COMPARISON OF THREE DIFFERENT BOOTSTRAP TECHNIQUES TO ESTIMATE THE AVERAGE HOUSEHOLD FOOD EXPENDITURES

Czaryda Palaganas and Demi Aira Dimatera

UP School of Statistics

Presented by:

Czaryda Palaganas

UP School of Statistics

MOTIVATION OF THE STUDY

- Resampling is a method that draws multiple samples from the original sample.
- Bootstrap is a nonparametric resampling method where repeated samples were drawn with replacement (SRSWR) from the original data samples
- According to Dimapilis (2013), at regional level, bootstrap estimation yields superior mean estimate than simple random sampling without replacement (SRSWOR) for the personal care and effect expenditures in the Philippines, where FIES 2006 was the population.

OBJECTIVE OF THE STUDY

QUESTIONS

- For a skewed population, is resampling yields a better estimate than simple random sampling?
- What if we use other sampling design rather than SRSWOR for drawing the original data samples?

OBJECTIVE

- This paper explores the use of resampling in estimating households' expenditure on food at regional level and compares the result with that of the same method without resampling.
- Using FIES 2015 as the population, the methods are SRSWOR, systematic sampling and probability proportional to size with replacement (PPSWR).

METHODOLOGY

1. The population (FIES 2015) is stratified into 17 regions (PSU). Then the 17 regions were stratified into three groups (SSU) with respect to the household size: 1-3 members for small household, 4-5 members for medium and 6+ members for large household.
2. From the SSUs, we will do three different sampling methods to obtain the original data samples with sampling rates 1%, 5% and 10%:
    1. Bootstrap Method (SRSWOR)
    2. Systematic resampling (lahiri method)
    3. PPSWR
3. Bootstrap resampling (SRSWR) is then performed on the data from the three different sampling methods 200 times.
4. The whole sampling scheme was replicated 100 times to ensure precision of the estimates.

Estimators of Population Mean and the Estimated Standard Error

Method	Estimator of the Population Mean	Estimated Standard Error
SRSWOR and Systematic	$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$	$\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$
SRSWR	$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$	$\frac{s}{\sqrt{n}}$
PPSWR	$\bar{y}_{HH} = \frac{\sum_{i=1}^{n} y_i/(Np_i)}{n}$	$\sqrt{\frac{1}{n}\frac{\sum_{i=1}^{n}\left(\frac{y_i}{Np_i} - \bar{y}_{HH}\right)^2}{n-1}}$

Where

- y_i = total food expenditure for each household
- p_i = number of members per household divided by sum of the number of members in every household in the population

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

RESULTS

	Total	Mean	Sd	Skewness	Kurtosis
Overall Statistics	3535359428	85099.16	51637.95	2.231	14.174
I - Ilocos Region	189366054	80649.94	44487.09	1.835	10.421
II - Cagayan Valley	167766071	75604.36	41362.29	2.360	16.605
III - Central Luzon	322815333	99726.70	51980.14	1.560	7.174
IVA - CALABARZON	438399898	105333.95	58274.66	1.568	8.164
V - Bicol Region	189877811	76811.41	37780.14	1.537	7.458
VI - Western Visayas	227592553	79829.03	43544.71	2.267	14.680
VII - Central Visayas	214224555	84307.18	59798.10	3.748	33.997
VIII - Eastern Visayas	163201892	69833.93	40543.00	2.404	13.764
IX - Zasmboanga Peninsula	124525829	69645.32	44465.59	2.179	11.961
X - Northern Mindanao	120980449	64112.59	39497.08	2.113	10.077
XI - Davao Region	198436464	81126.93	46112.90	2.002	11.490
XII - SOCCSKSARGEN	152228224	71738.09	44938.17	4.239	46.546
Caraga	128148360	71912.66	43471.03	3.925	41.453
NCR	524842286	127080.46	65586.97	1.429	6.882
CAR	138608546	80352.78	41194.36	1.438	6.716
ARMM	145965496	64931.27	26007.90	2.275	15.506
IVB - MIMAROPA	88379607	70760.29	41638.03	2.248	12.488

Estimated Average Household Food Expenditures and its Standard Error of Each Method

With Resampling						
Sampling Rate	1%		5%		10%	
Method	$\bar{y}$	$se(\bar{y})$	$\bar{y}$	$se(\bar{y})$	$\bar{y}$	$se(\bar{y})$
SRSWOR	85157.32	12463.66	84955.9	5971.38	85123.17	4213.218
Systematic	85225.68	12583.46	85156.89	6003.241	85062.71	4216.872
PPSWR	85165.78	12157.59	85062.07	5970.841	85110.05	4335.07

Without Resampling						
Sampling Rate	1%		5%		10%	
Method	$\bar{y}$	$se(\bar{y})$	$\bar{y}$	$se(\bar{y})$	$\bar{y}$	$se(\bar{y})$
SRSWOR	85168.16	11770.71	84953.64	6047.163	85117.68	4411.7
Systematic	85226.95	11878.45	85152.93	6082.67	85067.28	4417.105
PPSWR	87687.45	11892.14	87692.64	6186.761	87801.32	4506.257

CV, Bias and Deff of the Estimated Average Household Expenditure of Every Sampling Method

With Resampling									
	$CV(\%)$			$Bias$			$deff$		
Rate	1%	5%	10%	1%	5%	10%	1%	5%	10%
SRSWOR	14.64	7.02	4.95	58.16	-143.25	24.02	1.12	0.97	0.91
Systematic	14.79	7.05	4.95	126.51	57.73	-36.45	1.14	0.98	0.91
PPSWR	14.29	7.07	5.09	66.61	-37.08	10.89	1.06	0.97	0.96

Without Resampling									
	$CV(\%)$			$Bias$			$deff$		
Rate	1%	5%	10%	1%	5%	10%	1%	5%	10%
SRSWOR	13.83	7.11	5.18	69.00	-145.52	18.52	NA	NA	NA
Systematic	13.96	7.15	5.19	127.79	53.77	-31.88	1.02	1.01	1.01
PPSWR	13.97	7.27	5.29	2588.29	2593.48	2702.16	1.02	1.05	1.04

CONCLUDING REMARKS

- In determining the average household food expenditure, bootstrap resampling yields superior estimate than without resampling.
- Specifically, ‘basic’ bootstrap best estimates the mean at 1% sampling rate and PPSWR with resampling is the best method for 5% and 10% sampling rate.
- Moreover, the bootstrap estimation procedures have greater gain in precision and accuracy as shown in the results.
- At regional level, systematic resampling estimates was superior in terms of accuracy and precision.