

Estimation Weights for School Longitudinal Surveys: The case of Geres

Gabrielle Palermo

ITACOSM Florence

6 June 2019



Introduction

Changes in the composition of schools can be computed via sample surveys.

The Longitudinal Study of the School Generation 2005 (Geres) was used as an example.

Geres was applied from 2005 to 2008 in Brazilian schools, in five waves.

Geres: School Panels with population changes

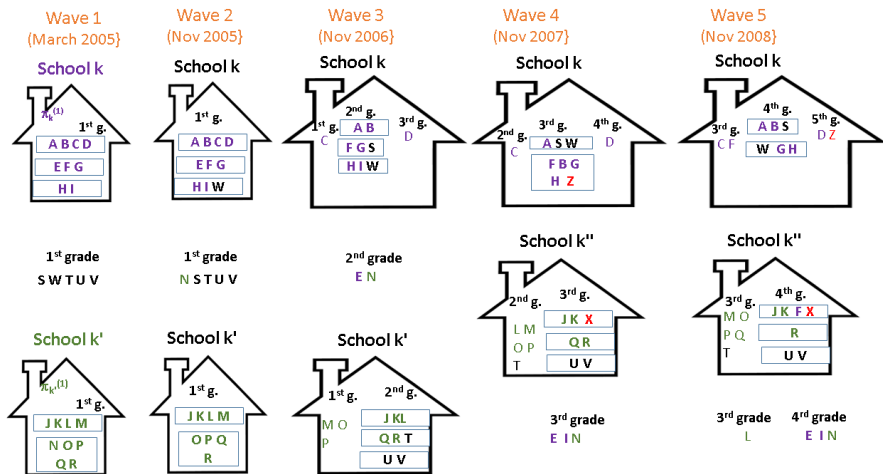


Figure: Summary of pupils mobility over time

Aim and research questions

The research aim is to estimate the total of repeaters at the fourth wave, attending municipal schools in the Rio de Janeiro city.

That is, pupils that were attending grades 1 and 2, given they had attended these grades before.

However, the survey data has weights only for the first wave, i.e., for the original sample. Then:

Is it possible to make accurate estimations for the population represented by the wave 4?

- Using the sampling weights?
- Or recomputed weights?

Which weight applies for the additional pupils that were not part of the sampling population?

School Panels

Citizenship education in England Study (CELS)

- Longitudinal cohort sample of pupils, 2001 to 2010;
- New cross-sectional sample of schools and peers every 2 years.

High School Longitudinal Study of 2009 (HSLS)

- USA, from 2009 until 2025. It does not add new pupils;
- Initial sample: school representativeness, but lost over time.

German National Educational Panel Study (NEPS)

- No sampling frame for kindergartens in Germany, cohort from 2010 until end of primary school;
- The sample was selected through the sampling frame of primary schools via indirect sampling.

Household Panels

Understanding society - UK Household Longitudinal Study (UKHLS)

Panel Study of Income Dynamics (PSID)

National Socioeconomic Characterization Survey: Panel CASEN 2006-2009

- Annual/biennial waves, with broad follow-up rules and additions of persons that start to live with original sample members;
- Cross-sectional weights for the waves > 1 : estimated via weight share methods.

Probability Sample

On the classical survey sampling theory, the sample is selected randomly directly from a finite population.

The population total at the initial wave $Y^{(1)}$ can be estimated according to the Horvitz-Thompson estimator:

$$\hat{Y}_{HT}^{(1)} = \sum_{k=1}^{m^{(1)}} \frac{y_k}{\pi_k} \quad (1)$$

Indirect Sampling and Weight Share Methods

Some populations do not have a sampling frame.

However, a given target population $U^{(t)}$ can be reached through links with another sample of a population $U^{(b)}$.

The weights for the sample $s^{(t)}$ can be estimated via weight share methods.

The Generalised Weight Share Method (GWSM) presented by Lavallée (1995, 2007) is a generalisation of the previous methods:

- Multiplicity Estimation (Birnbaum and Sirken, 1965; Sirken, 1970);
- Snowball Sampling (Goodman, 1961);
- Fair Share Method (Huang, 1984);
- Weight Share Method (Ernest, 1986).

Estimation weight

A general estimation weight w_k can be associated to school k of sample $s^{(t)}$. Then:

$$\hat{Y}^{(t)} = \sum_{k=1}^{m^{(t)}} w_k y_k \quad (2)$$

Lavallée (2007) showed $\hat{Y}^{(t)}$ is unbiased when the estimation weight is computed via GWSM.

GWSM Stages

1. Compute the initial weight w'_k for each school $k \in \Omega^{(t)}$.

$$w'_k = \sum_{i=1}^{N^{(b)}} \sum_{j=1}^{N_k^{(t)}} l_{i,kj} \times t_i \times w_i \quad (3)$$

Where:

- $l_{i,kj} = 1$ if pupil $i \in U^{(b)}$ corresponds to pupil $j \in U_k^{(t)}$, and 0 otherwise;
- $t_i = 1$ if $i \in s^{(b)}$, and 0 otherwise;
- $\Omega^{(t)}$ is the set of schools $k \in U^{(t)}$ that has pupils $j = i$, given $i \in s^b$.

GWSM Stages

2. Sum of the links into each school $k \in \Omega^{(t)}$ with population $U^{(b)}$

$$L_k^{(t)} = \sum_{i=1}^{N^{(b)}} \sum_{j=1}^{N_k^{(t)}} l_{i,kj} \quad (4)$$

3. The final weight $w_k^{(t)}$

$$w_k^{(t)} = \frac{w'_k}{L_k^{(t)}} \quad (5)$$

Where:

$w_{kj}^{(t)} = w_k^{(t)}$ for all pupil $j \in U_k^{(t)}$.

Sampling Design

Geres is the first survey of its kind successfully achieved and the only one in Brazil.

Geres had a stratified one-stage cluster sampling design:

- Cities: Rio de Janeiro, Belo Horizonte, Campinas, Campo Grande and Salvador;
 - Schools' administration system: municipal, state, private and exceptional schools;
 - Nine groups of schools according to the schools' size and socio-economic status.

The stratum of municipal schools of Rio de Janeiro was chosen for the present exercise.

Data Description

- All pupils registered in the 1st grade in the selected schools were selected to be part of the first wave and were followed five waves, from 2005 to 2008.
- Also, the survey followed the changes in the pupils' composition into and among the selected schools, given the pupils were attending the respective main grade at wave t .

The main grade for each wave is:

Wave 1: 1st grade;

Wave 2: 1st grade;

Wave 3: 2nd grade;

Wave 4: 3rd grade;

Wave 5: 4th grade.

Sampling totals - waves 1 and 4

Wave	n° schools	n° strata	Pupils
1 st	30	9	2,524
4 th	36	9	4,451

Figure: Total of pupils and schools on the 1st and 4th waves - Geres - Municipal schools of Rio de Janeiro

	Total	1st wave Population	1st wave Sample	Additional Pupils
Pupils total	4,451	3,662	2,007	789
Pupils at grade 1 or 2	381	363	292	18

Figure: 4th wave Totals - Geres - Municipal schools of Rio de Janeiro

Initial and additional pupils can have an estimation weight computed via Generalised Weight Share Method.

GWSM for Geres

School ID	Stratum ID	GWSM	Sampling weight	Sampling weight Recalculated
Total	-	519.43	765.00	765.03
33062439	1101	9.21	19.50	19.50
33083053	1101	14.41	19.50	19.50
33078343	1111	14.75	34.25	27.40
33080666	1111	20.55	34.25	27.40
33070717	1111	20.76	34.25	27.40
33087210	1111	22.83	34.25	27.40
33086834	1111	12.39	NA	27.40
33083169	1114	7.18	14.67	14.67
33086478	1114	9.02	14.67	14.67
33083061	1114	9.24	14.67	14.67
...				...

Figure: 4th wave Weights per school and stratum - Geres - Municipal schools of Rio de Janeiro

Estimated Totals - Geres

	GWSM	Sampling weight	Sampling weight Recalculated
Estimated Totals	5,292.94	8,207.17	7,712.49

Figure: Estimated totals of pupils that were attending the 1st or the 2nd grades on 2007 (wave 4), given they had attended these grades before.
Municipal schools of Rio de Janeiro
Geres

Final remarks

- The GWSM considers the changes in the schools composition;
- Also, new schools can have their weight estimated when they are attended by initial pupils;
- Both sampling weights have higher estimated totals;

Next steps

- Jackknife variance estimation for all three estimators, according to the sampling design;
- Calibration of the totals, considering the pupils that had not been followed;
- Design based simulation to compute the bias of each estimator.
- Longitudinal weights.

References

- Birnbaum, Z. W. and Sirken, M. G. (1965). Design of sample surveys to estimate the prevalence of rare diseases: Three unbiased estimates. *National Center for Health Statistics: Vital and Health Statistics (Series 2)*, (11).
- Ernest, L. (1986). Weighting issues for longitudinal household and family estimates.
- Goodman, L. A. (1961). Snowball sampling. *The annals of mathematical statistics*, pages 148–170.
- Huang, H. (1984). Obtaining cross-sectional estimates from a longitudinal survey: Experiences of the income survey development program. pages 670–675.
- Lavallée, P. (1995). Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method. *Survey Methodology*, 21(1):25–32.
- Lavallée, P. (2007). *Indirect sampling*. Springer Science & Business Media.
- Sirken, M. G. (1970). Household surveys with multiplicity. *Journal of the*

Grazie per l'attenzione!