Revista Colombiana de Estadística
Diciembre 2011, volumen 34, no. 3, pp. 451 a 460

The Generalized Logistic Regression Estimator in a Finite Population Sampling without Replacement Setting with Randomized Response

El estimador de regresión logística generalizado en un muestreo sin reemplazo con respuesta aleatorizada en poblaciones finitas

Víctor Hugo Soberanis-Cruz[a], Víctor Miranda-Soberanis[b]

División de Ciencias e Ingeniería, Universidad de Quintana Roo, Quintana Roo, México

Abstract

The randomized response technique (RR), introduced by Warner (1965) was designed to avoid non-answers to questions about sensitive issues and protect the privacy of the interviewee. Some other randomized response techniques have been developed as the Mortons technique which was developed based on a finite population sampling without replacement. In this paper we are presenting an estimation of the population (total of individuals N) based on Mortons technique assisted for a logistic regression model and considering a specific sensitive characteristic A, with an auxiliary variable associated to the sensitive variable. Analyses were conducted assuming finite population sampling and based on the p-estimators theory through a model assisted estimator. In addition, we propose an estimator of the variance of the estimator, as well as the results of simulations showing that the model assisted estimator of the variance decreases compared with an estimator which depends of the sampling design.

***Key words*:** Model assisted inference, Randomized response, Sampling design, Sensitive question.

Resumen

La técnica de respuesta aleatorizada (RA) introducida por Warner (1965), fue diseñada para disminuir la no-respuesta sobre aspectos sensibles y para proteger la confidencialidad del entrevistado en muestreos con reemplazo. Otras técnicas RA para muestreos sin reemplazo en poblaciones finitas, como la de Morton, han sido desarrolladas y comparadas. En este trabajo se exponen los resultados de la estimación del total de individuos de una población

[a]Professor. E-mail: vsobera@uqroo.mx
[b]Professor. E-mail: vmsoberanis@gmail.com

finita con la técnica de Morton, considerando una característica específica sensitiva A en un muestreo sin reemplazo y asistido por un modelo de regresión logística, con una variable auxiliar asociada a la variable sensitiva. Se desarrolla en el contexto de poblaciones finitas y en el marco de la teoría de los estimadores-π, a través de un estimador asistido por el modelo. Asimismo, se propone un estimador para la varianza del estimador y se muestra, vía simulación, que este estimador para la varianza disminuye, comparado con otro estimador que depende únicamente del diseño de muestreo.

***Palabras clave*:** diseño de muestreo, inferencia asistida por modelo, pregunta sensitiva, respuesta aleatorizada.

1. Introduction

Surveys represent procedures used by researchers to obtain information about a sample of individuals. Sometimes surveys include one or more questions related to personal information that could be considered as "private" and cause the individual to feel at risk (Méndez, Eslava & Romero 2004), and therefore, the individual refuses to participate or provides untrue responses.

When interviewers try to obtain honest responses, in studies where some sensitive issues, such as drug use, tax evasion, or sexual preferences, through survey sampling, they may face difficulties that intrinsically belong to the interviewee: attitude, time available, a different way of thinking, among others. (Sánchez 1985).

Strategies to minimize resistance from the interviewee to provide the real response, when the topic might represent an invasion of privacy are classified into two types. The first strategy is based on the phrasing of the questions that contain the characteristic that wants to be measure in such a way that indirect questions are used to obtain the real response. The second strategy refers to the method of randomized response (RR), introduced by Warner (1965). The RR is a specially designed method to ensure the privacy of the interviewee when sensitive, delicate, or embarrassing topics are studied. With these two strategies, the researcher avoids that the behavior of those surveyed gets skewed toward socially desirable responses. In this regard, real responses are obtained about sensitive issues (true/false) while assuring the confidentiality of the responses. In other words, the interviewer will not know the real answer.

Warner (1965) developed a technique called Randomized Response that guarantees the anonymity of the interviewee. It consists of a random mechanism that selects one of two complementary questions, as follows: Question (1): "do you have a specific characteristic A?" whereas Question (2) is "do you have the complementary characteristic?" where A represents the sensitive characteristic of interest and, the absence of such characteristic (Estevao, Hidiroglou & Sarndal 1999). The interviewee will provide the answer (yes or no), however, the interviewer will not know which question is answered. In this way, the anonymity of the interviewee is protected. This technique is also known as the Complementary Question Model.

As an alternative to Warner's Model of Complementary Questions Greenberg, Abul-Ela & Horvitz (1969) proposed a Model of Randomized Responses that contain two unrelated questions. One question addresses the sensitive issue of interest and the second one is innocuous. In other words, the second question is non-sensitive.

Horvitz, Greenberg & Abernathy (1976) attributed to R. Morton the idea about the Randomized Response Technique in which the random selection of unrelated questions are made among three options: (1) the sensitive question itself, (2) an instruction that indicates "yes", and (3) an instruction that indicates "no", that can be chosen with their respective probabilities P_1,P_2,P_3, where $P_1 + P_2 + P_3 = 1$. In this paper, we will call this model MU (Morton Unrelated), in allusion to R. Morton who had the original idea.

In order to understand the objective of this paper, we will consider a finite population of N units with a sensitive characteristic y_k, $(k = 1, 2, \dots, N)$ in which the total, $T_y = y_1 + y_2 + \cdots + y_N$ wants to be estimated. The main objective of this paper is to obtain better estimations of T_y (Soberanis, Ramírez, Pérez & González 2008) through Sampling without Replacement and the design of Morton Unrelated (MU), exploring estimators assisted by the Generalized Logistic Regression Estimator (GLRE) proposed by Lehtonen & Veijanen (1998*a*). In addition, the standard deviations of the estimators will be compared through simulation and recommendations will be provided. Thus, in this paper, an estimator of T_y will be proposed using a super population model: The Model of Logistic Regression.

2. The Generalized Logistic Regression Model

Let $U = 1, \dots, k, \dots, N$ be a finite population of participants. The subset $A \subset U$ is defined by a sensitive characteristic A; therefore, the RR technique is used to protect the anonymity of the sample of individuals (Soberanis et al. 2008). We will estimate $T_A = \sum\limits_U y_k$ where $y_k = 1$ if $k \in A$, and $y_k = 0$ if $k \notin A$. The selection of sample S is conducted under the sampling design $p(s)$ with positive inclusion probabilities π_k and π_{kl}, where

$$\pi_k = Pr(S \ni k) = \sum_{S \ni k} p(s) \text{ y } \pi_{kl} = Pr(S \ni \ k \,\&\, l) = \sum_{S \ni \, k \,\&\, l} p(s)$$

The estimator GLRE is generated with the predicted values obtained through the adjustment of the following model (Estevao, Hidiroglou & Sarndal 1995): We need to estimate the total population $T = \sum_U y_k$. A sample S is obtained assigning to unit k the sampling weight $a_k = \frac{1}{\pi_k}$. $\underline{x}$ represents an auxiliary vector of dimension $J \geq 1$, and $\underline{x}_k$ represents the a *priori* value for unit $k \in U$. In this method, data $\{(y_k, \underline{x}_k) : k \in s\}$ are observed. For those units that are not included in the sample, $y - k$ is unknown but it is possible to obtain a value μ_k that approximates y_k for all population units even though the approximation is not the most precise. Now, let

$$T_A = \sum_U \mu_k + \sum_U (y_k - \mu_k)$$

where the sum $\sum_U \mu_k$ is the dominante term and the residual sum $\sum_U (y_k - \mu_k)$, even though it is small, it needs to be estimated. Let us assume that $\underline{x}_k$ is situated in the sampling frame for all $k \in s$. The predicted values $\widehat{y}_k$ are obtained from the supplementary information adjusting a model in such a way that $E_\xi(Y_k \mid \underline{x}_k; \underline{\beta}) = \mu(\underline{x}_k \mid \underline{\beta})$, where E_ξ is the expectation operator under the theoretical model $\xi, \mu(\underline{x}_k \mid \underline{\beta})$. In addition $\xi, \mu(\underline{x}_k \mid \underline{\beta})$ is a specific function, and $\underline{\beta}$ is an unknown vector of parameters in the model. The function of model ξ is to describe the elements in the population in a "reasonable" way as if they would have been generated by the model itself.

However, it is not expected that the population was generated by the model ξ, therefore, conclusions about population parameters, including $\widehat{T}_A$, are independent from assumptions underlying the model.

In this manner, using the data from sampling$\{(y_k, \underline{x}_k : k \in s\}$, $\widehat{\underline{\beta}}$ is obtained as the maximum likelihood pseudo-estimator (MLEP) of $\underline{\beta}$, as it includes the sampling weights. In addition, the predicted values $\widehat{y}_k = \mu(\underline{x}_k \mid \widehat{\underline{\beta}}) = \widehat{\mu}_k$ are calculated, for each $k \in U$. Further, using $\widehat{\mu}$ and the Horvitz-Thompson estimator (HT) for the residual sum, we obtain:

$$\widehat{T}_{LGREG} = \sum_U \widehat{\mu}_k + \sum_s \pi_k^{-1}(y_k - \widehat{\mu}_k) \tag{1}$$

Equation (1) is the GLRE from Lehtonen & Veijanen (1998a).

In practice, model ξ works as a way to find $\widehat{\beta}$, in order to use it in the estimation functions.

2.1. The LGREG for Model MU

The Generalized Logistic Regression Estimator $\widehat{T}_{A,LGREG}$ for $\widehat{T}_A$, proposed previously, is an application of the Lehtonen & Veijanen (1998a), which, as we have mentioned before, is an estimator assisted for a Logistic Regression Model. In other words, for $\underline{y} = (y_1, \dots, y_k, \dots, y_N)^t$, the following model is suggested:

$$Pr\{Y_k = 1 \mid \underline{x}_k; \underline{\beta}\} = \frac{e^{\underline{x}_k^t \underline{\beta}}}{1 + e^{\underline{x}_k^t \underline{\beta}}},\ k = 1, 2, \dots, N \tag{2}$$

From now on, this super population model will be referred as model ξ. In addition, for Morton's (MU) Random Mechanism (RR), it is defined:

$$Z_k = \begin{cases} y_k & \text{with probability } P_1 \\ 1 & \text{with probability } P_2 \\ 0 & \text{with probability } 1 - P_1 - P_2 \end{cases} \tag{3}$$

$k = 1, 2, \dots, N$. Thus,

$$\begin{aligned} E(Z_k) &= E_\xi E_{RC}(Z_k) \\ &= E_\xi[P_1 y_k + P_2] \\ &= P_1 E_\xi(y_k) + P_2 \\ &= P_1 \mu_k + P_2 \end{aligned} \tag{4}$$

where

$$\mu_k = E_\xi(Y_k) = Pr\{Y_k = 1 \mid \underline{x}_k; \underline{\beta}\} = \frac{e^{\underline{x}_k^t \underline{\beta}}}{1 + e^{\underline{x}_k^t \underline{\beta}}} \tag{5}$$

Therefore, if $\lambda_k = E(Z_k) = Pr\{Z_k = 1 \mid \underline{x}_k; \underline{\beta}\}$, we obtain:

$$\lambda_k = P_1 \mu_k + p_2 \tag{6}$$

Now,

$$t_A = \sum_U y_k = \sum_U \mu_k + \sum_U (y_k - \mu_k) \tag{7}$$

and,

$$\widehat{T}_A = \sum_U \widehat{\mu}_k + \sum_s \frac{(y_k - \widehat{\mu}_k)}{\pi_k} \tag{8}$$

where

$$\widehat{\mu}_k = \mu(\underline{x}_k^t \underline{\widehat{B}}) = \frac{e^{\underline{x}_k^t \underline{\widehat{B}}}}{1 + e^{\underline{x}_k^t \underline{\widehat{B}}}} \tag{9}$$

Thus, $\underline{\widehat{\beta}}$, satisfies the following equation

$$\sum_S \left[y_k - \mu(\underline{x}_k^t \underline{B} \right] \frac{\underline{x}_k}{\pi_k} = \underline{0}$$

where $\underline{\beta}$ represents the population parameter defined for the likelihood equation:

$$\partial \log L(\underline{\beta}) / \partial \underline{\beta} = \underline{0}$$

which is equivalent to the following equation

$$\sum_U \left[y_k - \mu\big(\underline{x}_k^t \underline{B}\big) \right] \underline{x}_k = \underline{0}$$

For practical purposes, we will use either $\underline{\beta}$ or $\underline{B}$.

For the open sampling, the likelihood function $L(\underline{\beta})$ is given by:

$$L\big(\underline{\beta} \mid \underline{y}\big) = \prod_{k \in A} \mu\big(\underline{x}_k^t \underline{\beta}\big) \prod_{k \in U - A} \left[1 - \mu\big(\underline{x}_k^t \underline{\beta}\big) \right]$$

Regarding the Random Responses problem, the vector of observations is $\underline{Z}_s = (Z_k)_{k \in s}$ its parameter, $\underline{\lambda} = (\lambda_k)_{k \in s}$, and the likelihood function is given by

$$\begin{aligned} L(\underline{\beta} \mid \underline{z}) &= \prod_U Pr\{Z_k = z_k\} \\ &= \prod_U \lambda_k^{z_k} (1 - \lambda_k)^{1 - z_k} I_{\{0,1\}}(z_k) \end{aligned}$$

as if $(Z_k, \underline{x}_k)$ was observed for each $k \in U$, as in a census. Thus,

$$\begin{aligned} l(\underline{\beta} \mid \underline{z}) &= \ln L(\underline{\beta} \mid \underline{z}) \\ &= \sum_U \Big[z_k \ln \lambda_k + (1 - z_k) \ln \{(1 - \lambda_k)\Big] \end{aligned}$$

It should be noted that the function $l(B \mid z)$ reaches a maximum in $\underline{B}$, and is defined and characterized by the following equation:

$$\begin{aligned} \frac{\partial l(\underline{B} \mid \underline{Z})}{\partial \underline{B}} &= \underline{0} \Leftrightarrow \sum_U \left[\Big(z_k - (P_1\mu_k + P_2) \Big) \frac{\mu_k(1-\mu_k)}{P_2(1-P_2) + P_1\mu_k(1 - P_1\mu_k - 2P_2)} \right] \underline{x}_k \\ &= \underline{0} \end{aligned} \tag{10}$$

Thus, $\underline{B}$ is defined implicitly as the one parameter that maximizes $l(B \mid Z)$.

Also, the estimator π of $l(\underline{B} \mid \underline{z})$ is given by:

$$\widehat{I}_\pi(\underline{\beta} \mid \underline{z}) = \sum_s \pi_k^{-1} \Big[z_k \ln \lambda_k + (1 - z_k) \ln (1 - \lambda_k) \Big]$$

Where λ_k is given by (6).

In addition,

$$\frac{\partial \lambda_k}{\partial \underline{\beta}} = P_1 \frac{\partial \mu_k}{\partial \underline{\beta}} = P_1 \mu_k (1 - \mu_k) \underline{x}_k$$

thus,

$$\begin{aligned} \frac{\partial \widehat{I}_\pi}{\partial \underline{\beta}} &= \sum_s \pi_k^{-1} \left[\frac{z_k}{\lambda_k} \frac{\partial \lambda_k}{\partial \underline{\beta}} - \Big(\frac{1 - z_k}{1 - \lambda_k} \Big) \frac{\partial \lambda_k}{\partial \underline{\beta}} \right] \\ &= P_1 \sum_s \pi_k^{-1} \left[\frac{(z_k - \lambda_k)\mu_k(1 - \mu_k)}{\lambda_k(1 - \lambda_k)} \right] \underline{x}_k \end{aligned}$$

but

$$\frac{\mu_k(1-\mu_k)}{\lambda_k(1-\lambda_k)} = \frac{\mu_k(1-\mu_k)}{P_2(1-P_2) + P_1\mu_k(1 - P_1\mu_k - 2P_2)}$$

then,

$$\begin{aligned} \frac{\partial \widehat{I}_\pi}{\partial \underline{\beta}} &= P_1 \sum_s \pi_k^{-1} \left[(z_k - \lambda_k) \frac{\mu_k(1-\mu_k)}{P_2(1-P_2) + P_1\mu_k(1 - P_1\mu_k - 2P_2)} \right] \underline{x}_k \\ &= P_1 \sum_s \pi_k^{-1} \left[\Big(z_k - (P_1\mu_k + P_2) \Big) \frac{\mu_k(1-\mu_k)}{P_2(1-P_2) + P_1\mu_k(1 - P_1\mu_k - 2P_2)} \right] \underline{x}_k \end{aligned}$$

Therefore, by solving the following equation $\dfrac{\partial \widehat{I}_\pi}{\partial \underline{\beta}} = \underline{0}$, we obtained $\widehat{\underline{\beta}}$.

Once $\widehat{\underline{\beta}}$ is obtained, the estimator proposed for T_A is

$$\widehat{T}_{A,LGREG} = \sum_U \widehat{\mu}_k + \frac{1}{P_1} \sum_s \frac{Z_k - (P_1\widehat{\mu_k} + P_2)}{\pi_k} \tag{11}$$

where $\widehat{\mu}_k$ is given by (9).

2.2. Estimation of the Estimator Variance

Base on the π estimators theory (Sarndal, Swensson & Wretman 1992, Wretman, Sarndal & Cassel 1977, Lehtonen & Veijanen 1998*a*), the following estimator is proposed for $Var(\widehat{T}_{A,LGREG})$:

$$\widehat{V}(\widehat{T}_{A,LGREG}) = \left(\frac{1}{P_1^2}\right) \sum\sum_{s} \frac{\Delta_{kl}}{\pi_{kl}} \left(\frac{Z_k - \widehat{\lambda}_k}{\pi_k}\right)\left(\frac{Z_1 - \widehat{\lambda}_1}{\pi_1}\right) \tag{12}$$

3. An Estimator that only Depends on the Sampling Design

Based on Morton's random response technique and a sampling design $p(s)$, a π-estimador for t_A, is given by Soberanis et al. (2008):

$$\widehat{T}_{A,\pi} = \frac{1}{P_1} \sum_{S} \frac{Z_k}{\pi_k} - \frac{NP_2}{P_1} \tag{13}$$

Its variance is given by

$$V(\widehat{T}_{A,\pi}) = \frac{1}{P_1^2}\left\{ \sum\sum_{U} \Delta_{kl} \widehat{\lambda_k}\widehat{\lambda_1} + \sum_{U} \frac{\lambda_k(1-\lambda_k)}{\pi_k} \right\}$$

Where $\widehat{\lambda} = \lambda_k/\pi_k$. So, the estimator proposed for the variance of the estimator is:

$$\widehat{V}(\widehat{T}_{A,\pi}) =$$
$$\frac{1}{P_1^2}\left\{ \sum\sum_{s} \breve{\Delta}_{kl} \left(\frac{Z_k}{\pi_k}\right)\left(\frac{Z_1}{\pi_l}\right) + P_1(1-P_1-2P_2)\sum_{s} \frac{\widehat{Z_k}}{\pi_k^2} + P_2(1-P_2)\sum_{U}\frac{1}{\pi_k} \right\}$$

Where $\widehat{Z}_k = \dfrac{Z_k - P_2}{P_1}$ y $\breve{\Delta}_{kl} = \dfrac{\pi_{kl} - \pi_k\pi_l}{\pi_{kl}}$

4. Simulations using Simple Random Sampling with No-Replacement

This section analyzes the properties of the estimator (11) in the specific case of Simple Random Sampling with No-replacement (SRSN). If the sampling design, $p(s)$, is the SRSN, then,

$$\pi_k = \frac{n}{N} = f; \quad \pi_{kl} = \frac{n(n-1)}{N(N-1)}, \, k \neq 1; \quad \pi_{kk} = \pi_k$$

$$\Delta_{kl} = \pi_{kl} - \pi_k\pi_l = \frac{-f(1-f)}{N-1}, k \neq 1; \quad \Delta_{kk} = \pi_k(1-\pi_k) = f(1-f) \tag{14}$$

5. Comparison of the Estimators $\widehat{T}_{A,\pi}$ and $\widehat{T}_{LGREG,\pi}$

In order to compare the estimators $\widehat{T}_{A,\pi}$ and $\widehat{T}_{LGREG,\pi}$, a population with $B = (B_0, B_1)' = (-3, 0.1)'$ with $A = 490$ "successes" using the Accept-Reject algorithm to generate random variables.

5.1. Simulations Results

Table 1: Mean, minimum, maximum and percentiles of estimators $\widehat{T}_{A,\pi}$ and $\widehat{T}_{A,LGREG}$, using $N = 700$, $A = 490$, $n = 140$ and $N = 800$ (Number of simulations).

					Percentiles		
Estimator	Mean	DE	Minimun	25%	50%	75%	Maximum
$\widehat{T}_{A,\pi}$	490.164	39.04	342.857	464.285	492.857	514.285	621.428
$\widehat{T}_{A,LGREG}$	489.818	36.22	383.185	466.997	491.953	513.498	594.419

6. Benefits of the Estimator's Variance of $\widehat{T}_{A,LGREG}$

For the simulated population, Section 6.1, shows that

$$\widehat{V}(\widehat{T}_{A,LGREG}) = \left(\frac{1}{P_1^2}\right) \sum\sum_{s} \frac{\Delta_{kl}}{\pi_{kl}} \left(\frac{Z_k - \widehat{\lambda}_k}{\pi_k}\right)\left(\frac{Z_l - \widehat{\lambda}_l}{\pi_l}\right)$$

is an excellent estimator of the variance of $\widehat{T}_{A,LGREG}$.

6.1. Simulation Results for the Estimators Variance of $\widehat{T}_{A,LGREG}$

According to Table 2, standard deviation of $\widehat{T}_{A,GLRE}$ is 36.22684, whereas our estimator given by (12) is, on average, 33.601760 and standard deviation 1.344364.

Table 2: Mean, minimum, maximum, and percentiles of the variance of the estimator $\widehat{T}_{A,LGREG}$ ($N = 700$).

					Percentiles		
Estimator	Mean	DE	Minimum	25%	50%	75%	Maximum
$\widehat{V}(\widehat{T}_{A,LGREG})$	33.601	1.344	29.431	32.717	33.660	34.591	38.248

7. Discussion and Conclusions

7.1. Discussion

This paper focuses on the use of auxiliary variables as well as on models for the random response sampling (RR) in finite populations, i.e. in populations where all observation units are identifiable. Furthermore, the sampling designs used were sampling designs with no replacement due to the sensitivity of the variable of interest.

The proposal is based on the work of Lehtonen & Veijanen (1998*a*), Lehtonen & Veijanen (1998*b*), Estevao et al. (1999), which was developed for sampling in finite populations. In other words, for a type of sampling where the researcher is focused on exhaustive data collection until the process to delimit and define the fundamental variables occurs constantly (Pandit 1996, Goulding 2002).

The use of auxiliary variables in a conventional way, i.e. when the variable of interest is correlated with the auxiliary variable in the context of RR, does not necessarily improve the Simple Random Sampling, at least for the Rao-Hartley-Cochran scheme. This happens because the auxiliary information is not used properly, as the variable of interest is a discrete variable. In fact, the proper use of the variable of interest is through a model that assists in the problem of estimating the population total, hence the use of the Generalized Logistic Regression Model.

7.2. Conclusions

If the logistic regression model describes the population adequately, then the estimators GLRE and MU should be used. The results suggest that by using this method, we will obtain a significant reduction in the estimator's variance. It should be noted that it is not necessary that the model is "true", as it represents a process in which the population being studied is generated. However, additional simulations under different conditions should be done in order to compare them to the results obtained in this paper. Only then specific recommendations on the most appropriate approach can be provided.

[Recibido: octubre de 2010 — Aceptado: julio de 2011]

References

Estevao, V., Hidiroglou, M. A. & Sarndal, C. E. (1995), 'Methodological principles for a generalized estimation system at Statistics Canada', *Journal of Official Statistics* **11**, 181–204.

Estevao, V., Hidiroglou, M. A. & Sarndal, C. E. (1999), 'The use of auxiliary information in design-based estimation for domains', *Survey Methodology* **2**, 213–221.

Goulding, C. (2002), *Grounded Theory, A Practical Guide for Management, Business and Market Researchers*, SAGE Publications Ltd.

Greenberg, B. G., Abul-Ela, Abdel-Latif, A. S. W. R. & Horvitz, D. C. (1969), 'The unrelated question RR model: Theoretical framework', *Journal of the American Statistical Association* **64**, 520–539.

Horvitz, D. C., Greenberg, B. G. & Abernathy, J. R. (1976), 'RR: a data gathering device for sensitive questions', *International Statistical Review* **44**, 181–196.

Lehtonen, R. & Veijanen, A. (1998*a*), 'Logistic generalized regression estimators', *Survey Methodology* **24**, 51–55.

Lehtonen, R. & Veijanen, A. (1998*b*), 'On multinomial logistic generalized regression estimators', *Survey Methodology* **24**(1), 51–55.

Méndez, I., Eslava, G. & Romero, P. (2004), 'Conceptos básicos de muestreo', *Monografías* **12**(27), 10–58.

Pandit, N. R. (1996), 'The creation of theory: A recent application of the grounded theory method', *The Qualitative Report* **2**(4).

Sarndal, C. E., Swensson, B. & Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer Verlag, New York.

Sánchez, J. L. (1985), 'El tratamiento de preguntas de carácter íntimo: modelo de respuesta aleatorizada', *Revista Estadística Española* **65**, 5–12.

Soberanis, V. H., Ramírez, G., Pérez, S. & González, F. V. (2008), 'Muestreo de respuestas aleatorizadas en poblaciones finitas: Un enfoque unificador', *Agrociencia* **42**, 537–549.

Warner, S. L. (1965), 'Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias', *Journal of the American Statistical Association* **60**, 63–69.

Wretman, J. K., Sarndal, C. E. & Cassel, C. M. (1977), *Foundations of Inference in Survey Sampling*, John Wiley and Sons, Inc., New York.