTEE ACTIVITIES**
- Effective program management and leadership
- Strategic partnerships, coordination, and collaboration
- Screening provision to NBCCEDP eligible populations
- Population-based interventions and health systems change
- Public education and targeted outreach
- Screening, diagnostic, and patient navigation services
- Quality assurance and quality improvement
- Professional development
- Data management and utilization

**OUTCOMES**
- Greater awareness among all populations and increased intentions to be screened for breast and cervical cancer
- Policies and systems that promote high quality breast and cervical cancer screening
- Provider practices and systems change that support high quality breast and cervical cancer screening
- Surveillance systems to track screening rates and quality
- Reduced barriers and increased access to breast and cervical cancer screening

**IMPACT**
- Reduced breast and cervical cancer morbidity and mortality
- Reduced health disparities in breast and cervical cancer

**PROGRAM MONITORING AND EVALUATION**

*This logic model outlines a strategic direction for NBCCEDP over the five years of DP12-1205. Incremental but definitive annual progress in incorporating population-based screening promotion strategies and interventions as a component of the current screening program is expected of all grantees. Adjustments will be made given changes in the healthcare environment and authorizing legislation for the NBCCEDP, as needed.*
NBCCEDP Overview

- **Grantee Survey**
  - Web-based survey completed by all 67 grantees
  - First time systematically assessing NBCCEDP program implementation
- **Additional Goals:**
  - Assess implementation in the changing landscape of healthcare reform
  - Assess training and technical assistance needs
  - Establish a “baseline” for the 5 year funding period
The Guide to Community Preventive Services
Evidenced-Based Interventions (EBIs)
to increase cancer screening

Client Reminders

Small Media

Provider Assessment and Feedback

Reducing Structural Barriers

Provider Reminders

Client-Oriented EBIs

Provider-Oriented EBIs

http://www.thecommunityguide.org/
Determining factors

Community Guide  Evidenced-Based Interventions

- Client Reminders
- Reducing Structural Barriers
- Provider Assessment and Feedback
- Provider Reminders

Small Media

Client-Oriented EBIs
Provider-Oriented EBIs
How to determine the number of clusters

1. Fit Statistics
   - R-Square (RSQ)
   - Semi-partial R-Square (SPRSQ)
2. Dendrogram
3. K Means
4. Practical and Sensible Interpretation
1. Determining the Number of Clusters: RSQ

The diagram illustrates the determination of the number of clusters using the RSQ method. The graph shows the RSQ values for different numbers of clusters, with Ward's as the clustering method. The RSQ values increase as the number of clusters increases, reaching a peak at around 8 clusters and then leveling off. The RSQ values are marked for clusters 2, 3, 4, and 5, with values of 0.394, 0.681, 0.943, and 0.971, respectively. The RSQ value for 9 clusters is 0.989, indicating a high level of explained variance. For 11 clusters and beyond, the RSQ values remain constant at 1.0, suggesting that additional clusters do not significantly improve the explained variance.
2. Determining the Number of Clusters: Dendrogram
3. Determining the Number of Clusters: K Means

These are the means for each variable by cluster:

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Provider Reminders</th>
<th>Provider Assessment and Feedback</th>
<th>Client Reminders</th>
<th>Reducing Structural Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (N=14)</td>
<td>0</td>
<td>0</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>2 (N=40)</td>
<td>.75</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>3 (N=7)</td>
<td>.86</td>
<td>1.0</td>
<td>.43</td>
<td>.29</td>
</tr>
<tr>
<td>4 (N=6)</td>
<td>1.0</td>
<td>0</td>
<td>1.0</td>
<td>.83</td>
</tr>
</tbody>
</table>

a. Based on the variables being Yes or No (1 or 0)
4. Determining the Number of Clusters: Practical and Sensible Interpretation

- Cluster 1: “High Client EBI users”
- Cluster 2: “High Overall EBI users”
- Cluster 3: “High Provider/Low Client EBI users”
- Cluster 4: “High EBI users with no Provider Assessment and Feedback”
What’s the right number?

- “This fundamental step is among the as yet unsolved problems of cluster analysis” (Aldenderfer and Blashfield, 1984).

- Optimize the within and between cluster heterogeneity

- “When moving from fewer to more clusters does not result in new clusters that are meaningfully distinct from the ones they split from, then one has an appropriate number of clusters” (Peck, 2005)
EVALUATION APPLICATION: USING CLUSTER ANALYSIS RESULTS
Evaluation Application: Subgroup Analysis

- Describe General Characteristics (Descriptive)
- Compare Means (t-tests/ANOVA)
- Compare Proportions (Chi-Square)
- Correlations
- Model (use as predictor or outcome variables)
Evaluation Application: Longitudinal Analysis

- Dush and Keen studied alcoholics entering a treatment/rehab program.
  - For this study, they performed a cluster analysis longitudinally to see whether the MMPI subtypes were stable over time: before treatment and 30 days after treatment.
  - Their findings showed that treatment programs could modify their programs as the alcoholics progress through the treatment.
Evaluation Application: Program Impact on Treatment Groups

- Peck (2005) described the diverse welfare population in a study about New York State’s welfare reform initiative:
  - Clustered individuals using demographic information
  - Compared baseline characteristics and outcomes/impacts across clusters
- Identified “different impacts for different population segments”
Cluster Entire Sample

Cluster Grouping Among Entire Sample (treatment and control members)

Assign to Treatment or Control Group for Each Cluster

Compare Cluster Effect Across All Clusters

Compare Cluster Treatment to Cluster Control Group

Figure 1: How Cluster Analysis Creates Experimentally Comparable Subgroups
Next Steps

- Future Program Planning/Improvement
  - Evaluation Use By Cluster
  - Technical Assistance/Training Need by Cluster
  - EBI use with Non-NBCCEDP clients
    - Understand changes in era of the Affordable Care Act
  - Changes in Clusters and Cluster Assignments over Time
    - With Year 3-5 Survey Data
THANK YOU!

Kristy Kenney, MPH
hsl7@cdc.gov
Aundrea Carter, MS
yma7@cdc.gov

For more information please contact Centers for Disease Control and Prevention
1600 Clifton Road NE, Atlanta, GA 30333
Telephone, 1-800-CDC-INFO (232-4636)/TTY: 1-888-232-6348
E-mail: cdcinfo@cdc.gov  Web: www.cdc.gov

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.
Dr. Mel Mark for recommending cluster analysis as a tool at The Evaluators Institute (TEI).
References