

Sampling Theory

Exam Project

Alexander James Ryan r0693898 Adityavarna Dehaleesan r0690673

14/6/2018

Introduction

Part One - Critical Reading (perhaps we can split into two parts?)

Introduction

Part Two - Modelling BMI

Introduction

Part Three - Design of Health Interview Survey

Introduction

The choice on where to stratify and group populations is difficult. The goal is to create a sample that is representative of different levels of the population, so that at each level of analysis, we can form accurate estimands. Should the focus be on weighting regions, then perhaps some large cities in smaller regions become under-represented, and small cities in large regions become over-represented. Ultimately the strategy in which you choose to stratify a population, will over-represent some communities, and under-represent others, an implicit bias that is impossible to evade. The goal of sampling populations is to reduce this imbalance as much as possible, so every community is given a fairly weighted representation.

In this paper, we will discuss a variety of stratifyign strategies, to illustrate the complexity and variation in choosing sampling priorities. The first division is in the three different regions in Belgium, Wallonia, Brussels and Flanders.

Table 1: Proportioned weights at Region level

Region	Weight
Wallonia	35%
Brussels	30%
Flanders	35%

Stratifying Strategy 1

Each Region weighting is devided equally between Provinces and Arrondissements

Of the 35% allocated to Flemish Region, each of the different Provinces receives 7%. This 7% is then proportionally allocated to each Care Sector based on their respective populations. The same strategy is then used for the Wallonia Region, but first the set allocation of 3% to the German community is taken from

35%, leaving 32% to be allocated to the Walloon Provinces. Thus each Walloon Province receives 6.4%, which is then proportionally allocated to each Arrondissement based on their respective populations.

To illustrate:

Table 2: Stratifying Strategy 1

Province	Care sector/Arrondissement	Flemish Region	Brussels	Walloon Region
West-Vlaanderen	Brugge	1,87%	-	-
-	Oostende	1,12%	-	-
-	Roeselare	2,13%	-	-
-	Kotrijk	1,88%	-	-
Oost-Vlaanderen	Gent	4,23%	-	-
-	Sint-Niklaas	1,14%	-	-
-	Aalst	1,63%	-	-
Vlaams-Brabant	West	2,90%	-	-
-	Leuven	4,10%	-	-
Antwerpen	Antwerpen	3,68%	-	-
-	Mechelen	1,59%	-	-
-	Turhout	1,74%	-	-
Limburg	Hasselt	4,93%	-	-
-	Genk	2,07%	-	-
Brussels	-	-	30%	-
Hainaut	Tournai	-	-	1,48%
-	Charleroi	-	-	2,06%
-	Thuin	-	-	0,73%
-	Mons	-	-	1,23%
-	Soignies	-	-	0,90%
Liège	Huy	-	-	1,20%
-	Liège	-	-	3,89%
-	Verviers	-	-	1,31%
-	German Community	-	-	3%
Luxembourg	Arlon	-	-	3,75%
-	Neufchateau	-	-	2,65%
Namur	Namur	-	-	4,09%
-	Dinant	-	-	2,31%
Brabant-Wallon	Nivelles	-	-	6,40%

Stratifying Strategy 2

The 35% allocated to the Flemish Region is proportionally allocated to each Care Sector, based the proportion of their populations in respect to the total population of the Province. The same strategy is used for the Wallonia region, and again, removing the 3% allocated to the German Community, leaving 32% to be proportionally allocated to the Arrondissements.

Table 3: Stratifying Strategy 2

Province	Care sector/Arrondissement	Flemish Region	Brussels	Walloon Region
West-Vlaanderen	Brugge	1,80%	-	-
-	Oostende	1,08%	-	-
-	Roeselare	2,04%	-	-
-	Kotrijk	1,80%	-	-
Oost-Vlaanderen	Gent	5,07%	-	-

Province	Care sector/Arrondissement	Flemish Region	Brussels	Walloon Region
-	Sint-Niklaas	1,36%	-	-
-	Aalst	1,95%	-	-
Vlaams-Brabant	West	1,95%	-	-
-	Leuven	2,77%	-	-
Antwerpen	Antwerpen	5,41%	-	-
-	Mechelen	2,33%	-	-
-	Turhout	2,55%	-	-
Limburg	Hasselt	3,45%	-	-
-	Genk	1,44%	-	-
Brussels	-	-	30%	-
Hainaut	Tournai	-	-	2,81%
-	Charleroi	-	-	3,93%
-	Thuin	-	-	1,38%
-	Mons	-	-	2,35%
-	Soignies	-	-	1,71%
Liège	Huy	-	-	1,73%
-	Liège	-	-	5,65%
-	Verviers	-	-	1,90%
-	German Community	-	-	3%
Luxembourg	Arlon	-	-	1,48%
-	Neufchateau	-	-	1,05%
Namur	Namur	-	-	2,84%
-	Dinant	-	-	1,60%
Brabant-Wallon	Nivelles	-	-	3,57%

Wallonia region

Given that we have to exclude 300 of our total 3500 sample in the Wallonia region for the Germany community, it is important how we allocate the remaining 3200 amongst the remaining communities. Depending on how we allocate, we could bias the results, over representing some regions, whilst under-representing others.

For example, if we simply divided the 3500 total allocated for the Wallonia region, and divided by 5, we would be sampling 700 people from each region. Since for the Liege province, we have set aside 300 of that for the German community, we only have 400 left for the Arrondissements Huy (pop=189,661), Liege (pop=617,551), Verviers (pop=208,249). Liege is the largest Arrondissement in the Wallonia region, and we would find it severely under-represented if we use this stratifying strategy.

Part Four - Missing Data in the Belgian Health Interview Survey

Introduction