Demonstration on Dietary Pattern Study Using Sparse Latent Model and PCA

Content

- Motivation
- Expectation
- Difficulties
- Solutions
- Results

Background & Motivations

- On-going research project collaborated with Nutritionists on exploring dietary patterns of senior NZ citizens and searching for healthier patterns that fits the population.

- Conducted two surveys with 367, 319 participants in a month apart. The weights of food intakes in grams per day and the frequency of the food intakes per day are collected.

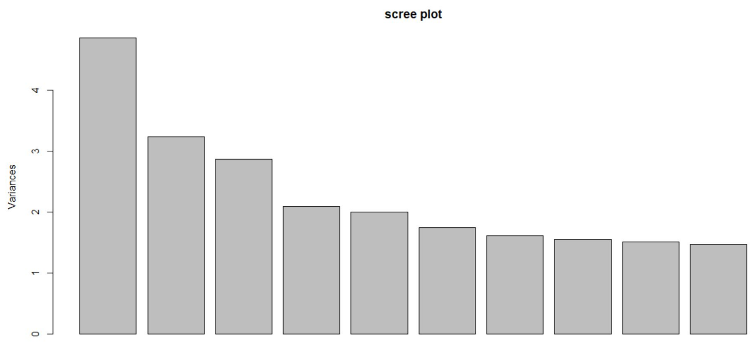
Background & Motivations

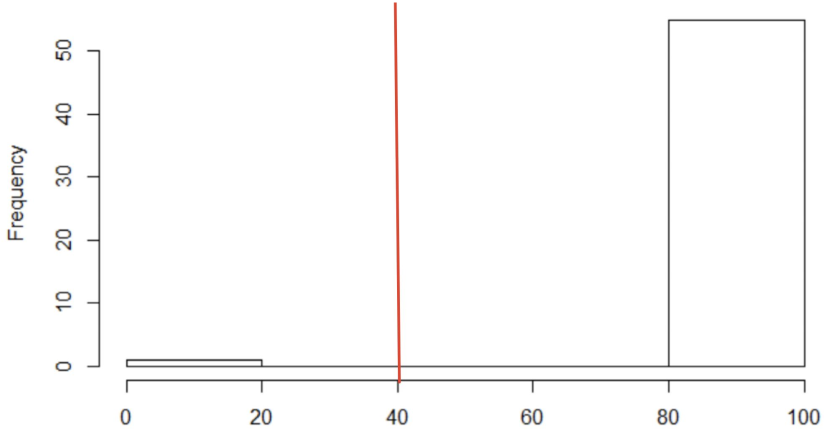
- PCA is very popularly used for dietary pattern studies because of the variety of different items (variables) and the **intuition** of PCA on creating arbitrary components represented by linear combinations of food items.
- PCA has its own **drawbacks** for including weakly associated food items, or cross load food items in different dietary patterns.
- We want to find a better way to **identify the association** between food items and dietary patterns **more precisely**.

Bayesian Latent Model with 5 Factors over FFQ1 Data	Food Groups
Factor 1: Dietary Pattern 1 (*healthy*)	Wholegrain; Alliums; Alternate; Fresh.frozen.Legumes; Refined grain; Root.starchy.vegetable

Goals & Expectations

- Find more dietary patterns, or more precisely defined dietary patterns
- Use **sparse latent model** vs PCA because the sparse latent model **shrinks** the weak association food items to 0 and eliminate them from the group.
- Expect sparse latent model to identify patterns and food combinations in a **more precise** way with lower cross loadings than PCA





Difficulties

- Data simulation wise
  - **Limited data** with a lot more variables (food items)
  - **Hard to assess the normality** of the data
- Modelling wise
  - What model to use
  - Model tuning - so the model identifies the right pattern without including irrelevant food groups

Strategy

Divide & Conquer

- **Divide** the problem into subproblems
  - E.g. list out the options for solve the problem of sparse data
  - Then narrow down the option that also solves non-normality
- **Rejoin** the solutions
  - Combining solutions without losing information

Solutions

- Data simulation wise
  - The use of Bayesian styled Sparse Latent Model
    - Allowing non-normality assumptions
  - MCMC simulation
    - Overcome the higher dimensionality than sample size
- Modelling wise
  - Bayesian styled Sparse Latent Model **allows hierarchical model** to fit age, race, gender on the level above food items
    - give more space for continuous research
  - Tuning the model by setting the correct sparsity on MCMC convergence
    - The model identifies the right pattern without including irrelevant food groups

Results

- Both PCA and Sparse Latent Model are **successfully applied** to our sample data
- **Two Dietary Patterns** are recognized by both methods
  - Western
  - Healthy
- Comparisons
  - Sparse Latent Model is able to identify dietary patterns with **less cross loading food items**
    - Each group of food items are more obvious categorized as one dietary pattern than PCA
  - Sparse Latent Model worked better on Non-linear data transformation
    - It is able to recognized dietary patterns for **both frequency and gram scaled** data
  - Sparse Latent Model might give outputs that is not as interpretable when fewer food items are included
    - I.e. only one food item in a group by itself

The table on the left is the result of PCA and the one on the right is of Sparse Latent Model

Two Rotated Components with FFQ1 Data	Food Groups
Dietary Pattern 1 (*healthy*)	salad veg; other veg; green leafy cruciferous; alliums; cruciferous; carrots; dried legumes; nuts and seeds; water; berry fruits; all other fruit; olives and avocados; alternate; root, starchy veg; fresh or frozen legumes; oily fish; wholegrain; white fish; spices, tomatoes, stone fruit
Dietary Pattern 2 (*western*)	sauces, chutneys; processed meats; dressings; biscuits, cakes and pastries; diet drinks; savoury; tomatoes; confectionery; chocolate; stone fruit; processed fish; cheese and creamy diary; beer

Table 3.1 *PCA outputs with two principal components on FFQ1 data*

Dietary Patterns Identified with FFQ1 Data	Food Items
Dietary Pattern 1 (*healthy*)	Berries; Dried.fruit; Apples.pears; Olives; Pickles; Coconut.oil; Vegetable.oils Marmite.vegemite; Brown rice; Non.milk.based.puddings; Carrots; Grain; Herbal.tea; Other.root.vegetable; Salad.vegetables; Tomatoes; Onions.leeks.garlic; Sausages; Seeds; Kumara.taro
Dietary Pattern 2 (*western*)	Hot.potato.chips.French.fries.wedges; Biscuits; Cakes; Port.sherry.liquors; Margarine; Coconut.cream; Creamy dressings; White.sauce; Sweets; Soft.fizzy.drinks

Table 4.1 *Sparse latent 5 factors model outputs with two dietary patterns identified on FFQ1 frequency scaled data*

Stone fruit is included in the both patterns for PCA but it is eliminated by Sparse Latent Model.