

Efficient designs for mean estimation in multilevel populations and test norming

Citation for published version (APA):

Innocenti, F. (2021). *Efficient designs for mean estimation in multilevel populations and test norming*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20210520fi>

Document status and date:

Published: 01/01/2021

DOI:

[10.26481/dis.20210520fi](https://doi.org/10.26481/dis.20210520fi)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Propositions belonging to the thesis

Efficient designs for mean estimation in multilevel populations and test norming

Francesco Innocenti

1. When cluster size is linearly related to the outcome variable of interest, and the cluster size distribution in the population is known before sampling, unbiased and efficient estimation of the outcome mean in a two-level population is attained by sampling clusters with probability proportional to size and computing the unweighted average of the sampled cluster means (this thesis).
2. When cluster size is unrelated to the outcome variable of interest, prior knowledge of the cluster size distribution in the population is not needed, because unbiased and efficient estimation of the population mean is simply attained by combining the unweighted average of cluster means with equal probability sampling of clusters and sampling the same number of individuals per selected cluster (this thesis).
3. Under linear regression of a quantitative outcome variable on a categorical predictor and a polynomial effect of order h for a quantitative predictor, the optimal design for norms estimation consists of $h + 1$ levels of the quantitative predictor, replicated at each level of the categorical predictor. Including additional levels of the quantitative predictor into the sample, as typically done in normative studies, yields a loss of efficiency under the constraint of a fixed total sample size (this thesis).
4. When prior knowledge about the effect of a quantitative predictor allows to rule out cubic or higher order polynomial effects, the balanced design consisting of three equidistant levels of the quantitative predictor guarantees both the highest minimum relative efficiency (i.e. relative to the optimal design for a model) and the highest minimum efficiency across all possible linear regression models for norming the outcome variable on a quantitative and a categorical predictor (this thesis).
5. In light of the practical importance of surveys and normative studies, efficient designs and sample size formulas for planning both types of studies have a direct impact on science because of unbiased and efficient estimation, and an indirect impact on society because of saving resources.
6. Statistics provides methodological tools for many other disciplines and this makes its impact on science substantial.
7. The societal impact of a statistical methodology is proportional to the societal impact of the disciplines that make use of it.
8. “The best thing about being a statistician is that you get to play in everyone’s backyard” (John Wilder Tukey).
9. The scientific debate between proponents of the design-based and model-based approach to survey sampling is one of the best epic sagas since Homer’s *Iliad*.
10. To deal with the extreme volatility of Dutch weather, a maximin approach is recommended.