

Tratamiento de la no respuesta en encuestas de panel en el caso de poblaciones finitas.

Margarita Antía - Ana Coimbra

Este trabajo pretende mostrar cómo tratar la no respuesta en un caso particular de las encuestas por muestreo como son las encuestas de panel. Éstas refieren a estudios basados en observaciones repetidas efectuadas sobre las mismas unidades de muestreo: personas, hogares, empresas, etc. La muestra es extraída por única vez al inicio del estudio y todas las unidades seleccionadas serán visitadas a lo largo de la duración del panel.

Los distintos momentos del tiempo en los que las encuestas son llevadas a cabo se denominan “olas”; la duración del panel y el período entre olas son definidos en la etapa del diseño de la encuesta.

La medición periódica de elementos permite realizar un seguimiento de la población objetivo, logrando captar su dinámica en el tiempo. Los resultados particulares en cada instancia de medición (estimaciones transversales) pueden ser obtenidos sin perjuicio de lo anterior y, aunque no sea el objetivo principal de las encuestas de panel suelen ser de interés en sí mismos.

Un problema usual en las encuestas de panel es la mortalidad de unidades a lo largo del tiempo. En términos generales, esta dificultad puede pensarse como un problema de no respuesta, definida como la imposibilidad de obtener toda o alguna información para una o más de las unidades seleccionadas en la muestra. Este es un fenómeno presente en la mayoría de las encuestas por muestreo y es imprescindible su tratamiento para evitar sesgos en las estimaciones. La inclusión del factor “tiempo” en las encuestas de panel provoca un agravamiento del problema de no respuesta respecto a las encuestas cross-section, reflejado en reducciones considerables en el “tamaño de muestra” período a período debido a la movilidad,

el fallecimiento y otros factores (como la pérdida de cooperación de unidades) que resultan en el “agotamiento” del panel. Otro efecto causado por la inclusión del factor “tiempo” es la potencial pérdida de representatividad de la muestra para inferir resultados transversales en olas posteriores a la primera.

1. Estimador de cambio bajo condiciones ideales

Las condiciones ideales en cualquier tipo de encuesta por muestreo están regidas por la obtención de respuesta completa de las unidades muestreadas a partir de un marco muestral perfecto acorde a la población objetivo, sin presencia de errores de medición. En las encuestas por panel, dichas condiciones requieren el supuesto adicional de que la población objetivo sea fija en el tiempo. Adicionalmente se requiere obtener respuesta completa de todas las unidades en todas las olas. Esto es, s es una muestra aleatoria de la población U de N individuos, tomada bajo un diseño $p(s)$ ¹ de tamaño n_s el cual genera probabilidades de inclusión π_k .

Bajo estos supuestos, el estimador de cambios de la variable de interés es :

$$\begin{aligned}\hat{A}_{j,j+h} &= (\hat{t}_{j+h} - \hat{t}_j) \\ &= \sum_s \left(\frac{y_{(j+h)k} - y_{jk}}{\pi_k} \right) \\ &= \sum_s \frac{a_{(j,j+h)k}}{\pi_k}\end{aligned}\quad (1)$$

donde $a_{(j,j+h)k}$ es cambio individual del elemento k en la variable de interés entre las olas j y $j+h$.

Los estimadores $\hat{A}_{j,j+h}$ tienen la forma de estimadores π (Horvitz-Thompson), por lo tanto, también comparten sus propiedades. La varianza de $\hat{A}_{j,j+h}$ también puede expresarse como

$$Var_{p(s)}(\hat{A}_{j,j+h}) = Var(\hat{t}_j) + Var(\hat{t}_{j+h}) - 2Cov(\hat{t}_j, \hat{t}_{j+h}) \quad (2)$$

y dado que se estos totales están calculados en base a las mismas unidades, se espera los totales estimados en olas sucesivas estén correlacionados positivamente resultando en una varianza pequeña del estimador de diferencias.

¹La elección del diseño a utilizar depende del objeto de estudio y no de la utilización de paneles.

Una alternativa a la utilización de paneles es la estimación de diferencias de la variable de interés utilizando totales estimados mediante encuestas cross-section en los momentos j y $j + h$. En este caso el estimador de diferencias también será insesgado pero su varianza será mayor que en el caso anterior, ya que la independencia entre muestras en una y otra instancia determina que el tercer término de (2) sea cero.

Este es un aspecto importante en la justificación de utilización de paneles frente a encuestas cross-section para medir cambios.

2. Calibración como tratamiento de la no respuesta

El estimador propuesto tiene la cualidad de ser insesgado y de sencillo cálculo, ya que es un estimador π . Para su desarrollo, se partió de supuestos muy restrictivos rara vez presentes en la práctica, a saber: la existencia de *respuesta perfecta* y población fija en el tiempo, reflejados en ponderadores constantes en el tiempo para cada elemento e iguales al inverso de su probabilidad de inclusión en la muestra.

El levantamiento del supuesto de respuesta perfecta plantea la necesidad de estudiar cuales son los efectos que provoca la imposibilidad de obtener toda o alguna información para todas o algunas unidades seleccionadas en la muestra.

Si la no respuesta se presentara de manera completamente aleatoria, el único inconveniente al que se enfrenta el investigador es la reducción del tamaño de muestra y un consiguiente aumento en la varianza de las estimaciones, que podría ser fácilmente contrarrestado mediante un “sobremuestreo” (fijando un tamaño de muestra mayor en la etapa de diseño). De esta manera el único efecto negativo de la no respuesta sería el incremento en la carga administrativa y los costos de recolección de datos. En la práctica, la situación anterior sería una “feliz” casualidad. Las unidades que no contestan “normalmente” difieren en algunas características de aquellas que sí lo hacen, y el sesgo introducido en las estimaciones por esta causa constituye el obstáculo más importante por corregir. Frente a la pérdida del insesgamiento de los estimadores, el incremento de su varianza es un disturbio menor: en presencia de sesgo significativo, un intervalo de confianza calculado estará centrado en un valor erróneo y no se logra el nivel de confianza requerido.

Sea \hat{t}_y el estimador de t_y cuando hay respuesta completa, o sea cuando el conjunto de respondentes r coincide con la muestra s ; sea \hat{t}_{yNR} el estimador de t_y en presencia

de no respuesta. El *error de estimación* de \hat{t}_{yNR} puede expresarse mediante un término que representa el *error muestral* y otro que representa el *error por no respuesta*.

$$Error = \hat{t}_{yNR} - t_y = (\hat{t}_y - t_y) + (\hat{t}_{yNR} - \hat{t}_y) \quad (3)$$

Siendo

- $\hat{t}_y - t_y$ el error muestral (el error que surge por elegir y observar una muestra, en vez de observar toda la población).
- $\hat{t}_{yNR} - \hat{t}_y$ el error por no respuesta (error que surge por la no existencia de respuesta completa).

El *sesgo total de \hat{t}_{yNR}* se obtiene calculando el valor esperado bajo los mecanismos de selección y de respuesta de los dos componentes de error previamente definidos:

$$\begin{aligned} B_{pq}(\hat{t}_{yNR}) &= B_{SAM} + B_{NR} \\ &= E(\hat{t}_{yNR} - \hat{t}_y) + E(\hat{t}_y - t_y) \end{aligned} \quad (4)$$

El término B_{SAM} (sesgo muestral) es cero o irrelevante para la mayoría de los propósitos prácticos, por lo tanto el sesgo de \hat{t}_{yNR} se convierte casi enteramente en el sesgo por no respuesta. Esto evidencia la necesidad de realizar algún tipo de tratamiento en la etapa de estimación.

En general se distinguen dos tipos de no respuesta: no respuesta al ítem, que refiere a faltantes en la respuesta para un ítem en particular del formulario debido a omisión (tanto del entrevistador como del entrevistado) o negativa del encuestado a contestar; y no respuesta de la unidad, que se da cuando la unidad seleccionada para ser entrevistada no es encontrada o se rehúsa a participar en la encuesta.

Las técnicas dominantes en la literatura actual para el tratamiento de la no respuesta son la calibración y la imputación (Särndal y Lundström (2005)). Usualmente la calibración predomina en el tratamiento para el caso de no respuesta de unidades; mientras que la imputación es más extensamente aplicada en los problemas de no respuesta de ítems. La primera es una estrategia global, tratando todas las variables de forma simultánea, mientras que la segunda es particular, específica de cada variable. La decisión acerca de utilizar una u otra no es obvia, y depende de distintos factores como lo son: la cantidad y el número de olas, el tipo de análisis

a llevarse a cabo, la disponibilidad de variables auxiliares con poder predictivo de los valores faltantes y el costo de implementar los procedimientos.

No obstante lo anterior, se han ensayado soluciones que aplican ambas técnicas en conjunto (Deville y Särndal (1994)).

La imputación es el procedimiento a través del cual los valores faltantes en una o más variables de estudio se completan con sustitutos. Los valores perdidos en la base de datos se reemplazan por los valores “plausibles” dando como resultado una matriz completa de valores. Existen varios métodos de imputación que básicamente difieren en como definen “plausible”, pero la mayoría coinciden en la necesidad de utilización de información auxiliar.

La calibración, o más precisamente el uso de estimadores calibrados, se basa fuertemente en el uso de información auxiliar tanto a nivel poblacional como a nivel de la muestra original. Su creciente popularidad puede explicarse porque no se basa en la especificación de un modelo de no respuesta, brinda un enfoque unificado dentro de la teoría del muestreo de poblaciones finitas, es computacionalmente sencilla de implementar y generaliza otras técnicas del tratamiento de la no respuesta como la post-estratificación, el raking y algunos casos de los ajustes basados en la teoría del muestreo en dos fases. El enfoque de calibración abarca a una familia de estimadores \hat{t}_{yw} cuyos miembros corresponden a diferentes inputs de información.

Es deseable que los estimadores afectados por la no respuesta sean útiles para estimar los totales de las variables de interés, que no estén sesgados y que tengan varianza reducida. El sistema de ponderadores que surge del enfoque de calibración verifica también, que al ser aplicado a las variables auxiliares reproduce el input de información auxiliar. La idea central de los estimadores calibrados es sencilla, consiste en modificar los ponderadores originales de la muestra minimizando alguna función de distancia entre dichos ponderadores y los ponderadores finales (o calibrados) y de manera que estos últimos estimen sin error algunas cantidades conocidas con los datos de los respondentes.

Sea w_k el peso calibrado para $k \in r$, luego, el estimador de $t_y = \sum_U y_k$ es:

$$\hat{t}_{yw} = \sum_r w_k y_k \quad (5)$$

Se busca el conjunto de valores w_k para todo $k \in r$ que satisfaga la *ecuación de calibración*:

$$\mathbf{X} = \sum_r w_k \mathbf{x}_k \quad (6)$$

Se dice que estos pesos w_k están calibrados al input de información \mathbf{X} , ya que cuando se aplican al vector auxiliar \mathbf{x}_k reproducen exactamente la información dada en \mathbf{X} .

Como resultado de la selección de la muestra, a cada elemento k le corresponde el peso $d_k = \frac{1}{\pi_k}$. En presencia de no respuesta, $\sum_r d_k y_k$ subestima $\sum_U y_k$, en una magnitud $\sum_{s=r} d_k y_k$ (en el caso que la variable de interés tome solamente valores positivos). Es por esto que los d_k deben ser modificados. Se buscarán nuevos pesos que sean mayores que d_k al menos para la mayoría de los respondentes, de manera de compensar la pérdida de unidades. Los nuevos ponderadores $w_k = d_k v_k$ se obtienen “aumentando” los pesos originales mediante el factor v_k , que reflejará las características individuales conocidas de los elementos $k \in r$ (resumidas en el vector \mathbf{x}_k), y puede pensarse como una función del vector auxiliar $v_k = F(\lambda' \mathbf{x}_k)$, donde λ es un vector de la misma dimensión que \mathbf{x}_k y se determinará para que se verifique la ecuación de calibración.

La clave para una calibración exitosa es el uso de información auxiliar poderosa; permitiendo reducir tanto sesgo como la varianza. La efectividad del estimador de calibración para controlar el sesgo ocasionado por la no respuesta dependerá de propiedades del vector auxiliar. Särndal y Lundström (2005) realizaron un estudio de Simulación Monte Carlo a través del cual se obtiene evidencia empírica de la fuerte relación existente entre el sesgo del estimador de calibración que proviene de la no respuesta y la información auxiliar utilizada para calibrar. Éste será menor cuanto más estrecha sea la relación entre la información auxiliar y la probabilidad de respuesta o la variable de interés. Si el sesgo es modesto el intervalo de confianza será válido y la probabilidad de cobertura será cercana al nivel de confianza requerido. Limitar el sesgo de las estimaciones en presencia de no respuesta se tornará en la mayor preocupación, la minimización de la varianza pasará a segundo plano ya que de nada sirve que un estimador presente varianza chica cuando está fuertemente sesgado.

3. Estimadores calibrados en encuestas de panel

La no respuesta de unidades en una ola es una forma de no respuesta parcial particular al muestreo por paneles, generando distintos patrones de respuesta a lo largo del estudio, como muestra el siguiente cuadro. Algunos miembros de la muestra pueden abandonar la encuesta en cierta ola y perderse para el resto del estudio (desertores); mientras que otros pueden perderse en una ola, y volver al panel en alguna de las siguientes (respondentes episódicos).

Patrón	Estado de respuesta	Ola 1	Ola 2	Ola 3	Ola 4	Ola 5
1	Respondentes	x	x	x	x	x
2	No	x	x	x	x	-
3	Respondentes	x	x	x	-	-
4	por	x	x	-	-	-
5	Desgaste	x	-	-	-	-
6	No	x	x	-	x	x
7	Respondentes	x	-	-	x	x
8	Episódicos	x	-	-	-	x

Ref: x: respuesta, -: no respuesta

La no respuesta en panel (manifestada bajo los patrones de *no respuesta por desgaste* y *no respuesta episódica*) genera en cada ola, un conjunto de respondentes r_i , todos incluidos en la muestra s . Esto requerirá el cálculo de ponderadores calibrados particulares a cada individuo respondente en cada ola.

Estimación transversal

Para la estimación transversal se calcularán los ponderadores de las unidades respondentes en cada ola. El estimador calibrado del total correspondiente a la ola i , \hat{t}_{ywi} se define por:

$$\hat{t}_{ywi} = \sum_{r_i} w_{k_i} y_{k_i} \quad (7)$$

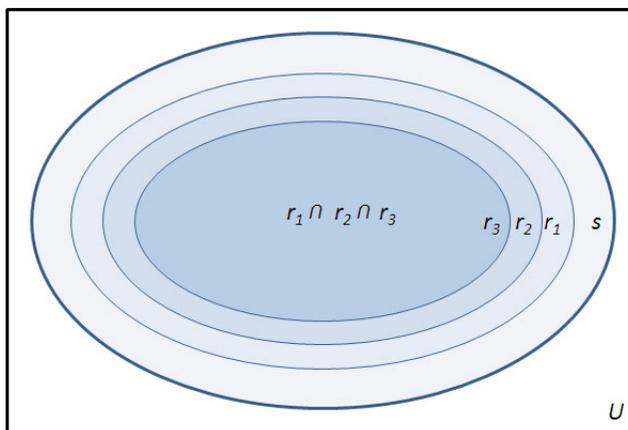
siendo w_{k_i} los ponderadores calibrados del elemento k respondente en la ola i , y y_{k_i} el valor de la variable de interés para este individuo en dicha ola.

Estimación longitudinal

Adicionalmente, para las estimaciones longitudinales será necesario el cálculo de un nuevo conjunto de ponderadores aplicables únicamente a las unidades respondientes en todas las instancias de las que se quiere medir el cambio: *respondentes simultáneos*.

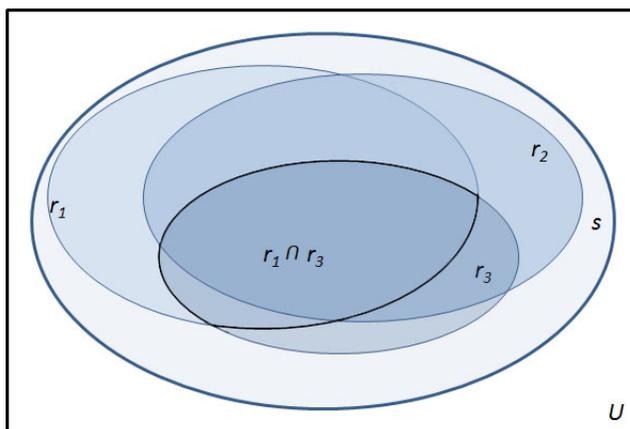
Cuando el patrón de respuesta admite únicamente no respuesta por desgaste, los respondientes de i -ésima ola también fueron respondientes en las olas anteriores $i - 1, i - 2, \dots, 1$, por lo tanto, los cambios solamente podrán ser medidos para las unidades respondientes en la ola más reciente. En la siguiente figura puede verse la representación gráfica de un estudio de panel de tres olas con patrón de respuesta por desgaste. A modo de ejemplo, si el interés radica en la estimación de cambios de una variable entre las olas 1 y 3, se calibrarán los cambios individuales para cada respondiente de la ola 3.

Figura 1: Conjunto de respondientes con patrón de respuesta: desgaste



Frente a un patrón de respuesta episódica, los cambios entre dos olas también serán medidos en los respondientes simultáneos, pero en este caso, este conjunto no necesariamente coincide con la ola más reciente de las sujetas a medición. Siguiendo el mismo ejemplo, la estimación de cambios entre la primera y la tercera ola se realizará en base al conjunto marcado en la figura (2).

Figura 2: Conjunto de respondentes con patrón de respuesta: episódica



Luego de definir el conjunto de respondentes simultáneos a las olas j y $j + h$, la estimación de las diferencias de la variable y entre dichas olas se obtiene mediante la siguiente fórmula:

$$\begin{aligned} \hat{A}w_{j,j+h} &= \sum_{r_j \cap r_{j+h}} (y_{(j+h)k} - y_{jk}) w_{(j,j+h)k}; \\ &= \sum_{r_j \cap r_{j+h}} a_{(j,j+h)k} w_{(j,j+h)k} \end{aligned} \quad (8)$$

Donde:

- $w_{(j,j+h)k}$ son los ponderadores obtenidos mediante algún método de calibración para las unidades respondentes en las olas j y $j + h$, y
- $a_{(j,j+h)k} = y_{(j+h)k} - y_{jk}$ es el cambio en la variable de interés y entre estas olas para cada elemento respondente $k \in \{r_j \cap r_{j+h}\}$.

La aplicación de la fórmula (8) en los patrones de respuesta episódicos y por desgaste difiere únicamente en la definición del conjunto de respondentes simultáneos $r_j \cap r_{j+h}$.

4. Aplicación: Las damas perdidas

Una aplicación concreta del uso de ponderadores calibrados para encuestas de panel se realizó para la “Encuesta sobre Situaciones Familiares y Desempeños Sociales en Montevideo y Área Metropolitana”, llevada a cabo por un equipo de investigadores de la Universidad de la República (Facultad de Ciencias Económicas y de Administración - Instituto de Economía y de la Facultad de Ciencias Sociales - Departamento de Economía y Programa de Población).

Este panel consiste en dos olas, la primera fue realizada entre marzo y octubre de 2001 a una muestra de 1806 mujeres y la segunda ola se realizó en el año 2008, logrando recontactar a 828 de ellas. El conjunto de respondentes simultáneos para estimaciones longitudinales está entonces compuesto por las 828 mujeres recontactadas en 2008.

El método de calibración aplicado fue el *raking* o *calibración en las marginales*. Para el cálculo de los ponderadores se utilizó el programa R Development Core Team (2008) con el paquete *survey* (T. Lumley (2009)) y la función *rake*. Los insumos necesarios para calcular estos nuevos ponderadores son el diseño que genera los expansores originales (estratificado por nivel socioeconómico) y los totales marginales sobre los cuales se calibrará, que en esta aplicación corresponden a la Edad y Nivel Educativo estimados a partir de la Encuesta Nacional de Hogares Ampliada del año 2006 (ENHA 2006).

Referencias

- [1] Bailar, B.A. (1975). “The Effects of Rotation Group Bias on Estimates from Panel Surveys”. *Journal of the American Statistical Association*, 70(349): 23–30.
- [2] Chhikara, R.S.; and Deng, L.Y. (1992). “Estimation Using Multiyear Rotation Design Sampling in Agricultural Surveys”. *Journal of the American Statistical Association*, 87(420): pp 924–932.
- [3] Copeland, K.R. (2004). “Nonresponse Adjustment in the Current Employment Statistics Survey”. U.S. Bureau of Labour Statistics, Washington DC.
- [4] De Leeuw, E.D.; Hox, J.; and Huisman, M. (2003). “Prevention and Treatment of Item Nonresponse”. *Journal of Official Statistics*, 19 (2): 153–176.
- [5] De Leeuw, E.D. (2006). “Introduction to Survey Nonresponse”. Summer Institute in Survey Research Techniques. *Survey Methodology* 988.223.
- [6] Dennis, J.M. and Li, R. (2003). “Effects of Panel Attrition on Survey Estimates”. For Presentation at the 2003 Annual Meeting of the American Association for Public Opinion Research in Nashville, Tennessee.
- [7] Deville, J.C. and Särndal, C.E. (1992). “Calibration Estimators in Survey Sampling”. *Journal of the American Statistical Association*, 87(418).
- [8] Deville, J.C.; Särndal, C.E.; and Sautory, O. (1993). “Generalized Raking Procedures in Survey Sampling”. *Journal of the American Statistical Association*, 88(423).
- [9] Deville, J.C. and Särndal, C.E. (1994). “Variance estimation for the regression imputed Horvitz – Thompson estimator”. *Journal of Official Statistics*. 10: 381–394.
- [10] Ernst, L.R (1986). “Weighting Issues for Longitudinal Household and Family Estimates”. Bureau of Census, SRD Research Report Number: CENSUS/SRD/RR-86/23.
- [11] Fuller, W.A. and Breidt, F.J. (1999). “Estimation for Supplemented Panels”. *The Indian Journal of Statistics Special Issue on Sample Surveys*, 61: 58–70.

- [12] Gross Sobol, M. (1959). "Panel Mortality and Panel Bias". *Journal of the American Statistical Association*, 54(285): 52 - 68.
- [13] Groves, R.M. and Couper M.P (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- [14] Kalton, G. and Brick, J.M . "Weighting Schemes for Household Panel Surveys: The Survey of Income and Program Participation". U.S. Department of Commerce Bureau of Census.
- [15] Kish, Leslie (1995). *Diseño Estadístico para la Investigación*. España: Siglo XXI de España Editores SA.
- [16] Little, R.J.A and David, H.M. (1983). "Weighting Adjustment for Nonresponse in Panel Surveys". Working Paper: U.S. Bureau of Census, Washington DC.
- [17] Little, R.J.A and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley & Sons.
- [18] Lumley, T. (2004) Analysis of complex survey samples. *Journal of Statistical Software* 9(1): 1-19
- [19] Lumley, T. (2009) "survey: analysis of complex survey samples". R package version 3.11-2.
- [20] Lynn, P.; Buck, N.; Burton, J.,; Jäckle, A.; and Laurie, H. (2005). "A Review of Methodological Research Pertinent to Longitudinal Survey Design and Data Collection". Working Papers of the Institute for Social and Economic Research, paper 2005 – 29. Colchester: University of Essex.
- [21] Rubin, Donald B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley & Sons.
- [22] R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [23] Särndal, C.E.; Swensson, B.; and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

- [24] Särndal, C.E. and Lundström, S. (1999). “Calibration as a Standard Method for Treatment of Nonresponse”. *Journal of Official Statistics*. 15(2): 305–327.
- [25] Särndal, C.E. and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. New York: John Wiley and Sons, Ltd.
- [26] Särndal, C.E. and Lundström, S. (2008). “Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator”. *Journal of Official Statistics*, 24 (2):167–191.
- [27] Sastry, N; Ghosh-Dastidar, B; Adams, J; and Pebley, A (2000). “The Design of a Multilevel Longitudinal Survey of Children, Families, and Communities: The Los Angeles Family and Neighborhood Survey”. Labour and Population Program, Working Paper Series 00-18.
- [28] Sharot, T. (1991). “Attrition and Rotation in Panel Surveys”. *The Statistician*, 40(3): 325–331.
- [29] Simard, M. (2002). “Generation and Gender Survey Sample Design Guidelines”. Statistics Canada.
- [30] Taylor, B.; Brook, L.; and Lynn, P. (1997). “Incentives Information and Number of Contacts: Testing the Effects of these Factors of Response to a Panel Survey”. *Survey Methods Centre Newsletter*, 17(3): 712.
- [31] Wadsworth, R.N. (1952). “The experience of a User of a Consumer Panel”. *Applied Statistics*, 1(3): 169–178.
- [32] Williams, W.H. and Mallows, C.L. (1970). “Systematic Biases in Panel Surveys Due to Differential Nonresponse”. *Journal of the American Statistical Association*, 65(331).
- [33] Woodruff, R.S. (1963). “The Use of Rotating Samples on the Census Bureau’s Monthly Surveys”. *Journal of the American Statistical Association*, 58(302): 454–467.