

UNIVERSIDAD NACIONAL DE INGENIERIA

FACULTAD DE INGENIERIA ECONOMICA Y

CIENCIAS SOCIALES



**“ERRORES NO MUESTRALES: EL ERROR DE NO RESPUESTA Y SU
INFLUENCIA EN LA VALIDEZ Y FIABILIDAD EN LAS ENCUESTAS
DE OPINIÓN EN HOGARES”**

INFORME DE SUFICIENCIA

PARA OBTAR EL TITULO PROFESIONAL DE:

LICENCIADO ESTADÍSTICO

**POR LA MODALIDAD DE ACTUALIZACION DE
CONOCIMIENTOS**

ELABORADO POR:

ALEJANDRINA SUSY CARAZA RIVERA

**LIMA – PERU
2004**

A mis padres Alejandro y Antonieta

A mi querida hermana Olinda

A mis hermanos Jorge y Alex

A mis sobrinos Abdul, Hellen, Alexei, Jhon

A mi gordo bello Alexander

**“ERRORES NO MUESTRALES: EL ERROR DE NO RESPUESTA Y
SU INFLUENCIA EN LA VALIDEZ Y FIABILIDAD EN LAS
ENCUESTAS DE OPINION EN HOGARES”**

INDICE

PRESENTACION	PAG
CAPITULO 1	1
1. OBJETIVOS	1
1.1. Objetivo General	1
1.2. Objetivos Especificos	1
CAPITULO 2	3
2. MARCO TEORICO	3
2.1. Introducción	3
2.2. Importancia de la asignación optima de recursos en el Diseño de Encuestas de hogares	4
2.3. La calidad de los datos en las encuestas de hogares	6
2.4. Clasificación de los errores no muestrales.	8
2.5. Definición de Falta de respuesta como un error no muestral	9
CAPITULO 3	11
3. METODOLOGÍA	11

3.1. Modelo planteado de error de no respuesta	11
3.2. Árbol de errores diversos	13
3.3. Causas de los errores de no respuesta	14
3.3.1. La falta de contacto	14
3.3.2. La negativa	18
3.4. Efectos de los errores de No respuesta	20
3.5. Métodos para el control de errores de No respuesta	23
3.6. Resultados Heurísticos de Tasas de No respuesta	42
3.6.1. Identificación del encuestador y la Tasa de No respuesta	42
3.6.2. Carga de respuesta y Tasa de No respuesta	43
3.6.3. Encuestadores y las Tasas de No respuesta	45
3.6.4. Características de los respondientes y las Tasas de No Respuesta	47
CAPITULO 4	52
4. CONCLUSIONES Y RECOMENDACIONES	52
4.1. Conclusiones	52
4.2. Recomendaciones	53
5. BIBLIOGRAFIA	55

PRESENTACIÓN

La presente monografía presenta la necesidad urgente de tomar conciencia de la magnitud e importancia de los errores no muestrales en especial del ERROR DE NO RESPUESTA que amenaza invalidar o sesgar en una dirección los resultados de cualquier investigación; es por ello que lleva como título: “ERRORES NO MUESTRALES: EL ERROR DE NO RESPUESTA Y SU INFLUENCIA EN LA VALIDEZ Y FIABILIDAD EN LAS ENCUESTAS DE OPINIÓN EN HOGARES”.

La monografía se presenta en cuatro capítulos. En el primer capítulo se presenta los objetivos de la monografía. En el segundo capítulo el marco teórico necesario para poder desarrollar el tema propuesto. En el tercer capítulo se desarrolla toda la metodología utilizada, desde la presentación del modelo planteado de error no muestral, seguido por la presentación del árbol de los errores no muestrales con la respectiva ubicación de los errores de no respuesta y por último el desarrollo de las causas, efectos, métodos para mejorar las tasas de no respuestas y resultados heurísticos del autor de la monografía. En el último capítulo se presenta las conclusiones y recomendaciones.

CAPITULO 1

1. OBJETIVOS

Los objetivos de esta monografía es sin duda hacer tomar conciencia a los investigadores sobre los errores de no respuesta. Es en este contexto que nos planteamos el siguiente objetivo general y específicos.

1.1. Objetivo General

Presentar los errores de no respuesta en los resultados de encuestas en hogares.

1.2. Objetivos Específicos

- Presentar el modelo planteado de error no muestral.
- Presentar el árbol de los errores diversos con la respectiva ubicación de los errores de no respuesta.
- Desarrollar las causas de los errores de no respuesta.
- Desarrollar los efectos de los errores de no respuesta.

- **Desarrollar los métodos para el control de los errores de no respuestas.**
- **Presentar resultados heurísticos de las tasa de no respuesta en encuestas de hogares.**

CAPITULO 2

2. MARCO TEORICO

A continuación presentamos toda la teoría necesaria para el desarrollo del tema de la presente monografía.

2.1. Introducción

Los investigadores deben estar concientes que en todo estudio de investigación se cometen errores. Estos errores generan información errónea y en consecuencia deben estar alertas de las posibles fuentes de error de manera que puedan tomar precauciones para disminuir la magnitud de dichos errores de acuerdo con la exactitud requerida.

Generalmente, cuando se habla de errores se refieren a los errores de muestreo. Cuando se utiliza las técnicas de muestreo siempre se utiliza una muestra para averiguar alguna característica de una población, en consecuencia se presenta casi

siempre diferencias entre el valor de la muestra (estimador) y el valor real de la población (parámetro). A esta diferencia se le conoce como error, pero el error esta compuesto por errores muestrales y errores no muestrales. Estos últimos se presentan en el proceso de investigación; se incluyen todos los aspectos del proceso del levantamiento de la información (trabajo de campo), donde se presentan todo tipo de errores involuntarios. Si bien es cierto los errores no muestrales se pueden medir fácilmente, las empresas encuestadoras no están acostumbrados en nuestro medio a calcular dichos errores debido a que no se implementan los controles de calidad debidas y en consecuencia los errores no muestrales sesgan los resultados en una dirección. El error no muestral puede invalidar los resultados de un estudio.

2.2. Importancia de la asignación optima de recursos en el Diseño de Encuestas de hogares ¹

El diseño de encuestas de hogares en su etapa de acopio de la información requiere la adopción de numerosas decisiones en cuanto a los procedimientos, instrumentos y personal que se utilizara para recoger y elaborar los datos. Muchas de estas decisiones son críticas para el control de los errores no muestrales pero están relacionados directamente con los recursos económicos asignados para el levantamiento de la información.

El especialista en muestreo¹ se tiene que enfrentar a numerosas decisiones no solo de tipo de la selección de la muestra si no como los datos son recogidos y no cuenta

¹ / Programa de las Naciones Unidas para el mejoramiento de la calidad de datos. ONU 1989

con información suficiente. Por tal motivo debemos ser conscientes que cuando se hace por primera vez un tipo de encuesta el diseño y los procedimientos utilizados se le puede considerar como una primera aproximación de la investigación.

Existen muchas limitaciones de orden práctico cuando se efectúa el trabajo de campo. Por ejemplo, la disponibilidad, calidad y actualización de los materiales de muestreo, así como la calidad del personal tienen efecto en la elección del diseño de la encuesta, la cual está relacionada con la asignación de los recursos económicos para esta etapa.

Los diseños muestrales óptimos no son muchas veces factibles su aplicación debido a que implican gastos muy diversos que incrementan los presupuestos fijados o disponibles de los clientes.

Otro aspecto a considerar en la asignación de recursos es que se puede elaborar un plan de mejoramiento del levantamiento de la información. Por ejemplo en encuestas de opinión pública sobre temas coyunturales las cuales se realizan mensualmente se puede mejorar la capacitación y controlar los errores cometidos debido al encuestador, de igual forma lo referente al material utilizado para el diseño de la muestra, el cual puede ser actualizado mes a mes en campo.

2.3. La calidad de los datos en las encuestas de hogares ²

Antes de iniciar un análisis detallado de los errores no muestrales en las encuestas de hogares será útil considerar un concepto mas amplio, a saber, la calidad de los datos. Esta depende de tres características: su adecuación a las necesidades de los usuarios, su oportunidad y su exactitud. Ninguna de estas tres características se puede ignorar al planificar y ejecutar las encuestas: los datos solo son útiles si resultan suficientemente adecuados, oportunos y exactos.

- **Adecuación:** En la práctica los clientes expresan sus necesidades de información y los especialistas de las empresas estadísticas elaboran en forma conjunta con las primeras un perfil de las necesidades del cliente. Las variables de estudio así como las preguntas utilizadas en el cuestionario deben estar definidas en forma conjunta. La empresa encuestadora debe precaverse de no prometer mas de lo que razonablemente se puede esperar de una encuesta de hogares, teniendo en cuenta las limitaciones de recurso y tiempo. Desde el punto de vista del muestreo el tamaño de la muestra raramente es el adecuado para producir inferencias en muchos niveles de desagregación. Por ejemplo para una muestra de 1200 entrevistados en Lima Metropolitana se puede obtener inferencias para los niveles de desagregación por nivel socioeconómico, sexo, cuotas de edad. Pero no podemos pretender realizar inferencias para niveles de desagregación por distrito.

² / Programa de las Naciones Unidas para el mejoramiento de la calidad de datos. ONU 1989

- **Oportunidad:** Es posible imaginar situaciones en las cuales los resultados de una encuesta serian de poco interés para la mayoría de los clientes si no se dispusiera de ellos en una fecha determinada. Las encuestas de tipo político, sobre elecciones específicas, caen en esta categoría.

Después de la etapa de recolección de datos la utilidad de los resultados de las encuestas disminuye gradualmente con el transcurso del tiempo. La tasa de declinación de la utilidad por el transcurso del tiempo depende del contenido y los objetivos de la encuesta.

A menudo los clientes insisten en la oportunidad, a expensas de la exactitud. La demora en la disponibilidad de los resultados es muy evidente, mientras que la falta de exactitud, especialmente cuando surgen errores no muestrales, es para los clientes mucho mas difícil de reconocer o no tienen conocimiento de las mismas.

La empresa encuestadora debe producir datos oportunos para facilitar su uso práctico, pero también tienen la responsabilidad de que sean exactos.

- **Exactitud:** El objetivo de una encuesta por muestreo es hacer estimaciones de ciertos valores para una población, utilizando observaciones obtenidas de un número limitado de unidades de una población.

En general se considera que la exactitud de una estimación significa su aproximación a un valor verdadero (parámetros de la población), este es siempre desconocido, el cual se obtendría si se pudiera recoger y procesar, sin errores, los datos referentes a todas las unidades de la población. El error de una estimación determinada es la diferencia entre la estimación y el valor verdadero.

2.4. Clasificación de los errores no muestrales

Los errores no muestrales se clasifican considerando las tres etapas para efectuar una encuesta:

- ✓ Errores no muestrales presentados en el diseño y preparación de la encuesta
- ✓ Errores no muestrales presentados en la recolección de los datos (trabajo de campo). Los cuales a su vez se clasifican según la fuente o tipo de error en:
 - ✓ Errores no muestrales de cobertura
 - ✓ Errores no muestrales de falta de respuesta
 - ✓ Errores no muestrales de respuesta
- ✓ Errores no muestrales presentados en el procesamiento y análisis de datos.

2.5. Definición de Falta de respuesta como un error no muestral

La falta de respuesta surge cuando los hogares u otras unidades de observación seleccionadas para que formen parte de la muestra no responden el cuestionario. Este hecho origina que las estimaciones que se hagan de la población no corresponda a una muestra probabilística sino que se asemeje a un muestreo por intención ³. Este hecho ocasiona un error no muestral debido a la falta de respuesta del encuestado que puede invalidar la investigación.

Existen diferentes situaciones:

1. La unidad de análisis elegida para la muestra en el hogar no se encuentra en casa.
2. La unidad elegida para la muestra en el hogar no desea contestar el cuestionario
3. No existe en el hogar elegido para la muestra una unidad de análisis con las características deseadas para el estudio.

³ / Muestreo por intención: Se considera muestreo por intención cuando las unidades que forman parte de la muestra son elegidas a conveniencia por el encuestador. Si un encuestado no desea colaborar es reemplazado por otro que si desea colaborar.

En esta monografía las dos primeras situaciones son consideradas como errores de no respuesta , mientras que la tercera no lo es.

CAPITULO 3

3. METODOLOGIA

A continuación presentamos la metodología utilizada en el desarrollo del tema de la presente monografía.

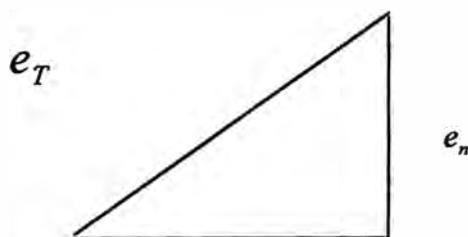
3.1. MODELO PLANTEADO DE ERROR DE NO RESPUESTA

Sea θ el valor del parámetro de la población (valor verdadero para la población estudiada) y $E(\hat{\theta})$ el valor esperado del parámetro de la población, entonces definimos al error total de la siguiente manera:

$$e_T = E(\hat{\theta}) - \theta$$

El modelo planteado que explica los errores no muestrales es el siguiente:

$$e_T = \sqrt{e_m^2 + e_n^2}$$



notación

e_T = Error total

e_m = Error muestral

e_n = Error no muestral

Donde

$$e_n = \sqrt{e_a^2 + e_b^2 + e_c^2}$$

Notación

e_a = Errores no muestrales presentados en el diseño y preparación de la encuesta

e_b = Errores no muestrales presentados en la recolección de los datos (trabajo de campo)

e_c = Errores no muestrales presentados en el procesamiento y análisis de datos.

Además

$$e_b = \sqrt{e_d^2 + e_e^2 + e_f^2}$$

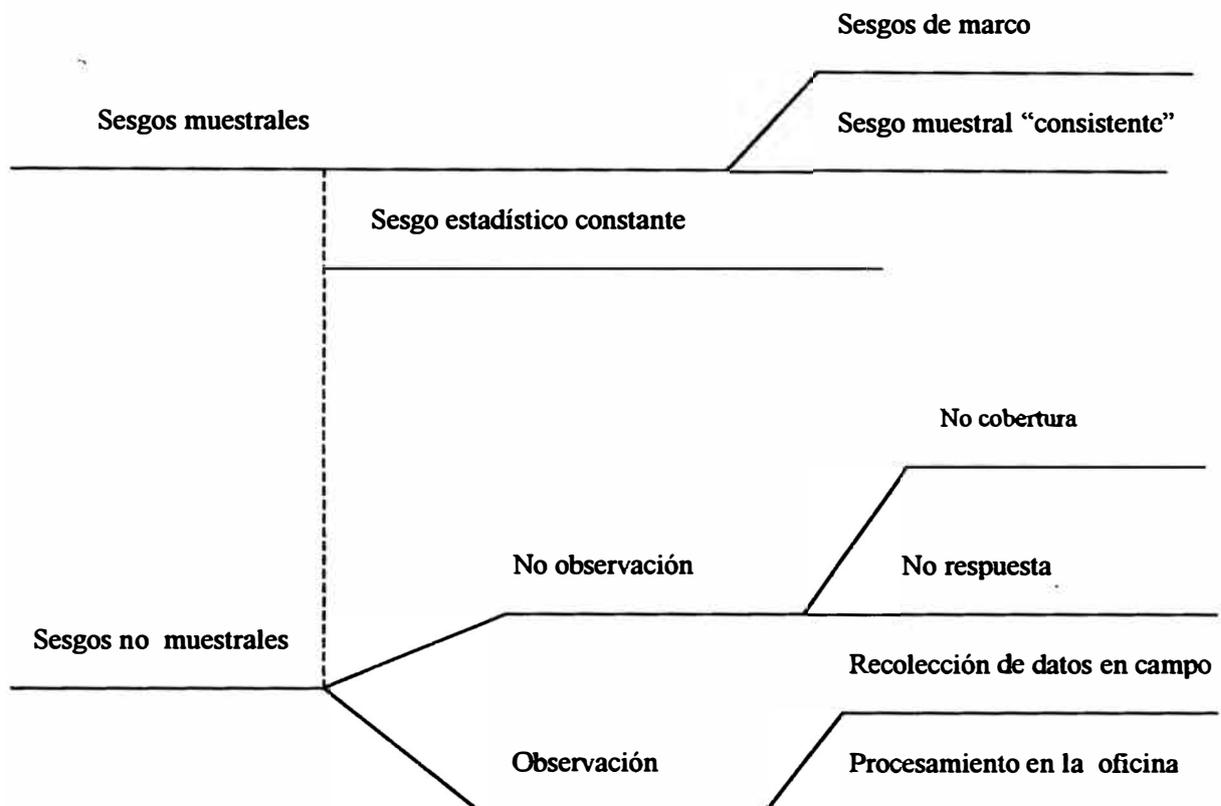
Notación

e_d = Errores no muestrales presentados por la no cobertura.

e_e = Errores no muestrales presentados por la no respuesta.

e_f = errores no muestrales presentados por la respuesta de los encuestados.

3.2. ARBOL DE ERRORES DIVERSOS ⁴



3.3. CAUSAS DE LOS ERRORES DE NO RESPUESTA ⁵

Hay muchos factores que pueden conducir a la falta de respuesta. La importancia relativa de los distintos factores y de las medidas necesarias para controlarlos varían de un país a otro, de una cultura a otra y de una encuesta a otra. En una situación cualquiera, el objeto de la identificación de las causas principales de falta de

⁴ Kish Leslie, 1982. Muestreo de Encuestas. Ed. Trillas S. A. México.

⁵ Hansen y Hurwitz, 1946. the problem of non response in simple surveys. Jour. Amer. Stat. Assoc.

respuesta y de las características del personal sobre el terreno y de los encuestados vinculados con la falta de respuesta, es idear medidas para controlarlas, hacer los ajustes respectivos y estimar sus efectos en los resultados de la encuesta.

Se pueden identificar dos categorías principales de falta de respuesta:

3.3.1. La falta de contacto.

De una manera general, el problema más importante en la mayoría de los países en desarrollo, especialmente en las áreas rurales, parece ser la falta de contacto es decir los inconvenientes para llegar a las unidades e identificarlas, defectos del marco muestral y/o otras circunstancias adversas.

- Inconvenientes para llegar a las unidades e identificarlas: Una posible causa de falta de contacto, es la inaccesibilidad de algunas áreas de la muestra debido a situaciones tales como disturbios civiles, problemas de tipo político o de seguridad, o inundaciones u otras calamidades naturales.

A veces, a causa de una publicidad desfavorable, la mayoría de los encuestados de un área particular de la muestra puede negarse a cooperar.

Esto también ocurre cuando algunas unidades caen en lugares peligrosos y el encuestador prefiere darlas por no respuestas

En esas circunstancias, la mayoría de los casos de falta de respuesta se puede producir en unas pocas áreas claramente identificadas. Debe hacerse un intento de prever esos inconvenientes en la fase de planificación y hacer arreglos especiales para asegurar la cobertura de las áreas con problemas, en la mayor medida posible. Por ejemplo, los inconvenientes se pueden reducir volviendo a programar la encuesta o empleando encuestadores que residan en el ámbito a ser encuestado. Si persiste algún problema de cierta magnitud, quizás sea necesario volver a definir la población en estudio, excluyendo explícitamente los lugares inaccesibles.

Igualmente, es un factor importante la imposibilidad de encontrar a un encuestado elegible (persona que necesitamos para forma parte de nuestra muestra) en casa después de reiterados intentos.

Una vez que el encuestador ha ubicado e identificado la unidad de la muestra (hogar) que debe visitar, puede sin embargo no llegar a establecer contacto con un encuestado adecuado. Hay varios factores de los cuales depende que se pueda contactar a un encuestado apropiado.

Son importantes las características de los encuestados que se consideran aceptable para la encuesta. Es probable que la falta de contacto sea menos común en encuestas de hogares en las que cualquier miembro

adulto puede dar la información requerida, y más común si es necesario establecer contacto, en forma individual, con un miembro especificado de cada hogar, o con más de uno. El tipo de trabajo del encuestado también es importante: por ejemplo, es más fácil encontrar a las personas que trabajan en su casa que a aquéllas que lo hacen fuera de casa.

El grado de dificultad para establecer contacto con un encuestado aceptable para la encuesta depende mucho de la instrucción de campo de los encuestadores, en algunos casos se les permite las respuesta por sustitutos es decir los datos correspondientes a la persona seleccionada como unidad muestral puede ser obtenida por medio de una tercera persona, como por ejemplo otro miembro del hogar.

En encuestas repetidas con una determinada frecuencia usualmente es posible y justificable en función de su bajo costo dedicar más atención y recursos a mejorar el marco muestral y, por consiguiente, reducir la incidencia de la falta de contacto. En este mismo sentido, también es útil el despliegue de encuestadores permanentes, ubicados frecuentemente dentro de las áreas de la muestra o cerca de ellas. Los encuestadores pueden familiarizarse con las condiciones del lugar y con los encuestados y pueden realizar visitas de insistencia (volver al hogar seleccionado hasta que se encuentre la unidad de análisis) en forma más económica.

En muchos casos se hace necesario emplear encuestadores contratados en el lugar, que conocen las condiciones e idioma.

- Los defectos o deficiencias del marco muestral: son otra causa común de falta de contacto. Por ejemplo, una descripción inadecuada de las unidades de área de la muestra puede provocar la no ubicación o identificación correcta de las unidades asignadas. Esos problemas depende de factores como los procedimientos de muestreo utilizados, la calidad de las listas de la muestra y de su grado de actualización, la movilidad de la población y, por supuesto, las calificaciones, capacitación y supervisión de los encuestadores en el trabajo de campo.
- Otro factor que influye en el logro de contacto con los respondientes es la oportunidad en que se realiza la encuesta. Por ejemplo, durante la época de vacaciones un gran número de personas puede encontrarse temporariamente ausente de su lugar habitual de residencia. Los trabajadores agrícolas pueden estar ausentes en la época de la cosecha. En realidad, la oportunidad del trabajo de campo puede constituir una importante decisión en materia de diseño, que provoque no sólo el éxito o el fracaso del intento de contactar a los encuestados, sino que también puede afectar la calidad y factibilidad del conjunto de las operaciones de la encuesta.

A este respecto, es útil distinguir dos categorías de falta de respuesta:

- ✓ Temporariamente ausente
- ✓ No está en casa.

El primer caso se refiere a los encuestados que se encuentran ausentes durante todo el período de las operaciones de la encuesta.

El segundo caso se refiere a los encuestados que no se encuentran en sus casas cuando llega el encuestador a realizar la aplicación de la encuesta.

Este caso depende de factores tales como el tipo de encuestado (por ejemplo, las horas del día, durante la semana, son particularmente inconvenientes para encontrar a los miembros del hogar que trabajan).

Se debe dar instrucciones al encuestador sobre la hora de visita al encuestado, la disponibilidad de informaciones acerca de los movimientos del encuestado y especialmente el número de visitas que debe realizar hasta ubicar al encuestado.

3.3.2. La negativa. Debemos destacar la incidencia de la negativa de los encuestados potenciales a cooperar.

Después de establecerse el contacto con el respondiente, el paso siguiente es obtener su cooperación para que proporcione la información requerida. Este esfuerzo puede tener éxito completo o parcial, o no tener ninguno; las dos últimas posibilidades conducen a una falta de respuesta parcial o total, respectivamente.

En esta monografía cuando hablemos de negativa de respuesta o cooperación del encuestado nos estaremos refiriendo a la falta de respuesta total.

Como se indicó antes, las negativas constituyen un sesgo en los resultados de las encuestas. En las operaciones en gran escala que cuentan con un buen presupuesto se puede utilizar una campaña publicitaria por los medios de difusión, para hacer conocer al público la operación, su oportunidad y su importancia. Esto puede dar como resultado tasas de respuesta más elevadas. Sin embargo, la mayoría de las encuestas no tiene un presupuesto que permitan realizar una amplia campaña publicitaria. No obstante, estas campañas pueden ser provechosas si se concentran en las áreas seleccionadas para la muestra. En muchas de ellas es esencial realizar análisis preliminares con autoridades locales, como un primer paso para obtener la cooperación de los residentes seleccionados para la encuesta.

Hay muchos factores que pueden influir en la negativa en responder del encuestado, incluyendo a los auspiciadores de la encuesta, la reputación de la entidad que la lleva a cabo, la índole de las preguntas a formular, la longitud o tiempo de la encuesta, la técnica y experiencia de los encuestadores. Por estas razones, es importante que toda organización encuestadora realice su trabajo de campo de tal manera que siempre se trate a los encuestados con cortesía y respeto y que no se les produzca ningún perjuicio por el hecho de haber proporcionado información con fines estadísticos.

3.4. EFECTOS DE LOS ERRORES DE NO RESPUESTA ⁶

Para estudiar los diversos efectos de los errores de no respuesta en las encuestas, denotaremos por:

$$\bar{Y} = \frac{Y}{N} = W_1 \bar{Y}_1 + W_2 \bar{Y}_2$$

la media de alguna característica, donde W_1 y W_2 son las proporciones de respuesta y no respuesta ($W_1 + W_2 = 1$) y \bar{Y}_1 y \bar{Y}_2 denotan las medias de las características en los dos segmentos. Estos son los valores promedio que se espera tener bajo las condiciones esenciales de la encuesta.

⁶ Kish Leslie, 1982. Muestreo de Encuestas. Ed. Trillas S. A. México.

El empleo de respuesta media \bar{Y}_1 para estimar la media \bar{Y} causa un sesgo ($\bar{Y}_1 - \bar{Y}$).

El sesgo relativo (SR) de la media muestral es

$$SR(Y_1) = \frac{(\bar{Y}_1 - \bar{Y})}{\bar{Y}} = \frac{(\bar{Y}_1 - W_1 \bar{Y}_1 - W_2 \bar{Y}_2)}{\bar{Y}} = W_2 \frac{(\bar{Y}_1 - \bar{Y}_2)}{\bar{Y}}$$

Si la media de no respuesta \bar{Y}_2 difiere en poco de la media de respuesta \bar{Y}_1 , el sesgo relativo conservara un valor pequeño, aunque los valores de W_2 sean moderados. Si tanto W_2 como $(\bar{Y}_2 - \bar{Y}_1)$, son pequeños, el sesgo deberá ser despreciable. Para que el sesgo sea importante, debe haber una coincidencia de un valor grande de la no respuesta con diferencias grandes entre las medias de los dos segmentos.

El total de la población $Y = N\bar{Y}$ se estima con la expansión simple Fy , cuyo el valor esperado es $N_1 \bar{Y}_1 = NW_1 \bar{Y}_1$. El sesgo relativo es

$$SR(N_1 Y_1) = \frac{(NW_1 \bar{Y}_1 - N\bar{Y})}{N\bar{Y}} = -\frac{NW_2 \bar{Y}_2}{N\bar{Y}} = -W_2 \frac{\bar{Y}_2}{\bar{Y}}$$

En consecuencia, si una subclase difícil de incluir en el marco contiene cantidades despreciables de la característica (porque $Y_2 = 0$), su exclusión mediante un método

de eliminación no sesgara sustancialmente el estimador de expansión simple $N_1 \bar{Y}_1$.

Sin embargo, cuando $\bar{Y}_2 = \bar{Y}_1$, el sesgo relativo de la expansión simple adquiere aproximadamente el tamaño de la no cobertura W_2 ; si se conoce el tamaño total de la población N , se puede emplear para calcular W_1 y, en consecuencia, para estimar Y con $(F/W_1) y_1$, cuyo valor esperado es NY . Por tanto, se puede incrementar el factor de expansión para la no respuesta, si se supone que las medias de las respuestas y de las no respuestas son iguales. En esta expansión ajustada, el sesgo relativo de $N\bar{Y}_1$ es igual que el de Y_1 ; es decir, $SR(N\bar{Y}_1) = SR(\bar{Y}_1)$

En la comparación de dos medias, la diferencia $(\bar{Y}_a - \bar{Y}_b)$ tiene el sesgo

$$\left[\left(\bar{Y}_1 - \bar{Y} \right)_a - \left(\bar{Y}_1 - \bar{Y} \right)_b \right] = \left[W_2 \left(\bar{Y}_1 - \bar{Y}_2 \right) \right]_a - \left[W_2 \left(\bar{Y}_1 - \bar{Y}_2 \right) \right]_b$$

En consecuencia, el sesgo relativo es

$$\frac{\left[W_2 \left(\bar{Y}_1 - \bar{Y}_2 \right) \right]_a - \left[W_2 \left(\bar{Y}_1 - \bar{Y}_2 \right) \right]_b}{\bar{Y}_a - \bar{Y}_b} = W_2' \frac{\left(\bar{Y}_1 - \bar{Y}_2 \right)_a - \left(\bar{Y}_1 - \bar{Y}_2 \right)_b}{\bar{Y}_a - \bar{Y}_b}$$

En la última expresión se supone que $W_{2a} = W_{2b} = W_2'$; es decir, que la proporción de no respuestas es aproximadamente igual en las dos clases.

Cuando esto sucede, puede ocurrir un sesgo relativo importante, si el efecto del sesgo de no respuesta en la media de una clase es muy diferente del de la otra.

Incluso, en los casos en que los estimadores separados tiene sesgos de no respuesta, sus diferencias suelen estar relativamente libres de sesgos, debido a la semejanza de los efectos de la no respuesta en las medias de clase que se comparan. La comparación de dos encuestas periódicas, que se ejecutan en las mismas condiciones esenciales, tendrá sesgo pequeño si cada encuesta tiene un sesgo semejante.

Sin embargo, no conviene suponer que esta cancelación de sesgos ocurre siempre. En primer lugar, la proporción de no respuestas puede ser diferente en las dos clases ($W_{2a} \neq W_{2b}$). En segundo lugar, tal vez difiera el efecto de la no respuesta de una clase a otra.

3.5. METODOS PARA EL CONTROL DE ERRORES DE NO RESPUESTA ⁷

Es poca la gente que nunca, o bien siempre, esta en su casa ; en un periodo de encuesta, la probabilidad (1-Pi) de convertirse en elemento ausente no es de 1 ó de 0 en la mayoría de los individuos. Para cada uno se supone la existencia de una probabilidad Pi de ser encontrado en su casa (respuesta) que se define para un triodo de entrevista y operación.

El valor de esa probabilidad esta entre 0 y 1 y es diferente de un individuo a otro. En las primeras visitas obtenemos una sobrerrepresentación de personas con

⁷ Kish Leslie, 1982. Muestreo de Encuestas. Ed. Trillas S. A. México.

probabilidades altas. Los resultados de las segundas, terceras y posteriores visitas contienen proporciones cada vez mayores de personas con probabilidades bajas.

Aunque en cada visita se recoge una mezcla de personas con probabilidades diferentes de respuestas, las mezclas varían entre sí; en las primeras visitas hay una proporción demasiado pequeña de probabilidades bajas. En visitas posteriores se recogen más de estas últimas, y el incremento del número de visitas hace disminuir la diferencia entre la muestra y las mezclas en la población.

Cuando se hace solamente una cantidad razonable de visitas, queda una discrepancia entre la muestra y las mezclas en la población. Además la gente cuya probabilidad de ser encontrada es cero, debido a sus horas de trabajo o a su ausencia durante el periodo de la encuesta, quedará sin representación en la muestra. Este tipo es raro en los casos que no están en casa, aunque frecuente en los rechazos de la entrevista; la porción de los que rehúsan definitivamente resistirá consistentemente todo esfuerzo.

Otros rechazos se parecen a los que no están en casa, pues son parciales, relativos y remediables, en lugar de ser totales, absolutos y definitivos; por consiguiente, es más adecuado un modelo que se base en una escala de varios grados de rechazos que otro de “todo o nada” (0 o 1). Este modelo da también mejores resultados, porque tiende a reducirle margen de ignorancia, y señala cuáles son los mejores remedios y controles. Sin embargo ese modelo es complicado, y la aproximación en

la que $P_i=0$ para la mayor parte de los rechazos resulta razonable. Por lo tanto, los tres métodos de remedio que dependen de que $P_i \neq 0$ son mucho menos útiles en los rechazos que en los que no están en casa : se trata de las visitas adicionales, lo reemplazos y el plan de Politz. Tenemos otros cuatro métodos que pueden remediar tanto los casos los caso que no están en casa como los rechazos: mejores procedimientos, submuestreo, estimación del efecto y sustitución.

Los siguientes métodos sirven en situaciones diferentes para reducir el porcentaje de no respuesta, o bien sus efectos.

1. El mejoramiento de los procedimientos de recolección de datos es el remedio mas obvio para incrementar la respuesta. Los mejoramientos indicados para la reducción de rechazos son:
 - a) La garantía del anonimato del entrevistado
 - b) Motivación para la cooperación del encuestado: despertar el interés del encuestado con comentarios acertados al iniciar las preguntas
 - c) Dar noticia por adelantado al encuestado, aunque a veces resulta contraproducente, es un procedimiento que puede incrementar la proporción de los que se encuentran en sus casas.

2. Las visitas repetidas son de gran efectividad para reducir los casos que no están en casa.

3. El submuestreo de las visitas que deben repetirse resulta económico, cuando el contacto con esas viviendas es mucho mas costoso que el primer intento.
4. La estimación del efecto de la no respuesta puede producir evidencia de la ausencia de sesgos grandes debido a la no respuesta. El reporte del tamaño y las causas de la no respuesta se ha convertido en nuestros días en practica común de las mejores encuestas. A veces se diseñan métodos formales que estiman el tamaño del sesgo y lo reducen.
5. La sustitución de las no respuestas suele ser sugerida como remedio. Con frecuencia se trata de un error puesto que los sustitutos se asemejan a las respuestas ya obtenidas, y no a los individuos de la no respuesta.
6. El procedimiento de reemplazo, es un método beneficioso de sustitución.
7. El esquema de Politz es un procedimiento con el que se obtienen diferencias en las probabilidades de las respuestas y se las pondera correspondientemente.

Los intentos para reducir el porcentaje o los efectos de la no respuesta se dirigen a reducir el sesgo causado por as diferencias entre los que responden y los que no lo hacen. No se debe confundir el sesgo de la no respuesta con la reducción del tamaño de la muestra causado por la no respuesta. El efecto último se resuelve con

facilidad, bien al anticipar el tamaño de la no respuesta en el diseño del tamaño muestral, o bien al compensarlo mediante un suplemento. Estos ajustes no hacen sino incrementar el tamaño de la respuesta, así como la precisión del muestreo, pero no reducen el porcentaje de no respuesta ni tampoco el sesgo que puede causar.

A continuación se detalla algunos de los métodos antes mencionados:

a. Trabajo complementario intensivo con una sub muestra de no respondientes.

Se ha propuesto, especialmente con respecto a las encuestas por correo, que después de un esfuerzo razonable por obtener respuesta mediante los procedimientos estándar se seleccione una sub muestra de los no respondientes restantes y se utilicen procedimientos más intensivos para conseguir la información correspondiente a esas unidades. Generalmente este método se adjudica a Hansen y Hurwitz, que lo trataron detalladamente en un artículo de 1946; sin embargo, Dalenius (1957) señala que la idea fue sugerida primero por Cornfield (1942).

Si en la práctica se obtienen entrevistas con todos los no respondientes de la sub muestra, y si las probabilidades de selección se reflejan adecuadamente en las estimaciones de la muestra, este enfoque eliminaría totalmente el sesgo debido a falta de respuesta. Sin embargo, esto se hace al costo de un incremento El uso del método en una encuesta con entrevistas personales complicaría sustancialmente el diseño de la muestra y los procedimientos de estimación y probablemente demoraría la terminación de la recolección de datos, siendo mas

complicado cuando el tema es de coyuntura política y pudiera pasar de un día al otro acontecimientos que cambien la opinión de los entrevistados.

a.1. Método de Hansen y Hurwitz (1946)

De las n_2 unidades que no han respondido, se selecciona una muestra aleatoria de tamaño $n_{21} \leq n_2$ y se envían entrevistadores para conseguir las respuestas de las n_{21} unidades seleccionadas en la segunda ocasión de entre las no respondientes a la primera.

El total muestral se puede descomponer como:

$$x = x_1 + x_2$$

donde x_2 corresponde a las unidades sin respuesta en la primera ocasión.

Si denotamos por $f_{21} = \frac{n_{21}}{n_2}$ la fracción de sub muestreo de la segunda ocasión, un estimador insesgado de x_2 , es:

$$\hat{x}_2 = \frac{1}{f_{21}} x_{21} = \frac{n_2}{n_{21}} x_{21}$$

siendo x_{21} el total de la segunda ocasión; en consecuencia un estimador insesgado del total poblacional, es :

$$\hat{X} = \frac{N}{n} (x_1 + \hat{x}_2) = \frac{1}{f} (x_1 + \hat{x}_2)$$

donde $f = \frac{n}{N}$ es la fracción de muestreo de la primera ocasión.

Para el calculo de la varianza del estimador utilizamos la técnica:

$$V(\hat{X}) = V E_w(\hat{X}) + E V_w(\hat{X})$$

donde E_w y V_w representan la esperanza y varianza condicional al conjunto de muestras w en las que n_1 y n_2 son fijos. Para el primer sumando tenemos:

$$E_w(\hat{X}) = E_w\left\{\frac{1}{f}(x_1 + \hat{x}_2)\right\} = \frac{1}{f}\left\{x_1 + E_w(\hat{x}_2)\right\} = \frac{1}{f}(x_1 + x_2) = \frac{1}{f}x = \frac{N}{n}x$$

$$V_w E_w(\hat{X}) = V\left(\frac{N}{n}x\right) = \frac{N^2}{n}(1-f)S^2; \text{ (muestreo sin reposición)}$$

Para el segundo sumando, denotamos por s_2^2 la cuasivarianza de las n_2 unidades no respondientes en la primera ocasión, entonces tenemos:

$$V_w(\hat{X}) = V_w\left\{\frac{1}{f}(x_1 + \hat{x}_2)\right\} = V_w\left(\frac{1}{f}\hat{x}_2\right) = \frac{1}{f^2}V_w(x_2) = \frac{1}{f^2}\frac{n_2^2}{n_{21}}(1-f_{21})s_2^2 = \frac{(1-f_{21})}{f^2 \cdot f_{21}}n_2s_2^2$$

$$E(V_w(\hat{X})) = \frac{(1-f_{21})}{f^2 \cdot f_{21}}E(n_2s_2^2)$$

El calculo de $E(n_2s_2^2)$ puede hacerse

$$E(n_2E(s^2/n_2)) = E(n_2S_2^2) = fN_2S_2^2$$

Por tanto

$$V(\hat{X}) = N^2(1-f)\frac{S^2}{n} + \frac{1}{f}\left(\frac{1}{f_{21}} - 1\right)N_2S_2^2 \dots \dots \dots (1)$$

donde S^2 es la cuasivarianza de las N unidades de la población y S_2^2 es la correspondiente al estrato de N_2 no respondientes.

De los dos sumandos de la fórmula (1), el primero representa la varianza del estimador del total si no hubiera falta de respuesta; el segundo expresa el incremento de varianza debido al sub muestreo de las no respondientes.

Si se trata de estimar la media X , formamos el estimador:

$$\hat{\bar{X}} = \frac{1}{Nf} (x_1 + \hat{x}_2) = \frac{1}{n} (x_1 + \hat{x}_2)$$

y su varianza, será:

$$V\left(\frac{\hat{\bar{X}}}{X}\right) = (1-f) \frac{S^2}{n} + \frac{N_2}{nN} \left(\frac{1}{f_{21}} - 1\right) S_2^2$$

Consideremos ahora la función de costo total:

$$C = c_0 n + c_1 n_1 + c_2 n_{21}$$

donde el primer sumando representa el costo de selección de la muestra de tamaño n y los otros dos, los costes de las entrevistas en primera y segunda fase. Puesto que n_2 es aleatorio, fijada una fracción de sub muestreo f_{21} también $n_{21} = f_{21} \cdot n_2$ es aleatorio, entonces utilizamos el coste esperado que calculamos de la siguiente forma:

$$C = n \left(c_0 + c_1 \frac{n_1}{n} + c_2 \frac{n_{21}}{n} \right) = n \left(c_0 + c_1 \frac{n_1}{n} + c_2 f_{21} \frac{n_2}{n} \right)$$

$$E(C) = n(c_0 + c_1W' + c_2f_{21}W)$$

Siendo $W' = (1 - W)$ la proporción teórica de respondientes.

Para optimizar el procedimiento podemos fijar la precisión (varianza del estimador) en un valor V minimizar el costo esperado.

Aplicando el método de multiplicadores de Lagrange para optimizar respecto a n , tamaño total de la muestra, y de la fracción f_{21} de sub muestreo de no respondientes, escribimos:

$$\phi = n(c_0 + c_1W' + c_2Wf_{21}) + \lambda \left\{ \frac{N(N-n)}{n} S^2 + \frac{NN_2}{n} \left(\frac{1}{f_{21}} - 1 \right) S_2^2 - V \right\}$$

y derivando respecto de n , de f_{21} y junto con la ecuación de condición resulta el sistema de ecuaciones:

$$(c_0 + c_1W' + c_2Wf_{21}) + \lambda \left\{ \frac{-N}{n^2} S^2 - \frac{NN_2}{n^2} \left(\frac{1}{f_{21}} - 1 \right) S_2^2 \right\} = 0$$

$$c_2Wn + \lambda \left(\frac{-NN_2S_2^2}{f_{21}^2} \right) = 0$$

$$\frac{N(N-n)}{n} S^2 + \frac{NN_2}{n} \left(\frac{1}{f_{21}} - 1 \right) S_2^2 = V$$

despejando λ de la primera y segunda, igualamos y después de simplificar resulta:

$$Wc_2f_{21} \{ NS^2 f_{21} + (1 - f_{21})N_2S_2^2 \} = (c_0 + c_1W')N_2S_2 + N_2S^2Wc_2f_{21}$$

$$Wc_2f_{21}(N^2S^2 - N_2S_2^2) = (c_0 + W'c_1)N_2S_2^2$$

de donde se puede despejar explícitamente f_{21} :

$$f_{21} = \sqrt{\frac{c_1}{c_2} \cdot \frac{S_2^2 (W' + c_0 / c_1)}{S^2 - WS_2^2}}$$

donde se observa que cuando mayor es c_2 respecto de c_1 , menor es la fracción de sub muestreo de no respondientes que debemos tomar.

También puede escribirse f_{21} en la forma:

$$f_{21} = \sqrt{\frac{S_2^2 (W' c_1 + c_0)}{c_2 (S^2 - WS_2^2)}}$$

y la hipótesis de trabajo $S_2^2 = S^2$ (cuasivarianza del estrato de no respondientes igual a la cuasivarianza poblacional), nos da:

$$f_{21} \approx \sqrt{\frac{W' c_1 + c_0}{W' c_2}}$$

Para despejar el valor óptimo de n , de la ecuación de condición obtenemos:

$$\frac{N^2 S^2}{n} - NS^2 + \frac{NN_2 \left(\frac{1}{f_{21}} - 1 \right) S_2^2}{n} = V; \text{ de donde}$$

$$n = \frac{N^2 S^2 + NN_2 \left(\frac{1}{f_{21}} - 1 \right) S_2^2}{V + NS^2}; \text{ que puede escribirse de la siguiente forma:}$$

$$n = \frac{N^2 S^2}{V + NS^2} + \frac{NN_2 \left(\frac{1}{f_{21}} - 1 \right) S_2^2}{V + NS^2}$$

Donde el primer sumando representa el tamaño de la muestra necesario para estimar el total con un error de muestreo $E = \sqrt{V}$ si no hubiera falta de respuesta. El valor numérico del óptimo tamaño n , se obtiene sustituyendo el f_{21} calculado antes.

En caso de muestreo con reposición operando de forma análoga, se llega a los siguientes valores óptimos:

$$n = \frac{N^2 \sigma^2}{V} + \frac{NN_2 \sigma_2^2 / f_{21}}{V}$$

Donde f_{21} debe tomarse de la fórmula anterior.

b. El método de visitas sucesivas

Un procedimiento para aminorar la no respuesta, especialmente cuando es producida por ausencia de la persona específicamente seleccionada para la encuesta, consiste en realizar visitas sucesivas; para ello se fija un número mínimo de “revisitas” que deben hacerse a cada unidad antes de abandonarla como “contacto imposible”. Un elemento importante a considerar es si la entrevista se dirige a un adulto cualquiera del hogar seleccionado, o bien el cuestionario debe ser respondido por un adulto específico (cabeza de familia, por ejemplo). En el primer caso el número de respuestas en la primera visita es mucho mayor que en el segundo, pero a medida que aumenta el número de revisitas el porcentaje total de respuestas tiende a igualarse, si bien es mayor en el primer caso. Para la aplicación y análisis del rendimiento de esta técnica es indispensable un buen estudio de los costos promedios por unidad en las

sucesivas revisitas. El procedimiento de revisitas retrasa los resultados finales, aunque no siempre será necesariamente un retraso excesivo si de antemano se planifica la duración de la recogida de datos.

b.1. Modelo de Deming

Deming desarrollo el siguiente modelo, muy útil para el estudio de esta técnica de revisitas. La población se divide en L clases de acuerdo con la probabilidad de que el encuestado sea encontrado en casa; sean:

w_{ij} = Probabilidad de que un encuestado de la j clase sea encontrado en i visitas

P_j = Proporción de la población en la j clase

\bar{X}_j = Media poblacional de la j clase

σ_j^2 = Varianza poblacional de la j clase

y supongamos $w_{ij} > 0$ para todas las clases, aunque el modelo puede adaptarse para incluir personas imposibles de encontrar. Después de i visitas la composición de la muestra es la siguiente: L clases, a cada una de las cuales pertenecen los elementos del estrato j de los cuales se ha obtenido respuesta durante alguno de los i primeros intentos, mas una clase que incluye todos los elementos de la muestra inicial de los que no se ha obtenido respuesta en los mencionados intentos. Si es n_0 el tamaño inicial de la muestra, el numero n_{ij} de elementos observados de clase j después de i visitas, es una multinomial de parámetros:

$$n_0; \left\{ w_{i1}P_1; w_{i2}P_2; \dots; w_{iL}P_L; \left(1 - \sum_{j=1}^L w_{ij}P_j \right) \right\}$$

El numero total de respuestas después de i visitas en una binomial con parámetros:

$$n_i : \left\{ n_0; \sum_{j=1}^L w_{ij}P_j \right\}$$

por lo tanto el numero esperado de respuestas en i visitas:

$$E(n_i) : n_0 \sum_{j=1}^L w_{ij}P_j$$

Para un n_i fijo, el numero de entrevistas obtenidas en cada una de las clases, viene dada por la multinomial condicionada que como sabemos también es multinomial de parámetros:

$$\left\{ n_i; \frac{w_{i1}P_1}{\sum w_{ij}P_j}, \frac{w_{i2}P_2}{\sum w_{ij}P_j}, \dots, \frac{w_{iL}P_L}{\sum w_{ij}P_j} \right\}$$

por tanto:

$$E(n_{ij} | n_i) = \frac{n_i w_{ij} P_j}{\sum_j w_{ij} P_j}$$

y la esperanza matemática de la media muestral condicionada a n_i , es:

$$E(\bar{x}_i | n_i) = E\left(\frac{\sum n_{ij} \bar{x}_{ij}}{n_i} \right) = \frac{\sum_j n_i w_{ij} P_j \bar{X}_j}{n_i \sum_j w_{ij} P_j} = \frac{\sum_j w_{ij} P_j \bar{X}_j}{\sum_j w_{ij} P_j} = \bar{X}_i$$

por tanto vemos que no depende de n_i y en consecuencia podemos afirmar que la media de la muestra después de i visitas,

independientemente de cual sea el numero n_i de respuestas, es también

\bar{X}_i . El sesgo de la media muestral después de i visitas es por tanto:

$$B(\bar{X}_i) = \bar{X}_i - \bar{X}$$

La varianza de \bar{x}_i condicionada a un n_i , es:

$$V(\bar{x}_i | n_i) = \frac{\sum_j w_{ij} P_j \left(\sigma_j^2 + (\bar{X}_j - \bar{X}_i)^2 \right)}{n_i \sum_j w_{ij} P_j} \dots\dots\dots(1)$$

y la varianza de \bar{x}_i , cualquiera que sea el numero n_i de respuestas obtenidas después de i visitas, viene dada aproximadamente (despreciando términos de orden $1/n_i^2$ por la formula (1) sustituyendo en ella n_i por su valor esperado, resultando:

$$V(\bar{x}_i) \approx \frac{\sum_j w_{ij} P_j \left(\sigma_j^2 + (\bar{X}_j - \bar{X}_i)^2 \right)}{n_0 \sum_j w_{ij} P_j} \dots\dots\dots(2)$$

Finalmente el Error cuadrático Medio de la media \bar{x}_i obtenida después de i visitas, es:

$$ECM(\bar{x}_i | i) = V(\bar{x}_i | i) + (\bar{X}_i - \bar{X})^2$$

Veamos ahora el costo de hacer i visitas. Sea c_r el costo por unidad en la r^a visita. Por otra parte el número esperado de entrevistas conseguidas al pasar de la visita $(r - 1)$ a la r , es:

$$\sum_j (w_{rj} - w_{r-1,j}) P_j$$

Por tanto el costo esperado para i -visitas, siendo n_0 el tamaño inicial de la muestra es:

$$n_0 \{ c_1 \sum w_{1j} P_j + c_2 \sum (w_{2j} - w_{1j}) P_j + \dots + c_i \sum (w_{ij} - w_{i-1,j}) P_j \}$$

c. El método de eliminación de visitas de insistencia.

Politz y Simmons (1949) propusieron un método ingenioso pero, en último análisis, sesgado y poco práctico para eliminar las visitas de insistencia.

Consistía en visitar las unidades una sola vez, en horas del día seleccionadas al azar. A los respondientes contactados se les preguntaba cuán a menudo, en los cinco días previos, habían estado en sus casas a esa misma hora, y se ponderaba consiguientemente las respuestas de la muestra; por ejemplo, los que se encontraban en sus casas los seis días recibían sólo la mitad de la ponderación que aquellos que sólo estaban en sus casas en tres de los seis días. Son muy evidentes los sesgos e inconvenientes potenciales de ejecución de este método y no se lo ha empleado en ninguna encuesta patrocinada por gobiernos.

c.1. El método Politz-Simmons

No se realiza mas que una visita pero se ponderan los resultados de acuerdo con la probabilidad de encontrar al encuestado. Dado que una persona que esta en casa una proporcion π del tiempo tiene una probabilidad relativa de ser entrevistado igual a π , su respuesta deberia recibir una ponderación $\frac{1}{\pi}$. Suponiendo entonces que las visitas se realizan durante 6 tardes de cada semana, al entrevistado se le pregunta si estaba en casa cada una de las cinco tardes anteriores a la de la entrevista; si declara que estuvo $t = 0,1,2,3,4,5$ se toma como estimador de π :

$$\hat{\pi} = \frac{1+t}{6}$$

El estimador ponderado de Politz y Simmons para la media es, entonces:

$$\bar{x}_{PS} = \frac{\sum_{t=0}^5 6n_t \bar{x}_t / (1+t)}{\sum_{t=0}^5 6n_t / (1+t)}$$

Donde n_t es el numero de entrevistados que declaran haber estado en casa t de las cinco noches anteriores y \bar{x}_t su media. Este estimador es menos sesgado que la media simple de la muestra de respuestas obtenidas en la visita pero su varianza será mayor ya que las ponderaciones son estimadas.

d. Añadido de una muestra de unidades no respondientes de encuestas previas a la muestra inicial. Este método, que ha utilizado la Universidad de Michigan

(SurveyResearch Center) fue descrito en un artículo de Kish y Hess (1959). Se aumenta la muestra inicial de una encuesta con direcciones de las cuales no se obtuvo respuesta en estudios similares recientes. Se hace un número especificado de visitas para obtener respuestas de los dos conjuntos de direcciones. Posteriormente, las respuestas obtenidas en la encuesta en curso, para unidades no respondientes de las anteriores, se consideran como "reemplazantes" de unidades no respondientes de la nueva muestra de la encuesta que se está realizando. Los autores sostienen que "...en términos aproximados...las visitas a las direcciones reemplazantes se pueden considerar como adicionales de las realizadas originalmente a ellas".

Bajo ciertos supuestos razonables esta técnica puede reducir el sesgo de falta de respuesta. Sin embargo, no es probable que tenga efecto si una gran parte de la falta de respuesta se origina en negativas y su uso sólo es factible en los casos en que se realizaron encuestas previas con contenido y procedimientos de recolección muy similares. Además, introduce ciertas complicaciones en el diseño de la encuesta y requiere controles muy cuidadosos para su adecuada ejecución.

e. Sustitución. La sustitución de unidades no respondientes por otras se ha utilizado en varias encuestas de hogares, tanto de los países en desarrollo como de los desarrollados. La razón de su uso consiste, por lo general, en asegurar la

obtención de entrevistas completas para el número exacto de hogares de la muestra especificada en el diseño inicial.

Se han empleado muchos métodos distintos de sustitución. Un procedimiento común consiste en dividir la muestra de hogares de cada conglomerado definitivo, al azar (o sistemáticamente), en dos grupos: la muestra asignada y la muestra de reserva. En determinadas condiciones, que usualmente comprenden la realización de cierto número de visitas de insistencia, si no ha sido posible concretar una entrevista el agente puede reemplazarla por la primera unidad disponible de la muestra de reserva.

La sustitución no elimina el sesgo de falta de respuesta. Esto se puede comprender si se considera el universo de la encuesta como dividido en dos grupos: los hogares de los cuales, siguiendo procedimientos especificados, es posible obtener una entrevista, . y aquellos en los que, con esos mismos procedimientos, no es posible obtenerla. La sustitución aumenta el tamaño de la muestra correspondiente al primer grupo, pero no da ninguna representación a los hogares del segundo. Seguramente las características de los dos grupos han de diferir, y el proceso de sustitución no hace nada para reducir el sesgo resultante de las diferencias.

La sustitución, en efecto, controla el error de muestreo al obtenerse el tamaño deseado de la muestra, pero esto se puede conseguir casi tan bien con una muestra inicial lo suficientemente grande como para admitir una proporción de hogares que no se puede entrevistar.

El principal argumento contra, la sustitución es que, en la práctica, su uso probablemente introducirá sesgos adicionales.

Esto ocurre porque:

- i) Frecuentemente las reglas establecidas para la sustitución son sesgadas; por ejemplo, cuando se sustituye una unidad de vivienda vacía por otra ocupada, o cuando se utiliza un hogar de reserva porque en la unidad de vivienda previamente ocupada por el hogar especificado de la muestra existe ahora un hogar nuevo.
- ii) Es extremadamente difícil impedir que los entrevistadores hagan sustituciones no autorizadas. Será muy tentador hacer una sustitución después de visitar sin éxito un hogar, en lugar de regresar las veces que se ha especificado.
- iii) El uso de la sustitución desvía la atención del problema del sesgo de falta de respuesta, sin hacer nada por aliviarlo. Esencialmente, excluye todo esfuerzo por reducir los efectos de ese sesgo mediante el uso de procedimientos apropiados de estimación e imputación.

3.6. RESULTADOS HEURISTICO DE TASAS DE NO RESPUESTA

Existen factores que son importantes para mejorar la cooperación de los encuestados. Entre ellos tenemos:

3.6.1. Identificación del encuestador y la Tasa de No respuesta

Es importante que todo encuestador lleve consigo una identificación en un lugar visible, esta identificación debe tener el nombre de la institución que organiza el estudio y los teléfonos donde se puede dar mayores informes sobre los objetivos y finalidad del estudio, su nombre completo, su número de DNI con una foto nítida y reciente.

En el año 2002 el Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería realizó el seguimiento de este factor “identificación”, encontrándose los siguientes resultados provenientes de Encuestas de Opinión Política.

TABLA N° 1

TASA DE NO RESPUESTA SEGÚN EL FACTOR “IDENTIFICACIÓN DEL ENCUESTADOR”

Fecha de trabajo de campo de las encuestas	TASA DE NO RESPUESTA	
	POSEE IDENTIFICACION	NO POSEE IDENTIFICACION
Encuesta de Abril del 2002	10%	30%
Encuesta de Mayo del 2002	7%	20%
Encuesta de Junio del 2002	5%	15%
Encuesta de Noviembre del 2002	11%	40%

FUENTE: Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería - 2002

Podemos observar que la diferencia entre el grupo que posee identificación y los que no poseen es significativa, es decir que el efecto de no tener la debida identificación del encuestador durante el trabajo de campo aumenta la negativa del encuestado a ser entrevistado y por consiguiente aumenta la tasa de no respuesta.

3.6.2. Carga de respuesta y Tasa de No respuesta

A menudo se vincula la falta de respuesta al tiempo que se tarda el encuestado para diligenciar una encuesta. Este tiempo esta en función del

numero de preguntas del cuestionario; en preguntas que el encuestado tiene que recurrir a su memoria.

La cantidad de preguntas que contiene el cuestionario a aplicarse en una encuesta puede influir no sólo en la tasa de respuesta sino también en la calidad de la información obtenida.

En una encuesta de Opinión Política realizada en octubre del 2002 por el Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería se realizó el seguimiento de este factor “carga de respuesta”, encontrándose los siguientes resultados:

TABLA N° 2

**TASA DE NO RESPUESTA SEGÚN EL FACTOR
“CARGA DE RESPUESTA DEL ENCUESTADO”**

Carga de respuesta: Tamaño del cuestionario medido en Número de Preguntas	TASA DE NO RESPUESTA
5 preguntas	2%
10 preguntas	5%
22 preguntas	12%
30 preguntas	20%
30 preguntas	30%

FUENTE: Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería - 2002

Podemos observar que la diferencia es significativa entre el grupo de encuestadores que tuvieron que utilizar un cuestionario cuyo número de preguntas era mayor tuvieron una mayor tasa de no respuesta con respecto a los que su cuestionario tenía menos preguntas. El valor promedio de tasa de no respuesta de esta encuesta es de 13.8%. Según la experiencia empírica del autor de esta monografía un tamaño de cuestionario recomendado es aquel que contenga en promedio 22 preguntas, lo cual pudo comprobarse en esta encuesta.

3.6.3. Encuestadores y las Tasas de No respuesta

Otro factor que ha sido objeto de estudio es la relación entre las características de los entrevistadores tales como la edad, el sexo, el nivel de educación, la experiencia y la raza, y las tasas de respuesta. Estos estudios describen situaciones específicas y no es posible hacer generalizaciones. En un estudio de salud efectuado por el Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería donde se utilizó como encuestadores a las amas de casa, graduados y enfermeras no se pudo determinar la eficiencia del diligenciado de las encuestas, es decir los porcentajes de no respuesta no tenían un comportamiento similar dentro de cada segmento. A pesar que se esperaba que los graduados y las enfermeras tendrían las tasa de no respuesta mínimas o respecto a las amas de casa.

La tabla N° 3 muestra como en muchos casos una proporción relativamente pequeña de encuestadores puede ser responsable de una parte sustancial de la falta de respuesta.

Es importante identificar a esos encuestadores y volver a adiestrarlos o reemplazarlos cuando sea necesario.

TABLA N° 3

TASA DE NO RESPUESTA SEGÚN EL FACTOR “ENCUESTADOR”

	ENCUESTADOR													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
% de No respuesta	7%	5%	15%	10%	20%	6%	4%	12%	8%	25%	40%	11%	8%	5%

FUENTE: Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería - 1999

Podemos observar que el porcentaje de no respuesta promedio en esta encuesta es de 12.6%. El valor mínimo lo tiene el encuestador 7 con una tasa de no respuesta de 4% y el valor máximo lo tiene el encuestador 11 con una tasa de no respuesta del 40%. Si utilizamos el criterio de cambiar o capacitar a todos aquellos encuestadores cuya tasa de no respuesta es mayor al promedio, esperando que tengan después de ello un comportamiento similar

al otro grupo que tuvieron una tasa de no respuesta igual o menor al promedio tendríamos el siguiente cuadro:

TABLA N° 3A

**TASA DE NO RESPUESTA SEGÚN EL FACTOR “ENCUESTADOR”
(Eliminando los encuestadores 5, 10 y 11)**

	ENCUESTADOR											
	1	2	3	4	6	7	8	9	12	13	14	
% de No respuesta	7%	5%	15%	10%	6%	4%	12%	8%	11%	8%	5%	

FUENTE: Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería - 1999

En este cuadro hemos eliminado a los encuestadores 5, 10, 11 por tener su tasa de no respuesta por encima del promedio inicial (12.6%); obteniendo con ello una nueva tasa de no respuesta promedio de 8.3%. Según experiencia empírica es una tasa de no respuesta buena para un estudio.

3.6.4. Características de los respondientes y las Tasas de No Respuesta

Es importante tratar de identificar las características de los no respondientes. Esto es necesario para idear medidas para reducir la tasa de no respuesta y

hacer los ajustes necesarios a fin de tenerla en cuenta y estimar sus efectos en los resultados de la encuesta.

En una encuesta de Opinión Política realizada el 31 de Mayo del 2003 por el Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería se investigo las tasas de no respuesta de algunas características consideradas según experiencia empírica factores que influyen en las tasa de no respuesta promedio de una encuesta.

Los factores que se consideraron influyentes fueron los siguientes:

- Factor distrito
- Factor Nivel Socioeconómico
- Factor grupo de edad
- Factor Sexo

A continuación presentamos algunas tablas de las tasa de no respuesta según estos factores.

TABLA N° 4

**TASA DE NO RESPUESTA SEGÚN EL FACTOR
“DISTRITO”**

DISTRITO	TASA DE NO RESPUESTA
VILLA EL SALVADOR	5%
SAN JUAN DE LURIGANCHO	8%
BREÑA	10%
CALLAO	15%
SAN ISIDRO	20%

FUENTE: Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería - 2003

La tasa de No respuesta menor de este grupo de distritos lo tiene el distrito de Villa el Salvador (5%) , siendo el distrito de San Isidro el que tiene la mayor tasa de no respuesta (20%). Según experiencia empírica del autor de esta monografía una tasa de no respuesta buena es aquella que es inferior al 12%, siendo por lo tanto los distritos de San Juan de Lurigancho y el de Breña, distritos cuya tasa de no respuesta esta por debajo del 12%.

TABLA N° 5

**TASA DE NO RESPUESTA SEGÚN EL FACTOR
“NIVEL SOCIOECONOMICO”**

NIVEL SOCIOECONOMICO	TASA DE NO RESPUESTA
ALTO	20%
MEDIO ALTO	15%
MEDIO ALTO	10%
MEDIO BAJO	8%
BAJO	7%

FUENTE: Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería - 2003

Según la clasificación de niveles socioeconómicos en Lima Metropolitana, el autor de esta monografía sabe por experiencia empírica que los niveles socioeconómicos alto y medio alto son los mas difíciles de conseguir que contestes las preguntas de una encuesta, siendo los niveles mas bajos los mas accesibles en términos generales. Esta experiencia empírica se pudo comprobar en esta encuesta donde se encontró que a medida que los niveles socioeconómicos disminuyen también lo hacen las tasa de no respuesta. Es así que el nivel socioeconómico alto tiene una tasa de no respuesta del 20% frente al nivel socioeconómico bajo que tiene una tasa de no respuesta del 7%. El valor promedio de tasa de no respuesta de este encuesta fue del 12%.

TABLA N° 6

TASA DE NO RESPUESTA SEGÚN EL FACTOR

“GRUPO DE EDAD”

GRUPO DE EDAD	TASA DE NO RESPUESTA
MENOR DE 50 AÑOS	9%
50 AÑOS O MAS	29%

FUENTE: Instituto de Investigaciones Económicas y Ciencias Sociales de la Universidad Nacional de Ingeniería - 2003

Las personas de 50 años o mas de edad tuvieron una tasa de no respuesta mayor (29%) al grupo que tenia menos de 50 años de edad (9%). No se encontró diferencia en el sexo de las personas.

CAPITULO 4

4. CONCLUSIONES Y RECOMENDACIONES

4.1. Conclusiones

Las conclusiones mas relevantes de esta monografía son las siguientes:

1. Se planteo y utilizo el concepto de error total como reemplazo al error muestral en encuestas de hogares.
2. Las causas mas comunes de no respuesta son la falta de acceso a las unidades de la muestra, la falta de contacto con los respondientes y la falta de cooperación o negativa .
3. El efecto de la no respuesta de un grupo de unidades muestrales puede invalidar un estudio.

4. El factor identificación del encuestador influye significativamente en las tasas de no respuesta de los encuestados.
5. El factor carga de respuesta del encuestado influye significativamente en las tasas de no respuesta de los encuestados.
6. El Factor encuestador influye significativamente en las tasas de no respuesta de un estudio.
7. El factor distrito influye significativamente en las tasa de no respuesta de un estudio.
8. El factor nivel socioeconómico influye significativamente en las tasa de no respuesta de un estudio.
9. El factor grupo de edad influye significativamente en las tasa de no respuesta de un estudio.

4.2. Recomendaciones

Las recomendaciones de esta monografía son las siguientes:

1. Se recomienda realizar un estudio experimental secuencial para determinar las tasas de no respuesta promedio en encuestas políticas en hogares para cada uno de los factores considerados relevantes por el autor de esta monografía.

2. Se recomienda incluir como información básica en todo estudio las tasas de no respuesta deseada por el cliente o propuesta por el estadístico y las tasa de no respuesta real promedio del trabajo de campo.

3. Se recomienda por experiencia empírica que una tasa buena de no respuesta en una encuesta de hogares no debe exceder el 12%.

Bibliografía

1. Programa de las Naciones Unidas para el mejoramiento de la calidad de datos. ONU 1989.
2. Deming (1953). On a probability mechanism to attain an economic balance between the resultant error of non-response and the bias of non-response. Jour. Amer. Stat. Assoc. Assoc.
3. Hansen y Hurwitz (1946). The problem of nonresponse in sample surveys. Jour. Amer. Stat. Assoc.
4. Kish y Hess (1959). A replacement procedure for reducing the bias of nonresponse. Amer. Statistician.
5. Kish leslie, 1982. Muestreo de encuestas. Ed. Trillas S.A. México.

6. **Ricardo Hirata, Como implementar Técnicas Estadísticas de Calidad.**
7. **Cochran William, 1976. Técnicas de muestreo. Compañía Editorial Continental S.A. México.**