Efficient designs for mean estimation in multilevel populations and test norming

Chapter 8

Scientific and social impact of the thesis

A crucial step in the research process is the choice of the design of the study, because a poorly designed study can have serious consequences for science (e.g. biased or unreliable results) and society (e.g. a waste of resources or bad decisions in health and education based on invalid research conclusions). This thesis deals with the design of two types of studies, that is, surveys for mean estimation in multilevel populations (e.g. students grouped in schools, patients clustered in hospitals), and normative studies for estimating reference values (or norms) for tests (e.g. IQ test) and questionnaires (e.g. to measure patients' symptoms). Both types of studies are of practical importance. By allowing comparisons between different populations with respect to their means (e.g. comparing countries in terms of average length of stay for discharges from hospitals), surveys for mean estimation can be useful, for instance, for the implementation of new governmental policies (e.g. new interventions to reduce length of stay in public hospitals). Normative studies, instead, provide reference values that clinicians and educators need in order to compare individuals' performance on a test with the reference population (e.g. individuals with the same age, sex, and education), and to make decisions about individuals (e.g. assignment of a patient to a treatment or of a student to remedial teaching). Thus, it is important that population means and reference values are estimated with the highest possible precision, and without wasting resources (i.e. time and money). This goal can be attained by a careful design of the study. Hence, the main objective of this thesis is to provide guidelines for planning both types of studies in order to achieve precise estimates using minimum resources.

This thesis addresses three design issues of surveys for mean estimation in multilevel populations. First, it identifies the best strategy to draw a sample from the population (i.e. the most efficient sampling scheme) and to analyse the data (i.e. the unbiased mean estimation). Specifically, data should be collected by sampling clusters (e.g. schools, hospitals) with probability proportional to their size (i.e. number of individuals belonging to a cluster) first, and then by sampling the same number of individuals per selected cluster. Furthermore, the average of all individual outcomes in the population (i.e. the population mean) should be estimated by computing the average of the means of the sampled clusters. Second, this thesis provides formulas to compute optimal sample sizes (i.e. number of clusters and number of individuals per cluster) that allow to either maximize precision of mean estimation for the available budget for sampling and measuring, or to minimize the budget for the required precision of estimation, thus avoiding a waste of the limited resources. Third, a strategy is proposed to overcome the dependency of the optimal sample sizes on some unknown model parameters. This strategy consists of defining a range of plausible values for each unknown parameter first, and then deriving the sample sizes (i.e. number of clusters and number of individuals per cluster) that maximize precision of estimation under the worst-case scenario that can occur within these parameter ranges. These results have a direct impact on science by extending the results available in survey sampling literature (Chambers & Clark, 2012; Cochran, 1977; Lohr, 2010; Särndal et al, 1992; Sukhatme, 1954; Valliant et al, 2000) to a scenario where cluster size affects individuals' outcome variables (e.g. when the length of stay of a patient in a hospital depends on the number of patients admitted to the hospital), and have an indirect impact on society by helping researchers in planning surveys for monitoring important social issues, such as alcohol consumption among adolescents or government expenditure on health, without wasting resources (i.e. time and money).

Three design issues of normative studies are addressed in this thesis. First, it provides the sample composition (e.g. which age groups to include) that maximizes precision of the estimation of reference values (i.e. the optimal design) under five regression models for the reference population. Second, since a design that is optimal under one regression model can be very inefficient under another regression model, and at the design phase of a study there is uncertainty about the "true" model, highly (but not maximally) efficient designs that are robust against the choice of the wrong model in the design stage are presented. Third, this thesis provides formulas to determine how many individuals must be sampled under the optimal design in order to achieve a desired statistical power for testing hypotheses on an individual's performance, or to achieve a desired margin of error in estimating an individual's performance. These results have a direct impact on science because they extend previous research on sample size requirements for normative studies (Oosterhuis et al., 2016), by providing formulas to compute the required sample size (instead of plots based on simulation studies), and by providing efficient robust designs, not only for studies that derive norms for a single test but also for studies norming multiple tests with the same sample. These results have an indirect impact on society because these guidelines can be used by researchers to carefully plan normative studies, thus preventing mistakes in the assessment of individuals (e.g. not recommending remedial teaching because imprecise norms based on a poorly designed normative study lead to an overestimation of a child's vocabulary size and arithmetic skills).

The results of this thesis are relevant for scientists (not necessarily statisticians) carrying out these two types of studies, because they can find in this work helpful guidelines for planning such studies. Since surveys are often performed by governmental institutions, such as national statistical institutes, chapters 2 and 3 might be more interesting for researchers based at these organizations. Normative studies, instead, are mainly conducted in educational, clinical, and neuropsychological research, so chapters 4 and 5 might be more relevant for researchers from these fields. Furthermore, chapters 2-5 of this thesis can also be used as teaching materials for mathematical statistics students, whereas the application sections of these chapters could be used to introduce students without a mathematical statistics background to the design of research studies.

In order to make the results of this thesis available to these target groups, chapters 2, 3, and 4 have been published in three international scientific journals, and chapter 5 has been submitted for publication. Furthermore, chapters 2 and 4 have been presented at two international scientific conferences. To compute the required size of the sample with the sample size formulas in chapters 3-5, R codes have been developed that will be made available when the chapter to which a code belongs has been published. However, more user-friendly software for applying the results of this thesis could be developed and made freely available to researchers. Such software could be accompanied by a non-technical summary or tutorial paper for a psychological, social, biomedical or health science journal, where the results of this thesis are further illustrated.