Using Factor Analysis and Cronbach's Alpha To Ascertain Relationships Between Questions of a Dietary Behavior Questionnaire

Eric Grau

MATHEMATICA
Policy Research, Inc.
OUTLINE

- Study overview
- Review of Methods
- Description of Analyses
- Results
- Summary
Study Overview: Goal

- **Objective:** Review and revise a draft questionnaire on dietary behavior that was developed by another contractor
  - Review content of questions
  - Review order of questions
  - Review whether questions should be dropped

- **Goal:** Develop a questionnaire (based on draft) that can assess respondents’ adherence to the Dietary Guidelines for Americans
Study Overview: Stages

- Our focus: assessing results of field test with respect to relationships between items
Study Overview: Field test

- Field Test
  - Size of test: 453 white, African-American, and Hispanic women food stamp recipients
  - Time limit: core questions in the instrument should take less than 15 minutes to administer
  - Resolve: Level of redundancy within topical modules
Study Overview: Questionnaire Organization

- Questionnaire organized into modules
  - Dietary modules: recording weekly consumption of various food groups
  - Attitude and behavior modules: questions about attitudes and behaviors related to food and nutrition
Review of Methods: Factor Analysis

Definition

- Describe a set of $p$ random variables in terms of a smaller number of unobserved random variables called FACTORS.

- Factors are determined by interpreting coefficients in factor model called LOADINGS.

- Orthogonal transformation of factor loadings, called a ROTATION, allows for easier interpretation of factors.
Review of Methods: Factor Analysis

- Variance of X can be decomposed into common variance + specific variance
- Data Assumed to be Multivariate Normal
- Methods of Estimation of Factor Loadings
  - Principle Component Analysis
  - Principle Factor Analysis
  - Maximum Likelihood Estimation
Review of Methods: Factor Analysis

Goodness-of-fit Measures

- MSA (measure of sampling adequacy): partial correlations between each pair of variables controlling for all other variables
  - indicates how well variables fit in factor model

- Kaiser’s MSA is overall measure:
  - Values below 0.5 are unacceptable
  - Values above 0.8 indicate factor model fits well
Review of Methods: Cronbach’s Alpha

- A correlation coefficient that describes how well a group of items focuses on a single idea or construct
- Establishes consistency of questions asked in different ways about a single attribute
- High levels indicate
  - relative absence of item error variance
  - Items contribute to a reliable scale for one attribute
Review of Methods: Cronbach’s Alpha

- Two calculations:
  - Raw alpha (based on correlations)
  - Standardized alpha (based on covariances)

Rule of thumb:

>= 0.70 considered acceptable

What does a lower level mean?
Description of Analyses

- Separate analyses were done within dietary and behavior/attitude topics

- Methods of estimation used for Factor Analysis:
  - Principle Factor Analysis
  - Maximum Likelihood

- Factors rotated using VARIMAX rotation
Description of Analyses: Data Issues

- Weekly consumption data positively skewed
  - Analyze data on square root scale
- Some variables have only four categories
  - Other methods may be more appropriate
  - Ordinal response problematic: assumes equal distance between levels
## Results: Fruit and Vegetables (with Fries)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>FACTOR 1</th>
<th>FACTOR 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit</td>
<td>0.58</td>
<td>0.08</td>
</tr>
<tr>
<td>Unsweetened Juice</td>
<td>0.32</td>
<td>0.09</td>
</tr>
<tr>
<td>All vegetable</td>
<td>0.68</td>
<td>0.26</td>
</tr>
<tr>
<td>Potatoes (not fries)</td>
<td>0.39</td>
<td>-0.26</td>
</tr>
<tr>
<td>French fries</td>
<td>-0.05</td>
<td>-0.29</td>
</tr>
<tr>
<td>Dark green vegetables</td>
<td>0.59</td>
<td>0.41</td>
</tr>
<tr>
<td>Orange vegetables</td>
<td>0.56</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Results: Fruit and Vegetables (with Fries)

- First factor: true fruits and vegetables
- Second factor: distinguishes between green vegetables and potatoes

- 89% of common variance explained by first two eigenvalues

- Kaiser’s Overall MSA = 0.74
Results: Fruit and Vegetables (with Fries)

- Cronbach’s alpha = 0.58

- What does this mean?
  - High level of error variance for items to be considered reliable for single construct scale
  - Not a single construct?
## Results: Fruit and Vegetables (no Fries)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>FACTOR 1</th>
<th>FACTOR 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit</td>
<td>0.35</td>
<td>0.51</td>
</tr>
<tr>
<td>Unsweetened Juice</td>
<td>0.29</td>
<td>0.17</td>
</tr>
<tr>
<td>All vegetable</td>
<td>0.65</td>
<td>0.34</td>
</tr>
<tr>
<td>Potatoes (not fries)</td>
<td>0.09</td>
<td>0.42</td>
</tr>
<tr>
<td>Dark green vegetables</td>
<td>0.73</td>
<td>0.13</td>
</tr>
<tr>
<td>Orange vegetables</td>
<td>0.39</td>
<td>0.39</td>
</tr>
</tbody>
</table>
Results: Fruit and Vegetables (no Fries)

- First factor: distinguishes type of vegetable
- Second factor: not clear

- 88% of common variance explained by first two eigenvalues (78% by first eigenvalue)

- Kaiser’s Overall MSA = 0.74
Results: Fruit and Vegetables (no Fries)

- Cronbach’s alpha = 0.68

- What does this mean?
  - Removing single item that “didn’t fit” improved alpha markedly
Results: Weight Consciousness

- Nine variables:
  - Five refer to food consumption
  - Four refer to behaviors

- Problems:
  - Some ordinal responses
  - Two behavioral variables are binary
  - Some levels needed collapsing
  - Kaiser’s Overall MSA = 0.56
## Results: Weight Consciousness

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>FAC. 1</th>
<th>FAC. 2</th>
<th>FAC. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit as dessert</td>
<td>0.03</td>
<td>0.09</td>
<td>0.48</td>
</tr>
<tr>
<td>Fruit/vegetables as snacks</td>
<td>0.10</td>
<td>0.12</td>
<td>0.46</td>
</tr>
<tr>
<td>Sweetened fruit drinks</td>
<td>0.11</td>
<td>0.26</td>
<td>0.15</td>
</tr>
<tr>
<td>Soda</td>
<td>0.07</td>
<td>0.43</td>
<td>0.16</td>
</tr>
<tr>
<td>Fast food</td>
<td>0.02</td>
<td>0.45</td>
<td>0.08</td>
</tr>
<tr>
<td>Switched to healthier diet</td>
<td>0.69</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>Attempted to lose weight</td>
<td>0.71</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td>Snack or eat meals at TV</td>
<td>0.05</td>
<td>0.35</td>
<td>-0.11</td>
</tr>
<tr>
<td>Eat breakfast in morning</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.30</td>
</tr>
</tbody>
</table>
Results: Weight Consciousness

- **First factor:** actions to improve health
- **Second factor:** unhealthy eating habits
- **Third factor:** healthy eating habits

- **Two clusters across the three factors:**
  - Switched to healthier diet/attemted to lose weight
  - Fruit as dessert/fruit or vegetable as snack

- All of common variance explained by first three eigenvalues
Results: Weight Consciousness

- Cronbach’s alpha = 0.48
  - High level of error variance for items to be considered reliable for single construct scale
  - Not a single construct?
Summary

- Redundancies
  - Eating fish and eating dry beans
  - Behaviors: switched to healthier diet and attempted to lose weight
  - Eating fruit as dessert and eating fruit or vegetables as snacks

- In many cases, alpha may not be an appropriate measure, given the number of underlying factors is greater than one.

- Final recommendation for ERS: Remove one of the fish and dry bean questions
Acknowledgements

- This work was done as part of a contract that MPR had with the U.S. Department of Agriculture, Economic Research Service

- Project: Development of a Questionnaire on Dietary Behavior for Use in Low-Income Populations (MPR project number 6191)
  - Project Officer: David Smallwood, ERS
  - MPR Project Director: Rhoda Cohen
Review of Methods: Factor Model

Factor Model:

\[ X_i = a_{i1}F_1 + a_{i2}F_2 + \ldots + a_{im}F_m + \varepsilon_i; \quad i = 1, 2, \ldots, p \]

- \( X_i \) = ith variable, centered with mean 0 variance 1
- \( \varepsilon_i \) = ith error (specific factor)
- \( a_{ij} \) = jth factor loading for \( X_i \)
- \( F_j \) = uncorrelated common factors with unit variance
Review of Methods: Factor Analysis
Rotations

- Factor Model
  - Orthogonal transformation of factor loadings corresponds to a rotation of the coordinate axes
  - Communalities and specific variances remain unchanged
  - Original loadings may not be readily interpretable—rotate until simple structure is achieved
Rotations

Factor Model VARIMAX Rotation:

- Maximize $V$, which is proportional to
  $$\sum_j \text{Var}(a_{ij}^2)$$
- Spreads out the squares of the loadings on each factor
- Forces large and negligible coefficients in any column of the rotated loadings matrix (i.e., associated with each factor)
## Results: High Protein Foods

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>FACTOR 1</th>
<th>FACTOR 2</th>
<th>FACTOR 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poultry</td>
<td>0.09</td>
<td>0.51</td>
<td>-0.07</td>
</tr>
<tr>
<td>Red meat/pork</td>
<td>0.65</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>Deli meats</td>
<td>0.62</td>
<td>0.22</td>
<td>0.07</td>
</tr>
<tr>
<td>Fish</td>
<td>0.09</td>
<td>0.47</td>
<td>0.24</td>
</tr>
<tr>
<td>Eggs</td>
<td>0.24</td>
<td>0.12</td>
<td>0.29</td>
</tr>
<tr>
<td>Peanut butter</td>
<td>0.03</td>
<td>0.04</td>
<td>0.46</td>
</tr>
<tr>
<td>Dry beans</td>
<td>0.10</td>
<td>0.41</td>
<td>0.28</td>
</tr>
</tbody>
</table>
Results: High Protein Foods

- First factor: measure of less healthy proteins
- Second factor: healthier proteins
- Third factor: peanut butter

- Beans and fish cluster together across all three factors
- All of common variance explained by first three eigenvalues
- Kaiser’s Overall MSA = 0.63
Results: High Protein Foods

- Cronbach’s alpha = 0.56
  - High level of error variance for items to be considered reliable for single construct scale
  - Not a single construct?