AM-Week2 Case Study2

Victoria Ruan

Who are the obese? A cluster analysis exploring subgroups of the obese

- Journal of Public Health
- Oxford Academic

Introduction

Limitations of BMI

1. such a distinction fails to account for the **variation within this group** across other factors such as health, demographic and behavioral characteristics.

Methods

1. Individuals with a BMI of ≥30 were included.
2. **A two-step cluster analysis** was used to define groups of individuals who shared common characteristics.

Data Source

Taken from Yorkshire Health Study (2010–12)

- a longitudinal observational study
- self-reported
- total 27806 individuals
- 4144 were classified as having a BMI ≥30.

Demographic Variables

- **age**
- **sex**
- **ethnicity**
  - ‘White’ or ‘Non-White’
- **socioeconomic deprivation**
  - the area individuals lived in
  - multidimensional measure

Health-related variables

- **whether an individual reported**
  - fatigue
  - pain
  - insomnia
  - anxiety
  - depression
  - diabetes
  - breathing problems
  - high blood pressure
  - heart disease
  - osteoarthritis
  - stroke
  - cancer
- **EuroQoL EQ5D**
  - a measure of an individual's **health-related quality of life**
- **Well-being**
  - asking individuals **how satisfied** they were of their life
  - from 0 (completely dissatisfied) to 10 (completely satisfied).
- **Behavioral Characteristics**
  - whether **smoking**
  - amount of **alcohol** consumed in the previous week
  - **sedentary** characteristics:
    - choose the lower of the following:
      - whether engaged in >1 h of **physical activity** a week
      - whether an individual **walked** for >1 h in a week
  - whether engaged in **active weight management:**
    - slimming clubs
    - increasing exercise
    - controlling portion size

- eating healthier
- using over-the-counter weight loss medication
- using meal replacements

Analysis

Tool:

- SPSS

A two-step cluster analysis

- exploratory and hypothesis generating
- cannot identify causation
- can be used to drive future research
- The data included **both binary and continuous variables**

Binary:

1. Scanning the data in a pre-classificatory stage
2. identifying cluster features
    - the ‘dense’ regions of data
    - data points that **share similar values** across a range of variables
3. use **agglomerative hierarchical clustering** method
    - classify data
4. use **log-likelihood**
    - as a distance measure
    - normalizes distance between different data types

Continuous:

1. standardized using ***z* -scores**
    - allow for **greater comparability** between the different scales

Clustering:

Prerequisites:

- The number of clusters needs to be **large enough**

    - capture the important features in the data
- but **not too large**
    - interpretation becomes difficult
- Use **Bayesian Information Criterion** (BIC)
    - best represents the underlying structure of the data

Interpretation:

- calculate the **mean values** of the variables for each cluster
- calculate the **coefficient of variation**
    - a **normalized measure of the variation** in variables
    - help assess **contribution** to cluster formation

Stability:

- a **replication analysis** is conducted
- use **Blashfield and Macintyre's split sample**
    - randomly divides the sample into half
    - performs the cluster analysis
        - using the same rules and parameters from the main cluster analysis on each sample
    - use Cohen's kappa coefficient
        - measure the agreement between two sub groups' equivalent clusters

Results

Clusters

there are two obvious ❓ **kinks** in the plot

- suggest that a **six-cluster** solution offers greater **discriminatory** power
    - capture further variation

Table 1

- Description of the demographic factors (%) of individuals whose body mass index (BMI) was ≥30

Variable	*Obese sample (BMI ≥30)*
Gender	
Female	57.6
Male	42.4
Age	
≤24	4.9
25–34	7.7
35–44	11.6
45–54	16.8
55–64	23.7
65–74	23.3
≥75	11.9
Deprivation quintile	
1 (Least deprived)	8.9
2	19.5
3	16.1
4	20.9
5 (Most deprived)	34.6
Ethnicity	
White	95.2
Non-White	4.8

Table 2

The mean values of variables split by clusters

Variable	***Clusters***						**All individuals**	***Coefficient of variation***
	Physically sick but happy	*Affluent healthy*	*Younger healthy*	*Unhappy anxious middle*	*Heavy drinking*	*Poorest*		

	elderly	*elderly*	*females*	*aged*	*males*	*health*		
Sample size	794	555	1021	577	887	310	4144	
Mean body mass index	34.41	33.68	34.06	34.32	32.98	36.49	34.07	0.03
Mean age	67	62	49	52	52	62	56	0.13
Proportion male	0.48	0.53	0.00	0.27	1.00	0.56	0.46	0.72
Proportion non-White	0.01	0.03	0.03	0.03	0.03	0.02	0.03	0.28
Mean deprivation score	27.07	23.78	24.38	27.48	24.37	33.94	25.96	0.15
Mean life satisfaction score	7.45	7.99	7.55	5.62	7.6	4.76	7.12	0.18
Mean EQ5D	0.60	0.87	0.88	0.59	0.87	0.21	0.73	0.36
Proportion with fatigue	0.40	0.03	0.02	0.70	0.04	0.82	0.25	1.44
Proportion with pain	0.76	0.03	0.07	0.58	0.09	0.91	0.33	1.18
Proportion with insomnia	0.08	0.01	0.00	0.32	0.01	0.36	0.09	1.84
Proportion with anxiety	0.03	0.03	0.01	0.56	0.01	0.58	0.13	2.19
Proportion with depression	0.02	0.03	0.02	0.46	0.01	0.69	0.13	2.28
Proportion with diabetes	0.32	0.18	0.04	0.04	0.08	0.38	0.15	0.98
Proportion with breathing problems	0.27	0.07	0.07	0.15	0.06	0.47	0.15	1.08
Proportion with high blood pressure	0.62	0.99	0.00	0.15	0.02	0.70	0.33	1.25
Proportion with heart disease	0.23	0.04	0.02	0.01	0.04	0.36	0.09	1.61
Proportion with osteoarthritis	0.38	0.08	0.03	0.11	0.03	0.44	0.15	1.22
Proportion with stroke	0.04	0.01	0.00	0.02	0.01	0.13	0.02	2.42
Proportion with cancer	0.07	0.03	0.01	0.03	0.01	0.05	0.03	0.78
Proportion who smoke	0.08	0.06	0.12	0.16	0.13	0.21	0.12	0.45
Mean alcohol intake (units/week)	5.31	8.03	4.98	4.85	11.86	6.57	7.03	0.38
Proportion who walk >1 h/week	0.26	0.46	0.44	0.36	0.43	0.08	0.37	0.40
Proportion who do								

physical exercise >1 h/week	0.31	0.49	0.51	0.40	0.48	0.12	0.42	0.36
Proportion who actively manage their weight								

The coefficient of variation

- Variables with **greater variation** will be **more important** in cluster formation
- highest among the health-related variables
    - stroke
    - anxiety
    - depression

Replication:

- Blashfield and Macintyre's split sample method
    - clusters that were fairly similar
- Cohen's kappa coefficient
    - 0.41 (*P* < 0.001)
    - suggesting **moderate agreement**
    - cases that altered were mostly found on the **boundaries** of each cluster
- The clusters remained **consistent**
    - if the **morbidly obese were removed** from the sample
-

Limitations

- BMI may **not always accurately classify individuals as obese**
    - does not directly measure **body fat**
    - **underestimated** prevalence of obesity compared with body fat
- Bias
    - self-reported information
- **cannot generaliz**e to other population
    - Cluster analysis is a data-driven method
-

Conclusions

1. It is important to account for the important **heterogeneity** within individuals who are obese.

- A focus on subgroups of individuals may allow a much **more efficient** targeting of **scarce healthcare and health promotion resources**.
- weight loss may not be the primary clinical focus for different groups
- Interventions introduced by clinicians and policymakers should **not target obese individuals as a whole** but **tailor** strategies depending upon the subgroups that individuals belong .