POVERTY MEASURES ESTIMATION FOR RESPONDENT-DRIVEN SAMPLING

Poverty measures are crucial socio-economic indicators that are used to evaluate the economic well-being of a region or a state. In this paper we investigate the problem of estimating poverty measures for data collected through Respondent-Driven Sampling (RDS), a refined form of snowball sampling commonly used to survey hard-to-reach populations. We propose estimators of the distribution function and in particular, estimators of poverty measures. We explore their performance in both simulated and empirically collected network datasets. We discuss the applicability of the methods and suggest areas of future research.

KEYWORDS: Poverty measures; Respondent-driven sampling; Distribution function.

Statement of Significance

Some populations are difficult to reach and survey due to social prejudice and stigma. Indigenous populations are often affected by poverty and poor living conditions and are often underrepresented in oficial statistical agencies, making traditional sampling impractical for surveying these groups. In this paper, we present a method for estimating poverty measures using Respondent-Driven Sampling, a survey methodology that is used to recruit hard-to-reach populations, such as marginalized groups, hidden populations, or stigmatized communities. We derive some properties of the proposed methods. We evaluate the proposed method using Monte Carlo simulations and a case study of Indigenous people in Ecuador.

1. INTRODUCTION

Economic indicators such as gross domestic product, unemployment rate, and inflation rate, offer valuable insights into the economic prosperity of a population for analysis and prediction purposes. These indicators can be utilized to evaluate the economic performance of a nation or region, monitor fluctuations in poverty levels over a period of time and forecast future economic growth. Economic indicators can also be used to compare the economic performance of different countries or regions, which can be useful for policymakers and organizations looking to understand the poverty levels and the characteristics of these communities. Poverty can have a significant impact an individual's physical and mental health, education, and overall quality of life ([?, ?]) Understanding poverty is crucial for evaluating the well-being of a society. This is why many fields, like social sciences, have dedicated a lot of research to studying poverty. Reducing poverty is a major goal for governments and organizations that want to improve people's lives. International organizations, such as the United Nations and the European Union, have established poverty reduction targets with fixed time frames as part of their development goals.

Many government agencies conduct surveys to gather information about people's income and living conditions. This has led to a growing interest in studying poverty and its causes. These surveys provide information about people's characteristics such as income, age, gender, health, education, and employment status. With more data available, researchers can now study poverty among vulnerable groups like children, women, less-educated people, and the elderly.

Poverty measures are important as economic indicators as they provide valuable insights into the socio-economic conditions and level of inequality in a society. They are used to assess the poverty status of a population and to monitor changes in poverty over a period of time. They provide useful information for policy making, monitoring progress and for targeting of resources and programs to those most in need.

There exists a wide literature about poverty measurement ([?], [?], [?], [?]). A discussion of competing definitions of poverty and different ways of measuring it is given by [?]. Some recent research on poverty measures are given by...

Indigenous households are affected by poverty and unmet basic needs, compared to those of mixed origins in Ecuador([?]), which are a majority in this country, resulting in limited access to education and challenges in obtaining decent housing [?]. According to studies and official statistical agency data, Indigenous people and other ethnic minorities have a higher rate of unemployment and less access to health, education, and housing than those of mixed origins. ([?, ?, ?, ?]). Several studies have examined the gender and ethnic wage gaps, providing evidence of discrimination in the job market ([?, ?]). Additionally, research have shown discrimination against Indigenous people and Afro-descendants in the labor market in Ecuador ([?, ?, ?])

Obtaining reliable information about minority groups, particularly ethnic minorities, poses a challenge for survey research professionals. These groups often have a small representation in the overall population, making it difficult to access them and leading to lower response rates. [?] demonstrated through a RDS survey in Canada that the Indigenous population in Toronto is underrepresented, as indicated by undercounts in the census of Statistics Canada. Additionally, Indigenous people may also be underrepresented in the Ecuadorian National Statistics Agency due to social prejudice and stigma that discourages identification as members of these groups ([?, ?, ?, ?, ?, ?, ?, ?]). Therefore, Indigenous people are difficult to reach and they are frequently underrepresented in official statistics surveys ([?]). Utilizing traditional sampling methods to survey them can be financially demanding and unfeasible in practice.

To overcome the challenges of surveying small groups within a population, alternative methods such as snowball sampling or Respondent-driven sampling (RDS) can be utilized. These non-probabilistic approaches address the issues of underrepresentation of such groups. Respondent-driven sampling (RDS) is a snowball-type sampling method that is used to recruit hard-to-reach, stigmatized and/or elusive populations. It was first introduced by [?] and it utilizes a chain-referral method, where participants recruit other members of the population of interest. This process continues until a sufficient number of participants are obtained for the survey. RDS is useful for recruiting hard-to-reach populations and obtaining a representative sample.

RDS has been used in a variety of research contexts, including public health, sociology, and psychology, to study populations that may be hard to reach or that are less likely to participate in traditional research studies. Some examples of these types of populations include sex workers, people who inject drugs, and marginalized or stigmatized communities. RDS is useful in these cases because it can help researchers access and study these populations without relying on traditional sampling methods that may not be effective.

Poverty studies use different measures to determine poverty levels like the headcount index and percentile ratios. Most of these measures are usually based on a poverty line, which is a which is a certain income level that separates poor people from non-poor people. The poverty gap index shows how severe poverty is by measuring the difference between poor people's incomes and the poverty line. The poverty gap index is a measure of the intensity of poverty that indicates how much would have to be transferred to the poor to bring them out of poverty, that is, the minimum cost to eliminate poverty if transfers were perfectly targeted.

In this article, we consider the problem of estimating poverty measures for hard-to-reach populations through Respondent-driven sampling (RDS). The proposed methods are based on the estimation of the distribution function and quantiles. We compare the proposed methods empirically to the popular corresponding probabilistic estimator of the distribution funcion. To our knowledge, no such an approach has been done to study these economic indicators for RDS. The article is organized as follows. In section 2, we present the proposed estimators of poverty measures. Then, some theoretical properties og these methods such as... are derived. Artificial populations based on the normal and lognormal distributions are also used to analyze the practical performance of the estimators in Section 4. A case study involving an Indigenous population RDS survey in Ecuador is studied in Section 5. . Finally, concluding remarks are provided in Section 7.

2. METHODS

2.1 Basic setup in estimation of the distribution function

Let $U = \{1, \ldots, N\}$ be a finite population with size N. Let Y be the variable of interest and let y_k be the value of Y for unit k. We assume that y_k is known for all sample units. Our aim is to estimate the distribution function $F_y(t)$ for the study variable Y which can be defined as follows:

$$F_y(t) = \frac{1}{N} \sum_{k \in U} \Delta(t - y_k) \tag{1}$$

where $\Delta(\cdot)$ denotes the Heaviside function, given by:

$$\Delta(t - y_k) = \begin{cases} 1 & \text{si } t \geq y_k \\ 0 & \text{si } t < y_k. \end{cases}$$

Let $s = \{1, 2, \dots, n\}$ be a random sample of fixed size n, drawn from U with a specified sampling design $p(\cdot)$ that assigns known inclusion probabilities of first and second order denoted by $\pi_k > 0$ and $\pi_{kl} > 0$ $k, l \in U$ respectively. The value $d_k = \pi_k^{-1}$ denotes the corresponding sampling design weight for unit $k \in U$. Under this scenario the distribution function $F_y(t)$ can be estimated by the Horvitz Thompson estimator, defined by

$$\widehat{F}_{YHT}(t) = \frac{1}{N} \sum_{k \in s} d_k \Delta(t - y_k). \tag{2}$$

This estimator $\widehat{F}_{YHT}(t)$ is design-unbiased for $F_y(t)$ and is a genuine distribution function [?]. If the size of the population N is unknown, the estimator used is the Hájek estimator:

$$F_{YHa}(t) = \frac{\sum_{k \in s} d_k \Delta(t - y_k)}{\sum_{k \in s} d_k}.$$

Several methods for estimating population distribution functions from sample survey data using auxiliary information are described in [?], [?], [?], [?], [?] or [?] among others. Several of these estimators are not monotone and therefore it is necessary to transform them to guarantee monotonicity and to be able to take the inverse and calculate the desired estimation of the quantile.

One of the most considered approaches when incorporating auxiliary information to develop new estimators of the distribution function is the use of the calibration method [?]. Thus, there is an extensive previous literature that uses different implementations of the calibration approach to obtain estimators of the distribution function and the quantiles [?, ?, ?, ?, ?, ?, ?, ?, ?, ?], and specifically for poverty measures [?, ?].

2.2 Proposed estimators of poverty measures in RDS

Now we consider the problem of the estimation of the distribution function in the RDS context. Thus, we assume the target population is connected by a network of mutual relations with $N \times N$ adjacency matrix $\mathbf{Z}$. This means that $z_{kj} = z_{jk} = 1$ if k and j are connected and 0 otherwise. We define the nodal degree of a the person k, $\delta_k = \sum_j z_{kj}$, as the number of network ties or alters of node k.

The RDS selection process starts with a set of initial members of the target population called seeds, that represent the wave 0 of the sample. These respondents are given recruitment coupons (typically three), so that they recruit the next wave of participants, among their known contacts within the hidden group, usually with incentives [?]. When these respondents return their coupons, they recruit the next wave of participants. This process is repeated until the desired sample s of size n, is attained [?].

Based on data observed in the RDS sample s we can estimate the distribution function $F_y(t)$ by using the ordinary sample empirical distribution function:

$$\widehat{F}_{Yn}(t) = \frac{1}{N}\sum_{k\in s} d_k \Delta(t - y_k). \tag{3}$$

We will denote this estimator as naive estimator.

The naive estimator can not be unbiased for $F_y(t)$ and do not use the information given by the network. In order to use this information we propose an estimator of the distribution function $F_y(t)$ in RDS as

$$\widehat{F}_{Hyd}(t) = \frac{\sum_{k\in s} \delta_k^{-1}\Delta(t - y_k)}{\sum_{k\in s} \delta_k^{-1}}, \tag{4}$$

with δ_k the degree reported by respondent k. This estimator is similar to the Hájek estimator of the distribution function and the degree plays a similar role in a RDS setting to the role played by the first-order probability in a probabilistic survey sampling context. The proposed estimator is also similar to the RDSII mean estimator proposed in [?] but changing the values y_k to $\Delta(t - y_k)$.

In estimating the distribution function, it is important to know whether a proposed estimator satisfies the distribution function properties, that is, whether the new estimator is a genuine distribution function. The estimator $\widehat{F}_{Hyd}(t)$ will be a genuine distribution function if the following conditions are satisfied:

1. $\widehat{F}_{Hyd}(t)$ is continuous on the right.
2. $\lim\limits_{t\to-\infty} \widehat{F}_{Hyd}(t) = 0$,
3. $\lim\limits_{t\to+\infty} \widehat{F}_{Hyd}(t) = 1$
4. $\widehat{F}_{Hyd}(t)$ is monotone nondecreasing.

The proposed estimator $\widehat{F}_{Hyd}(t)$ satisfies all these conditions, thus we can estimate quantiles and hence achieve appropriate poverty measures based on quantiles. We define an estimator of the quantiles by considering the inverse of $\widehat{F}_{Hyd}(t)$:

$$\widehat{Q}_y(\alpha) = inf\{t : \widehat{F}_{Hyd}(t)) \geq \alpha\} = \widehat{F}_{Hyd}^{-1}(\alpha) \tag{5}$$

Once we have estimators of the distribution function and the quantiles, we can now estimate poverty measures, such as the Gini coefficient, the poverty risk HCI and the interquantile and interdecile ratios.

The Gini coefficient is a measure of inequality of a distribution. Eurostat [?] defined the Gini coefficient as the relationship of cumulative shares of the population arranged according to the level of income, to the cumulative share of the income received by them. The Gini coefficient [?] is estimated by

$$\widehat{G}_y = \frac{\sum_{k\in s} \delta_k^{-1}(2\widehat{F}_{Hyd}(y_k) - 1)y_k}{\sum_{k\in s} \delta_k^{-1} y_k}. \tag{6}$$

The poverty risk or Headcount index HCI [?] considers the proportion of individuals with an income below the at risk-of-poverty threshold, which is set at 60 % of the national median equivalised disposable income. It is estimated as

$$\widehat{HCI} = \frac{\sum_{k\in s} \delta_k^{-1} I(y_k < 0{,}6\widehat{Q}_y(0{,}5))}{\sum_{k\in s} \delta_k^{-1}}. \tag{7}$$

The poverty gap index PGI [?] measures the extent to which people fall below the poverty line. It is estimated as

$$\widehat{PGI} = \frac{\sum_{k\in s} \delta_k^{-1} \frac{0{,}6\widehat{Q}_y(0{,}5) - y_k}{0{,}6\widehat{Q}_y(0{,}5)} I(y_k < 0{,}6\widehat{Q}_y(0{,}5))}{\sum_{k\in s} \delta_k^{-1}}. \tag{8}$$

We consider the estimation of the Sen index SEN [?], which aims to take into account the number of individuals living in poverty, the proportion of the population affected by poverty and the distribution of poverty. This estimate is given as

$$\widehat{SEN} = \widehat{HCI} * \widehat{G}_y + \widehat{PGI}(1 - \widehat{G}_y). \tag{9}$$

The interquartile and the interdecile ratios are measures of spread of a distribution. The first one is estimated as

$$\widehat{IQR} = \frac{\widehat{Q}_y(0{,}75)}{\widehat{Q}_y(0{,}25)}, \tag{10}$$

and the interdecile ratio is estimated as

$$\widehat{IDR} = \frac{\widehat{Q}_y(0{,}90)}{\widehat{Q}_y(0{,}10)}. \tag{11}$$

3. VARIANCE ESTIMATION

The estimation of the variances for the defined poverty measures is not a simple problem: on the one hand, there is the complication of the RDS design and, on the other, the complexity introduced by not being linear functions. In this section we will consider some possible strategies.

First, we analyse the problem of the estimation of a quantile by using the methodology used in [**?**] and [**?**].

We consider the linear approximation:

$$\widehat{Q}_y(\alpha) - Q_y(\alpha) \approx \frac{1}{f_y(Q_y(\alpha))}(\widehat{F}_{Hyd}(Q_y(\alpha)) - F_y(Q_y(\alpha))$$

and

$$f_y(Q_y(\alpha)) \approx 2z_{\frac{\gamma}{2}}\sqrt{V(\widehat{F}_{Hyd}(Q_y(\alpha)))}/L_y(\gamma)$$

where $f_y(\cdot)$ is the derivative of the limiting value of $F_y(\cdot)$ as N tends to infinite and $L_y(\gamma)$ is the length of the Woodruff interval whose lower and upper limits are given by

$$\text{ínf}\left\{t; \widehat{F}_{Hyd}(t) \geq \alpha - z_{\frac{\gamma}{2}}\sqrt{\hat{V}(\widehat{F}_{Hyd}(Q_y(\alpha)))}\right\}$$

$$\text{ínf}\left\{t; \widehat{F}_{Hyd}(t) \geq \alpha + z_{\frac{\gamma}{2}}\sqrt{\hat{V}(\widehat{F}_{Hyd}(Q_y(\alpha)))}\right\}$$

where $z_{\frac{\gamma}{2}}$ is the upper $\frac{\gamma}{2}$-point of a $N(0,1)$ variable (see [**?**]).

Thus an estimator to the variance estimator for $\widehat{Q}_y(\alpha)$ can be obtained by:

$$\hat{V}(\widehat{Q}_y(\alpha)) = (\frac{L_y(\gamma)}{2z_{\frac{\gamma}{2}}})^2 \tag{12}$$

The problems thus is to obtain an estimator of the variance $V(\widehat{F}_{Hyd}(Q_y(\alpha)))$. [**?**] propose an analytical variance estimator of the RDSII estimator which accounts for nonuniform selection probabilities and the structure of the sample. Our proposed estimator is a particular case considering the y_i value as $\Delta(t-y_i)$. Thus one can use formula 17 in [**?**] for estimating $\hat{V}(\widehat{F}_{Hyd}(Q_y(\alpha)))$ in (**??**).

Using the above variance formulae for $\widehat{Q}_y(\alpha))$ and a similar expression for the covariance between $\widehat{Q}_y(\alpha)$ and $\widehat{Q}_y(\beta)$ (see [**?**]) the variance estimators of interquartile and the interdecile ratios are obtained by using the approximation of a ratio.

The resulting variance estimators are not unbiased and are based on linear approximations, thus the their behavior in practice may not be good. Therefore it is useful to use resampling methods that take into account both the variability of the sample and the network. [**?**] says that any technique which attempts to estimate the uncertainty should ensure that the seed node is allowed to be random and proposes a bootstrap technique for variance estimation in RDS. In the application we will use a resampling method (the jackkniffe technique) adapted to RDS context for the construction of confidence intervals.

4. MONTE CARLO SIMULATIONS

In this section, we study the performance of the proposed estimators in simulation experiments in different scenarios. The main factor of interest is the estimation of poverty measures. We used our own code written in `R` to compute the proposed estimators. Programming details and code are available from the authors.

Two scenarios were used to generate the values of the variable of interest y. In the first scenario, y_j was taken from a normal distribution with mean 1500 and standard deviation 500, for $j = 1, \dots, 1000$. In the second scenario, y_j was generated from a lognormal distribution with mean 0 and standard deviation 1, for $j = 1, \dots, 10000$. Other population parameters were also tested, but the results were similar and not reported here. Sample size was $n = 500$ and samples were selected using simple random sampling without replacement, similar to how RDS is typically conducted. ESTO NO PUEDE SER. SE SUPONE QUE LAS MUESTRAS SON ELEGIDAS CON RDS COMO DETALLA EN EL PaRRAFO SIGUIENTE

The simulated population size was $N = 10000$. A NxN network connection indicator matrix C was randomly generated, with c_{ij} either 0 or 1, a connection indicator between node i and j, for $i, j = 1, \dots, N$. The degree will be determined by the c_{ij} as $\sum_{i \in U, i \neq j} c_{ij} = \delta_j$. Ten seeds were selected at random from the network with probability proportional to their degree, with three maximal coupons issued for each participant. HABRiA QUE INCLUIR OTRA FORMA DE SIMULAR LA RED, YA QUE AL HACERLA ALEATORIA, EL ESTIMADOR USUAL DE LA FUNCIoN DE DISTRIBUCIoN FUNCIONA MEJOR, QUE ES LO QUE SALE EN LA SIMULACIoN Y POR TANTO EL ARTiCULO NO TIENE VALOR. MIRA EN OTRO ARTiCULO UNA RED NO ALEATORIA, QUE ES DONDE TIENE SENTIDO USAR EL RDS Y EL ESTIMADOR PROPUESTO

The proposed method is evaluated in the problem of the estimation of poverty measures. It is empirically compared to the reference naive estimator $\widehat{F}_{Yn}$ CAMBIALO EN EL RESTO. We investigated the percent relative bias

$$rb\% = \frac{E_{MC}(\widehat{\theta} - \theta)}{\theta} * 100,$$

and the percent relative mean squared error

CAMBIA rmse PARA QUE LA BASE SEA EL PARaMETRO Y NO EL ESTIMADOR NAIVE

being $\widehat{\theta}$ the proposed estimator of the population poverty measures and $\widehat{\theta}_H$ the estimator of the poverty measures using the Horvitz-Thompson estimator. E_{MC} denotes the average of the Monte Carlo replications. Simulation results were based on $B = 1000$ samples.

We evaluate the performance of the proposed estimator against the reference estimator $\widehat{F}_{YH}$. The reference estimator makes use of the population size and sample sizes available and therefore uses the first-order inclussion probabilities, as in a probabilistic SRSWOR setting.

Cuadro 1: Percent relative mean squared errorrmse % for poverty measures estimation in the two scenarios

Measure	$\widehat{G}_y$	$\widehat{HCI}$	$\widehat{PGI}$	$\widehat{SEN}$	$\widehat{IQR}$	$\widehat{IDR}$
Scenario 1	150.32	133.51	158.85	159.49	128.82	152.20
Scenario 2	114.64	095.76	125.18	143.70	130.75	140.13

Cuadro 2: Percent relative bias rb % for poverty measures estimation in the two scenarios

			Scenario 1			
Measure	$\widehat{G}_y$	$\widehat{HCI}$	$\widehat{PGI}$	$\widehat{SEN}$	$\widehat{IQR}$	$\widehat{IDR}$
$\widehat{F}_{Hyd}$	-0.6086	-0.7331	-1.1624	1.6167	0.0568	-0.9161
$\widehat{F}_{YH}$	-0.4929	-2.0887	-3.4415	-0.3574	0.0906	-1.3482
			Scenario 2			
Measure	$\widehat{G}_y$	$\widehat{HCI}$	$\widehat{PGI}$	$\widehat{SEN}$	$\widehat{IQR}$	$\widehat{IDR}$
$\widehat{F}_{Hyd}$	0.2751	0.5705	0.3411	-0.1132	1.2266	0.5481
$\widehat{F}_{YH}$	0.3247	-3.5486	-3.6978	-0.4751	1.3601	0.1072

Cuadro 3: Estimated poverty measures, ENEMDU survey values and sampled values for the RDS survey on ethnic minorities in the Canton of Riobamba, Ecuador

Method	$\widehat{G}_y$	$\widehat{HCI}$	$\widehat{PGI}$	$\widehat{SEN}$	$\widehat{IQR}$	$\widehat{IDR}$
Proposed estimator	0.4420	0.2995	0.1209	0.1998	2.5333	7.1429
HT estimators	0.4613	0.3146	0.1538	0.2280	3.2000	8.2857
ENEMDU survey	0.2916	0.1785	0.1214	0.1380	1.7303	16.8333

Tables 1 and 2 show the rmse % and the rb % values for the estimators being compared. The results indicate that the proposed method performs well with similar relative bias values than the reference estimator. Predictably, the proposed method has inferior efficiency scores to those given by the unfeasible reference estimator $\widehat{F}_{YH}$.

5. CASE STUDY

We evaluate the performance of the proposed estimators in a real-world RDS dataset on ethnic minorities surveyed in the Canton of Riobamba, Ecuador [?]. The survey aimed to investigate the living conditions and socioeconomic issues of young Indigenous, Montubios, and Afro-Ecuadorians. The RDS methodology is well-suited for studying these populations, as they are a hidden population because of stigma and they form a well-connected group.

A total of 814 people were recruited in six waves and questioned on their social and economic background and living conditions using a dual system of incentives to motivate recruitment. The reported income of the household is the variable of interest and we consider the estimators of the distribution function and quantiles to compute poverty measures of this group of young ethnic minorities in Ecuador. explicar como se han obtenido los estimadores de las varianzas, haz alusion al trabajo en el que usamos ese jackkniffe

Table ?? shows the proposed estimators applied to the above-mentioned RDS network dataset, together with the reference estimator and the poverty measures computed for the ENEMDU 2019 National Ecuadorian survey.

The Gini coefficient is commonly used to measure the distribution of income or wealth within a population, it ranges from 0 to 1, the higher the Gini coefficient, the greater the inequality. ($\widehat{G}_y$=0.4420) indicates an intermediate level of inequality among young ethnic minorities in Ecuador. Additionally, the headcount index measures the proportion of individuals living below the poverty line, $\widehat{HCI}$ = 0,29956 reveals that 29.95 % of participants have an income below the at-risk-of-poverty threshold. Furthermore, the poverty gap index, which measures the depth of poverty or the average shortfall of the poor from the poverty line, is 12.09 %. This value represents the minimum cost required to eliminate poverty in relation to the at-risk-

Cuadro 4: Estimated poverty measures by the Proposed and the reference estimator, Lower (LB) and upper bounds (UB) of the 95 % Confidence Interval obtained with the Chain Jackknife variance estimator, ENEMDU survey values for the RDS network data on ethnic minorities in the Canton of Riobamba, Ecuador

Method	Proposed	LB	UB	HT	LB	UB	ENEMDU
$\widehat{G}_y$	0.4420	0.4012	0.4828	0.4613	0.3190	0.6036	0.2916
$\widehat{HCI}$	0.2995	0.0645	0.5345	0.3146	0.2477	0.3815	0.1785
$\widehat{PGI}$	0.1209	0.0777	0.1641	0.1538	0.1189	0.1887	0.1214
$\widehat{SEN}$	0.1998	0.1543	0.2453	0.2280	0.1881	0.2679	0.1380
$\widehat{IQR}$	2.5333	0.8349	4.2317	3.2000	1.2868	5.1132	1.7303
$\widehat{IDR}$	7.1429	2.4047	11.9411	8.2857	3.4192	13.1378	16.8333

of-poverty threshold. The SEN index, is a measure of poverty that considers both the severity of poverty and the distribution of income. A higher value of the index indicates a greater level of poverty severity within a population, a value of 0.1998 indicates a relatively low level of poverty severity.

The Interquartile Ratio $\widehat{IQR}$ is a measure of dispersion in a statistical distribution. The closer the $\widehat{IQR}$ value is to 1, the smaller the spread in the central 50 % of the distribution. $\widehat{IQR}$ value of 2.5333 indicates a significant gap in incomes within the central portion of the distribution. Additionally, the $\widehat{IQR}$ value of 7.1429 further supports the presence of important differences along the distribution of incomes. The ENEMDU survey also supports these findings, as it reveals even larger disparities in income.

Table **??** reports the estimated poverty measures with respect to Gender. It shows similar overall values for the poverty measures for both men and women. However, there is a greater dispersion in income for women as compared to men, as indicated by both the estimated Interquartile Ratio $\widehat{IQR}$ and $\widehat{IDR}$ values. Furthermore, we calculated the above-mentioned poverty measures with respect to Education, but were not reported here as no significant differences were obtained.

5.CONCLUSION

Acknowledgments

The research was partially supported by a grant from Ministerio de Educación y Ciencia (PID2019-106861RB-I00, Spain), from IMAG-Maria de Maeztu CEX2020-001105-M/AEI/10.13039/501100011033 and from FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades (FQM170-UGR20, A-SEJ-154-UGR20).

Cuadro 5: Estimated poverty measures, ENEMDU survey values and sampled values for the RDS survey on ethnic minorities in the Canton of Riobamba, Ecuador considering Gender

Method	Gender	$\widehat{G}_y$	$\widehat{HCI}$	$\widehat{PGI}$	$\widehat{SEN}$	$\widehat{IQR}$	$\widehat{IDR}$
Proposed estimators	Woman	0.4518	0.2898	0.1097	0.1911	3.600	7.5000
	Man	0.4213	0.2349	0.1071	0.1609	2.0000	5.6000
ENEMDU	Woman	0.3115	0.2565	0.1547	0.1864	2.0987	12.5575
	Man	0.2858	0.2252	0.1350	0.1607	1.9483	7.4859
HT estimators	Woman	0.4668	0.3038	0.1256	0.2088	3.7000	9.1667
	Man	0.4497	0.2434	0.1154	0.1729	2.0000	6.000

Bibliografía

[1] Atkinson, T. (1987) On the Measurement of Poverty. Econometrica, vol 55: 749-764.

[2] Araki, Hidekazu: Movimientos étnicos y Multiculturalismo en el Ecuador: Pueblos Indígenas, Afrodescendientes y Montubios. (2012) Master's Dissertation, University of Kanagawa, Kanagawa, Japan (2012)

[3] Beckett, M.; Firestone, M.A.; McKnight ,C.D. A cross-sectional analysis of the relationship between diabetes and health access barriers in an urban First Nations population in Canada. *BMJ Open* **2018**, *8*, e018272, doi:10.1136/bmjopen-2017-018272.

[4] Clapham, D.; Foye, C.; Christian, J. The Concept of Subjective Well-being in Housing Research. *Hous. Theory Soc.* **2018**, *35*, 261â€“280, doi:10.1080/14036096.2017.1348391.

[5] Canelas, C., Salazar, S. Gender and ethnic inequalities in LAC countries. IZA J Labor Develop 3, 18 (2014). https://doi.org/10.1186/2193-9020-3-18

[6] Clark, S., R. Hemming, and D. Ulph. 1981. â€˜â€˜On Indices for the Measurement of Poverty.â€™â€™ Economic Journal 91:515-26.

[7] Chisaguano, S. 2006. *La población indígena del Ecuador (Análisis de Estadísticas Socio-Demográficas).* Available online: https://www.acnur.org/fileadmin/Documentos/Publicaciones/2009/7015.pdf. (accessed on 10 August 2020).

[8] Flores Morales, S. T. (2019) Discriminación en el mercado laboral de pueblos indígenas y afrodescendientes en el Ecuador. Periodo 2009-2017. Universidad San Francisco de Quito. Proyecto de Investigación

[9] Fraze, T.K.; Brewster, A.L.; Lewis, V.A.; Beidler, L.B.; Murray, G.F.; Colla, C.H. Prevalence of Screening for Food Insecurity, Housing Instability, Utility Needs, Transportation Needs, and Interpersonal Violence by US Physician Practices and Hospitals. *JAMA Netw. Open* **2019**, *2*, e1911514, doi:10.1001/jamanetworkopen.2019.11514.

[10] Gallardo, Lourdes, Ńopo, Hugo (2009) : Ethnic and gender wage gaps in Ecuador, Working Paper, No. 679, Inter-American Development Bank, Research Department, Washington, DC.

[11] Garcia-Aracil, A., Winter, C. (2006) Gender and ethnicity differentials in school attainment and labor market earnings in Ecuador, World Development, 34, 2, pp. 289-307. https://doi.org/10.1016/j.worlddev.2005.10.001.

[12] Haughton, J. and S. R. Khandker. 2009. Handbook on Poverty and Inequality. Washington, DC: The World Bank.

[13] Erens, B. Designing high-quality surveys of ethnic minority groupsin the United Kingdom. In book *Surveying Ethnic Minorities and Immigrant Populations: Methodological Challenges and Research Strategies*; Font, J.; MÃ©ndez, M., Eds.; Amsterdam University Press: Amsterdam, Holland, 2003; pp. 45â€“68.

[14] Espinoza, M.A., Gallegos, D. (2018). Discriminación laboral en Ecuador. Revista espacios, 39, 23, pp 32

[15] Feskens, R.; Hox, J.; Lensvelt-Mulders, G.; Schmeets, H. Collecting Data among Ethnic Minorities in an International Perspective. *Field Methods* **2006**, *18*, 284â€“304, doi:10.1177/1525822X06288756.

[16] Firestone, M.; Smylie, J.; Maracle, S.; McKnight, C.; Spiller, M.; Oâ€™Campo, P. Mental health and substance use in an urban First Nations population in Hamilton, Ontario. *Can. J. Public Health* **2015**, *106*, e375â€“e381.

[17] Font, J.; MÃ©ndez, M. (Eds.) *Surveying Ethnic Minorities and Immigrant Populations: Methodological Challenges and Research Strategies*; Amsterdam University Press: Amsterdam, Holland, 2013.

[18] Secretaría Nacional de Planificación y Desarrollo. 2017. *Plan Nacional de Desarrollo 2017–2021.*

[19] Ramírez-Luzuriaga, M.; Belmont, P.; Waters, W.; Freire, W. Malnutrition inequalities in Ecuador: Differences by wealth, education level and ethnicity. *Public Health Nutr.* **2019**, 1–9, doi:10.1017/S1368980019002751.

[20] Pablo, Q.S.; Paloma, T.L.P.; Francisco, J.T. Energy Poverty in Ecuador. *Sustainability* **2019**, *11*, 6320, doi:10.3390/su11226320.

[21] Vásquez, Rommel. Compañías Incluyentes: Â¿una Nueva Receta Para Doblegar a la Pobreza en el Ecuador? Master's Dissertation, Universidad Andina Simón Bolívar, Quito, Ecuador, 2009.

[22] Martínez, Luciano. Desarrollo Rural y Pueblos IndíGenas: Las Limitaciones de la Praxis Estatal y de las ONG en el Caso Ecuatoriano. Master's Dissertation, FLACSO, Quito, Ecuador, 2002.

[23] Handcock, M.S., Morris, M.: Relative Distribution Methods in the Social Sciences; Springer Science & Business Media, Berlin, Germany (2006)

[24] Larrea, C., Torres, F., López, N., Rueda, M.: Pueblos Indígenas, Desarrollo Humano y Discriminación en el Ecuador. Abya Yala, Quito, Ecuador (2007)

[25] Eurostat: Algorithms to compute social inclusion indicators based on EU-SILC and adopted under the Open Method of Coordination (OMC). Doc. LC-ILC/39/09/EN-rev.1, Unit F-3: Living conditions and social protection, Directorate F: Social and information society statistics. Eurostat, Luxembourg (2009)

[26] Hartley, D. (2016) Poverty and social exclusion. In: Platt, Lucinda and Dean, Hartley, (eds.) Social Advantage and Disadvantage. Oxford University Press, Oxford, UK, pp. 3-24. ISBN 9780198737070

[27] Heckathorn, D.: Respondent-driven sampling: A new approach to the study of hidden populations. Soc. Probl. **44**, 174â€“199 (1997)

[28] Martínez, S, Illescas, M., Martínez, H., Arcos, A.: Calibration estimator for Head Count Index. Int. J. Comput. Math. **97**, 51–62 (2020)

[29] Martínez, S. Rueda, M., Illescas, M.: The optimization problem of quantile and poverty measures estimation based on calibration. J. Comput. Appl. Math. **113054** (2020)

[30] Mullo, H.S., Sánchez-Borrego, I., Pasadas, S.: Respondent-driven sampling for surveying ethnic minority in Ecuador. Sustainability **12**, 9102 (2020)

[31] Rueda, M., Martínez, S., Illescas, M.: Treating nonresponse in the estimation of the distribution function Math Comput. Simulat. **186**, 3 (2020)

[32] Salganik, M., Heckathorn, D.: Sampling and estimation in hidden populations using respondent-driven sampling. Sociol. Methodol. **34**, 193â€“240 (2004)

[33] Uquillas, J., Carrasco, T., Rees, M.: Exclusión Social y Estrategias de vida de los Indígenas Urbanos en Perú, México y Ecuador (2003) Available online:

[34] Volz, E., Heckathorn, D.: Probability based estimation theory for respondent driven sampling. J. Off. Stat. **14**, 79â€“97 (2008)

[35] Hastings, W.K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika.* **1970**, *57*, 97â€“109.

[36] Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. The Monte Carlo Method. *Journal of Chemical Physics.* **1953**, *21*, 21, 1087.

[37] Hansen, H.; Hurwitz, W. On the Theory of Samplingfrom Finite Populations. *Annals of Mathematical Statistics.* **1943**, *14(4)*, 333–62.

[38] Abdesselam, K.; Verdery, A.; Pelude, L.; Dhami, P.; Momoli, F.; Jolly, A. The development of respondent-driven sampling (RDS) inference: a systematic review of the population mean and variance estimates. *Drug and alcohol dependence.* **2020**, *206*, 107702.

[39] Cochran, W. *Sampling Techniques*, 3d ed. New York: Wiley, 1977.

[40] Salganik, M. Variance estimation, design effects, and sample size calculations for respondent-driven sampling. J. Urban Health. **83**, 1 (2006), 10.1007/s11524-006-9106-x

[41] Foster, J.; Greer, J.; Thorbecke, E. A Class of Decomposable Poverty Measures. Econometrika **1984**, *52* 3 761-765.

[42] Sen, A.Poverty: An Ordinal Approach to Measurement. Econometrika **1976**, *44* 2, 219-231.

[43] Gile, K. Improved Inference for Respondent-Driven Sampling Data With Application to HIV Prevalence Estimation. Journal of the American Statistical Association. **106:493**, 135-146 (2011), DOI: 10.1198/jasa.2011.ap09475

[44] Yamanis, T.J.; Merli, M.G.; Neely, E.E.; Tian, F.F.; Moody, j.; Gao, E. An empirical analysis of the impact of recruitment patterns on RDS estimates among a socially ordered population of female sex workers in China. Sociological Methods and Res. **42**, 3 (2013), doi:10.1177/0049124113494576

[45] Verdery, A.M.; Mouw, T.; Bauldry, S.; Mucha, P.J. Correction: network structure and biased variance estimation in respondent driven sampling. PloS One. **11**, 1 (2016), 10.1371/journal.pone.0148006

[46] Baraff, A.J.; McCormick, T.H.; Raftery, A.E. Estimating uncertainty in respondent-driven sampling using a tree bootstrap method. Proc. Natl. Acad. Sci. U. S. A. **113**, 51 (2016), 10.1073/pnas.1617258113

[47] Rohe, K. A critical threshold for design effects in network sampling. Ann. Statist. **556-582** (2019), 10.1214/18-AOS1700

[48] Rotondi, M.A.; Oâ€™Campo, P.; Oâ€™Brien, K. Our Health Counts Toronto: using respondent-driven sampling to unmask census undercounts of an urban indigenous population in Toronto, Canada. *BMJ Open* **2017**, *7*, e018936, doi:10.1136/bmjopen-2017-018936.

[49] Neely, W. (2009). Statistical theory for respondent-driven sampling (Doctoral dissertation, University of Wisconsin–Madison). Available:

[50] Silva, P. L. D., Skinner, C. J. (1995). Estimating distribution functions with auxiliary information using poststratification. Journal of Official Statistics, 11 (3), 277–294.

[51] Rao, J. N. K., Kovar, J. G., Mantel, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. Biometrika, 77 (2), 365–375.

[52] Kuo, L. (1988) Classical and prediction approaches to estimating distribution functions from survey data. In *Proc. Survey Research Methods Section* 280-285. Amer. Statist. Assoc., Alexandria, VA.

[53] Kuk, A. (1993) A kernel method for estimating finite population distribution functions using auxiliary information. *Biometrika* **80**, 385–392.

[54] Wang, S. and Dorfman, A. H. (1996), A new estimator for the finite population distribution function. *Biometrika* **83**, 639-652.

[55] Woodruff, R. S. (1952) Confidence intervals for medians and other position measures, *Journal of the American Statistical Association* **47**, 635–646.

[56] Rueda, M., Martínez, S., Martínez, H., Arcos, A. (2007). Estimation of the distribution function with calibration methods. J. Stat. Plan. Inference, 137(2), 435-448. https://doi.org/10.1016/j.jspi.2005.12.011

[57] Singh, H. P., Singh, S., Kozak, M. (2008). A family of estimators of finite-population distribution function using auxiliary information. Acta applicandae mathematicae, 104(2), 115-130.

[58] Wu, C. (2003). Optimal calibration estimators in survey sampling. Biom., 90(4), 937-951. https://doi.org/10.1093/biomet/90.4.937

[59] Arcos, A., Martínez, S., Rueda, M., Martínez, H. (2017) Distribution function estimates from dual frame context, J. Comput. Appl. Math., 318, 242–252. https://doi.org/10.1016/j.cam.2016.09.027

[60] Martínez, S., Rueda, M., Arcos, A., Martínez, H. (2010). Optimum calibration points estimating distribution functions. J. Comput. Appl. Math., 233(9), 2265-2277. https://doi.org/10.1016/j.cam.2009.10.011

[61] Martínez, S., Rueda, M., Arcos, A., Martínez, H., Munoz, J. F. (2012). On determining the calibration equations to construct model-calibration estimators of the distribution function. Rev. Mat. Complut., 25(1), 87-95. https://doi.org/10.1007/s13163-010-0058-z

[62] Martínez, S., Rueda, M., Martínez, H., Arcos, A. (2017). Optimal dimension and optimal auxiliary vector to construct calibration estimators of the distribution function. J. Comput. Appl. Math., 318, 444-459. https://doi.org/10.1016/j.cam.2016.02.002

[63] Martínez, S., Rueda, M., Illescas, M. (2020). The optimization problem of quantile and poverty measures estimation based on calibration. Journal of Computational and Applied Mathematics, 113054.

[64] Martínez, S.,Illescas, M., Martínez, H. Arcos, A.(2020). Calibration estimator for Head Count Index, International Journal of Computer Mathematics, 97(1-2), 51-62

[65] Mayor-Gallego, J. A., Moreno-Rebollo, J. L., Jiménez-Gamero, M. D. (2019). Estimation of the finite population distribution function using a global penalized calibration method. AStA Advances in Statistical Analysis, 103 (1), 1–35.

[66] Deville,J.C., Särndal, C.E.(1992). Calibration estimators in survey sampling, J. Am. Stat. Assoc. 87(418), 376-382. https://doi.org/10.1080/01621459.1992.10475217

[67] Chambers, R. L., Dunstan, R. (1986). Estimating distribution functions from survey data. Biom., 73 (3), 597-604. https://doi.org/10.1093/biomet/73.3.597

[68] Harms, T., Duchesne, P. (2006). On calibration estimation for quantiles, Surv. Methodol., 32, 37–52.

[69] Francisco, C. A., & Fuller, W. A. (1991). Quantile estimation with a complex survey design. The Annals of statistics, 454-469.

[70] Woodruff, R. S. (1952) Confidence intervals for medians and other position measures, *Journal of the American Statistical Association* **47**, 635–646.

[71] Karl Rohe. A critical threshold for design effects in network sampling. The Annals of Statistics, 47(1) 556-582. 2019.

[72] Deville, J. C., and Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association* 87(418), 376–382.