

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Facultad de Estudios Superiores Acatlán



“DISEÑO MUESTRAL PARA UNA ENCUESTA DE
SEGUIMIENTO OPERATIVO A LAS FAMILIAS
BENEFICIARIAS DEL PROGRAMA DE DESARROLLO
HUMANO OPORTUNIDADES”

T E S I S A

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I O

PRESENTA:

MARTHA CUEVAS DURAN

ASESOR:

MTRO. JOSÉ ELIUD SILVA URRUTIA



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos y Dedicatorias:

A mis padres Antonio Cuevas López † y Elvira Durán Bojorges, por su apoyo en todo momento ya que sin ellos, nunca hubiera logrado terminar la carrera.

A mis hermanos Sofía, Roberto, Elvia, Irma, Emma, Irene, Ma. Antonia, Luis Antonio y Claudia, por todo el apoyo que siempre me brindaron.

A mis hijos Iván y Julio, por toda la felicidad y cariño que me dan.

A mi esposo Oswaldo Rodríguez Facio, por todo su apoyo y comprensión.

CONTENIDO

Introducción	4
1. El Programa de Desarrollo Humano Oportunidades	6
1.1 Antecedentes.....	6
1.2 Diseño del Programa de Desarrollo Humano Oportunidades.....	11
1.3 Esquema bajo el que opera el Programa Oportunidades.....	15
1.4 Proceso de evaluación.....	24
2. Técnicas de Muestreo	32
2.1 El por qué del muestreo.....	32
2.2 Conceptos básicos del muestreo	37
2.3 Muestreo simple aleatorio (msa).....	42
2.4 Muestreo sistemático	51
2.5 Muestreo estratificado.....	55
2.6 Muestreo por conglomerados y muestreo por etapas.....	64
3. El diseño muestral para el seguimiento operativo	75
3.1 Diseños muestrales	75
3.2 Diseño utilizado en Oportunidades	84
3.3 Diseño sugerido	95
Conclusiones	104
Anexo 1. Notas metodológicas utilizadas para el cálculo de muestra en Oportunidades	107
Anexo 2. Encuesta aplicada a las familias beneficiarias en las localidades en muestra.....	113
Bibliografía y referencias.....	127

Introducción

En distintas empresas privadas o dependencias del gobierno federal, existe una forma inadecuada de obtener los tamaños de muestra para distintos estudios, ya sea para realizar encuestas por muestreo o realizar alguna auditoría o revisar expedientes. En general, la forma de obtener el tamaño de la muestra es obtenida mediante porcentajes de la población o por cuotas o por diseños muestrales sin sustento estadístico.

Uno de los errores más frecuentes en el ámbito del muestreo no probabilístico es encontrar que se llega a afirmar que porque se tiene un tamaño de muestra determinado se tiene un margen de error específico (generalmente muy bajo), esto extrapolado de lo resultante de un diseño aleatorio simple. Sin duda para muchos operativos o tomadores de decisiones esto carece de relevancia alguna.

Aún en las encuestas en las que se dice: se estratificaron las entidades federativas, se seleccionaron aleatoriamente las localidades y en ellas se seleccionaron a juicio entrevistados -una etapa de diseño y dos de muestreo- se reporta que con 400 entrevistados “se tiene un margen de error de ± 0.05 ”. En este sentido se omite el efecto de la conglomeración de localidades. En el muestreo no probabilístico, el cálculo del tamaño muestral, siguiendo las referencias de los diseños probabilísticos, carece de validez estadística.

Tampoco tiene validez estadística el aplicar una “regla de dedo” o “tradicional” por ejemplo: “tomar el 10% de la población”, “buscar que haya 1000 entrevistas”, etcétera, debido a que, al igual que los casos mencionados, no hay ningún fundamento estadístico. Estas deficiencias, se desprenden en muchos casos de que se crea que el tamaño de muestra dependa de la población. Este es un error grave en muestreo no probabilístico, pero lo es mayor aún en el tipo de muestreo probabilístico.

Dentro de los programas de la administración pública, se tiene el Programa de Desarrollo Humano Oportunidades. En este programa es particularmente importante tener un diseño muestral que sea representativo que, dicho sea de paso, en la actualidad se carece. Así pues, el objetivo de la presente investigación es diseñar un plan de muestreo de seguimiento operativo sobre las familias beneficiarias del Programa de Desarrollo Humano Oportunidades que sea representativa tanto por Entidad Federativa y como a Nivel Nacional mediante técnicas de muestreo probabilístico. Es importante señalar que no se pretende un diseño muestral para la asignación de recursos dentro del Programa– tema sin duda de alta trascendencia-, sino exclusivamente para el seguimiento de las familias beneficiadas del mismo.

Si se utiliza un diseño muestral probabilístico para las encuestas a familias beneficiarias del Programa, se tendrán las bases técnicas para afirmar que el tamaño de muestra es suficiente para garantizar una muestra representativa en los dos niveles referidos. De hecho el documento metodológico realizado por Oportunidades señala que la muestra es representativa a nivel estatal con determinadas características, lo cual es incorrecto al no realizar el cálculo del tamaño de muestra mediante técnicas estadísticas.

La tesina está formada por tres capítulos. En el Capítulo I se describe de manera la forma en que opera el Programa de Desarrollo Humano Oportunidades y se describen sus antecedentes, su diseño y su forma de evaluar. Por otra parte, en el Capítulo II se da un panorama básico de las técnicas de muestreo más empleadas comúnmente (muestreo aleatorio simple, muestreo estratificado y muestreo de conglomerados). Finalmente en el Capítulo III se habla acerca de los diseños muestrales, se plantea la forma en que Oportunidades realiza el muestreo para el seguimiento operativo y se propone una alternativa de mejora sobre dicho diseño muestral. Asimismo se enmarcan algunas conclusiones derivadas de la presente investigación.

1. El Programa de Desarrollo Humano Oportunidades

1.1 Antecedentes

A lo largo de toda la historia han existido desigualdades sociales, no ha surgido aún, para muchos, la utópica sociedad igualitaria que supere un fenómeno tan ligado al devenir humano. En las diversas sociedades, el sistema de distribución del poder y del dinero ha asumido formas distintas. En la sociedad industrial a los pobres no les llega prácticamente nada y, en consecuencia, son atrapados en un círculo vicioso que parece no tener salida y prevalece en muchos aspectos hasta nuestros días.¹

En la actualidad, se han definido tres niveles de pobreza. El nivel más profundo se denomina pobreza alimentaria. Se considera que las personas y los hogares viven en esta condición cuando aún si asignaran todo el ingreso que perciben para cubrir sus necesidades de alimentación, éste es insuficiente para garantizar el consumo mínimo establecido en una canasta alimentaria normativa universal. Estas familias, por tanto, no cuentan siquiera con los recursos mínimos para la adquisición de los alimentos necesarios para su subsistencia, tampoco para desarrollar sus capacidades o cubrir cualquier necesidad adicional como vestido o vivienda.

El siguiente nivel se denomina pobreza de capacidades. Se considera que las personas y los hogares viven en condiciones de pobreza de capacidades cuando sus ingresos son insuficientes para cubrir conjuntamente sus necesidades básicas de alimentación, salud y educación. Por último, se denomina pobreza de patrimonio la que enfrentan las personas y los hogares cuando sus ingresos no son suficientes para satisfacer, en conjunto, sus necesidades de alimentación, salud, educación, vivienda, vestido y transporte.

¹ Madariaga Orozco, Camilo; Omar Sierra. 2000. "Redes Sociales y Pobreza". Psicología desde el Caribe, vol. num. 005 , p.128

La pobreza se ha *institucionalizado*, de modo que al no poder superarla, los integrantes de esos grupos han quedado fuera del sistema y se ha convertido en una institución con sus normas y valores propios, sin entrar a participar en el estratificado social. Tristemente, según lo indican las tendencias, la pobreza actual se transmite de generación en generación como una impronta: *los hijos de pobres serán pobres*.

La pobreza no conoce fronteras geográficas, se esparce por todos los continentes y está presente tanto en países industrializados como en países en desarrollo, aunque en diferentes grados. Causa niveles inadecuados de vida, salud precaria, hambre, viviendas insalubres o su falta, desempleo, exclusión social y analfabetismo. Trunca la vida de más de 1,500 millones de personas, número que se incrementa en por lo menos 25 millones al año. La mayoría de los pobres son mujeres, niños y ancianos y se ubican en el continente asiático. Asimismo la pobreza puede ser detonante para la migración del ámbito rural al urbano o internacional de países del tercer mundo a los del llamado mundo desarrollado.²

En América Latina, la situación de este grupo social se ha venido haciendo más crítica a raíz de la crisis económica de los países del área. Crisis que sobreviene después de lo que se conoce como los años dorados (1945-1973), periodo en el que los indicadores sociales como salud, vivienda, educación, servicios públicos presentaron repuntes importantes que aunque no redujeron el número de pobres sí su participación respecto del total de la población.³

A principios de 1995, a pesar de los esfuerzos realizados, la pobreza en México ya era un problema de suma gravedad. Más de la quinta parte de las familias no tenía

² Informe Mundial de Salud, 2005. <http://www.paho.org>

³ Madariaga Orozco, Camilo; Omar Sierra. 2000. "Redes Sociales y Pobreza". Psicología desde el Caribe, vol, num. 005 , p.128

ingresos suficientes para adquirir la canasta básica alimenticia con los nutrientes requeridos, lo que afectaba su salud, su capacidad de aprendizaje, el tipo de empleo al cual podía aspirar y, por tanto, sus posibilidades de acceder a mayores niveles de ingreso (fuentes estadísticas de información útiles para su estudio son la Encuesta Nacional de Ingreso Gasto (ENIGH) en sus distintos levantamientos). Asimismo, existían comunidades donde la desnutrición era todavía la tercera causa directa de mortalidad infantil. Por otro lado, alrededor de 10 millones de mexicanos carecían de acceso regular a los servicios básicos de salud. Igualmente, más de 1.5 millones de niños entre 6 y 14 años de edad no asistían a la escuela, y en las comunidades marginadas se observaban índices de analfabetismo que triplicaban el promedio nacional, y tasas de deserción y reprobación escolar que duplicaban las registradas a nivel nacional. Los indicadores de vivienda e ingreso, entre otros, mostraban una situación similar.⁴

La crisis económica de 1994-1995 podía agravar esta situación. En este contexto, el gobierno mexicano se propuso un plan que consistió en tratar de disminuir la pobreza y moderar la desigualdad que existía entre los diferentes estratos de la población. Al respecto, había que enfrentar cuatro retos. El primero era que alrededor de uno de cada cinco mexicanos vivía en condiciones de pobreza extrema; de ellos el 40% eran niños. A esto se unía la gravedad de los indicadores de bienestar de esa población, en especial de la que vivía en el medio rural. El segundo desafío era la relativa ineficacia de los instrumentos del gobierno federal disponibles en ese momento para combatir la pobreza, en particular los subsidios alimentarios. El tercero era el hecho de que la crisis económica presionaría fuertemente las finanzas públicas y limitaría los recursos disponibles para el gasto social. Por último, había una gran posibilidad de que aumentara el número de mexicanos en condiciones de pobreza extrema o que se profundizaran las

⁴ Levy, Santiago y Rodríguez, Evelyne, Sin Herencia de Pobreza, El Programa PROGRESA-Oportunidades de México, Banco Interamericano de Desarrollo, Planeta, 2005

condiciones de pobreza de los que ya estaban en ese momento como consecuencia de los efectos inmediatos de la crisis económica.⁵

Por otro lado, alcanzar un crecimiento económico sostenido era (y sigue siendo) indispensable para paliar el problema de la pobreza a través de generar empleos estables y bien remunerados, y crear condiciones materiales que permitieran un mejoramiento sustentable de los niveles de vida. Era claro que por sí solo sería y es insuficiente para ampliar las oportunidades de progreso en todas las regiones del país, para rediseñar una distribución más equitativa de la riqueza y para eliminar las condiciones de marginación en las que vivían y viven muchos mexicanos.

Por ello, a partir de 1995 era importante fortalecer las acciones de política social, en forma paralela a los programas para estabilizar la economía y recuperar el crecimiento. En particular, era necesario aumentar la intervención del Estado en materia social, pero modificando esta intervención de acuerdo con la experiencia adquirida en el pasado, adaptándola a la nueva realidad del país y al cambiante entorno internacional. Sin embargo, más allá de las opciones de respuesta de corto plazo a la crisis de 1994-1995, se requería una estrategia más amplia que permitiera avanzar de manera significativa y permanente en la mejora de las condiciones de bienestar de la población pobre.⁶

En consecuencia, a partir de 1996 se adecuaron y transformaron algunos de los programas existentes en este campo, además de crear otros nuevos, con el objeto de que simultáneamente se fortaleciera la conformación de una red de seguridad básica para proteger a la población de más bajos recursos, y se ampliaran sus oportunidades reales para participar y contribuir al crecimiento económico. En este contexto, y como parte de una estrategia más amplia para superar la pobreza

⁵ Ibid

⁶ Ibid

extrema y combatir la marginación, en 1997 se puso en marcha el Programa de Educación, Salud y Alimentación (PROGRESA) que más tarde fue nombrado Oportunidades.⁷

No toda la población pobre es igualmente pobre. Algunas de las familias pobres tienen niveles de ingreso y alcanzan satisfactores de educación y salud muy cercanos a los de las familias no pobres; otras viven en condiciones mucho más precarias. Esta información es importante para determinar a quién se debe ayudar primero o dónde debe darse el mayor esfuerzo, independientemente del apoyo que todas las familias pobres merecen recibir. Los índices de pobreza son una forma parcial e incompleta de capturar el fenómeno antes descrito. Pero a pesar de los problemas, tanto conceptuales como empíricos que tienen, aportan información que contribuye al diseño de las políticas de combate a la pobreza extrema.

Actualmente el programa Oportunidades intenta realizar un cambio de visión a la política de superación de la pobreza que se inició a mediados de los años noventa para transitar de acciones puramente asistencialistas que sólo transferían ingreso a través de diferentes medios, hacia programas que, al mismo tiempo que otorgan una transferencia de ingreso, fomentan la inversión en el desarrollo de capacidades, ampliando las oportunidades de las familias en pobreza para salir adelante por su propio esfuerzo.

Dicha óptica implicaría una transformación de los subsidios generalizados y de acciones aisladas de salud, alimentación y educación, en un esfuerzo integral, coordinado, continuo y de largo alcance, dirigido a las familias en condiciones de pobreza de capacidades. Lo anterior se sustenta en el supuesto de que estas familias están inmersas en un círculo vicioso que implica niveles altos de

⁷ Ibid

morbilidad, fecundidad y mortalidad infantil; incapacidad para enfrentar riesgos; incapacidad para demandar servicios educativos; pocos recursos dispersos entre familias numerosas; y, la transmisión de estas condiciones de una generación a otra. Las familias que enfrentan condiciones estructurales de pobreza se dice que no pueden beneficiarse plenamente de las acciones amplias de política social. Para romper con el círculo vicioso en que están inmersas sería necesario complementar y darle integridad a estas acciones:⁸

- concentrar esfuerzos adicionales en esta población con mecanismos de focalización transparentes y efectivos;
- promover la participación activa de los beneficiarios, fomentando su corresponsabilidad; y,
- enfrentar las condiciones que generan el círculo vicioso de la pobreza, considerando particularmente la interacción entre alimentación, salud y educación.

Actualmente, con las políticas del ejecutivo federal en cuestión de desarrollo social, se percibe la intención de que los diversos programas asistencialistas existentes confluyan en sus esfuerzos para que de manera integral se aspire a una mejora en la calidad de vida de las familias mexicanas (Vivir mejor).

1.2 Diseño del Programa de Desarrollo Humano Oportunidades

La Ley General de Desarrollo Social establece que son derechos para el desarrollo social, entre otros, la educación, la alimentación, el trabajo, la seguridad social y los relativos a la no discriminación; que toda persona tiene derecho a participar y beneficiarse de los programas de desarrollo social y que aquellos que se encuentren en situación de vulnerabilidad tienen derecho a recibir apoyo.

⁸ Ibid

Para lograr sus objetivos, la citada Ley establece el Sistema Nacional de Desarrollo Social, que es un mecanismo permanente de concurrencia, colaboración, coordinación y concertación de los Gobiernos Federal, Estatales y Municipales, así como con los sectores social y privado. Los ejes rectores definidos para la Política Social, se dirigen a: fomentar el desarrollo social con respeto a los derechos y dignidad de las personas, impulsar la superación de la pobreza desde una visión compartida e incluyente, que se concrete en las políticas públicas, fomentar la cultura de la corresponsabilidad, impulsar una coordinación eficiente entre los diferentes órdenes de gobierno, desde una perspectiva integral de promoción del desarrollo y la participación social.

Se trata de avanzar en el logro de la superación de la pobreza mediante el desarrollo humano integral incluyente y corresponsable, para alcanzar niveles suficientes de bienestar con equidad.⁹ La Ley General de Planeación por su parte, señala que la planeación nacional del desarrollo deberá llevarse a cabo como un medio para el eficaz desempeño de la responsabilidad del Estado sobre el desarrollo integral del país y deberá tender a la consecución de los fines y objetivos políticos, sociales, culturales y económicos contenidos en la Constitución Política de los Estados Unidos Mexicanos.

En algunas investigaciones, se ha encontrado que las familias pobres, por sus propias condiciones de marginación, pueden tener preferencias más marcadas por aumentar su consumo presente en relación con su consumo futuro. Por eso, los recientes programas contra la pobreza, vistos en conjunto, han tratado de eliminar, o al menos de reducir sustancialmente, el intercambio entre transferencias de ingreso que apoyen al consumo presente, y transferencias de ingreso que permitan hacer inversiones para un mayor consumo en el futuro.

⁹ Página electrónica de Oportunidades en: <http://www.oportunidades.gob.mx/>

Al mismo tiempo, se buscó adecuar los programas para combatir la pobreza extrema a la modernización económica, por un lado, y al cambio democrático, por el otro. De esta manera, se procuró un balance más adecuado entre subsidios a la oferta y la demanda y se buscó intervenir con instrumentos que eviten distorsiones en los precios relativos, reduciendo o eliminando las oportunidades para la búsqueda de rentas a través de arbitrajes o atesoramientos, o situaciones que propicien la corrupción; que generen incentivos adecuados al esfuerzo; y que distribuyan los riesgos apropiadamente, de un modo más general, las intervenciones públicas deben ayudar a las familias pobres a mejorar las condiciones en que participan en los mercados.

De forma esencial, también se intentó promover más transparencia y una mayor rendición de cuentas. Una forma de lograr lo anterior fue el establecimiento de la obligación legal, a partir de 1999 (a través del Decreto de Presupuesto de Egresos de la Federación), de expedir y publicar en el Diario Oficial de la Federación (DOF), así como en internet, reglas de operación para todos los programas que otorgaban subsidios. Ello obligó a precisar los objetivos de cada programa, la población a la que apuntaban, los montos de los beneficios, su mecánica de operación y las condiciones de acceso.¹⁰

Las disposiciones del Decreto de Presupuesto también obligaron a una evaluación sistemática de los resultados llevada a cabo por evaluadores externos y a la publicación de los resultados correspondientes. Por otro lado, se promovió la entrega directa de las transferencias a los beneficiarios, y no a través de gestores o intermediarios (como organizaciones de trabajadores o campesinos). Estos cambios intentan avanzar en la construcción de una relación más respetuosa y menos tutelar y clientelar entre las dependencias del Ejecutivo Federal y los beneficiarios de estos programas.

¹⁰ Levy, Santiago y Rodríguez, Evelyne. Op. Cit.

En ese marco se inscribe la acción del Programa Oportunidades, que tiene como objetivo general el de apoyar a las familias que viven en condición de pobreza extrema con el fin de potenciar las capacidades de sus miembros y ampliar sus alternativas para alcanzar mejores niveles de bienestar, a través del mejoramiento de opciones en educación, salud y alimentación, además de contribuir a la vinculación con nuevos servicios y programas de desarrollo que propicien el mejoramiento de sus condiciones socioeconómicas y calidad de vida. Es un programa interinstitucional en el que participan la Secretaría de Educación Pública, la Secretaría de Salud, el Instituto Mexicano del Seguro Social, la Secretaría de Desarrollo Social, y los gobiernos estatales y municipales.

Oportunidades inició su operación en 1997 en zonas rurales (bajo el nombre de PROGRESA) y desde entonces ha ampliado su cobertura hasta incursionar en las zonas metropolitanas. En su primer año, el Programa atendía a alrededor de 300 mil hogares rurales residentes en cerca de 11 mil localidades, y en el 2001 el padrón de beneficiarios ascendía a 3.2 millones de hogares residentes en alrededor de 68 mil localidades rurales y semiurbanas. Con esta dinámica, para 2004, el Programa ya había alcanzado su meta de crecimiento para el sexenio 2000-2006, al beneficiar a cinco millones de hogares en más de 86 mil localidades de toda la República Mexicana, es decir, alrededor de un cuarto de la población nacional. Actualmente el Programa tiene presencia en 93,065 localidades de todo el país, y continúa manteniendo un padrón de cinco millones de hogares beneficiarios.¹¹

El presupuesto del Programa se asigna en tres Secretarías: SEDESOL, SEP y SSA y la operación se rige por reglas establecidas por los titulares de esas secretarías, la SHCP y el IMSS. El presupuesto creció de \$9 mil millones a \$25 mil

¹¹ Página electrónica de Oportunidades en: <http://www.oportunidades.gob.mx/>

millones entre 2000 y 2004. Para 2007 el presupuesto se distribuye de la siguiente manera:

Presupuesto de egresos de la federación para el ejercicio fiscal 2007. Programa de desarrollo humano oportunidades (pesos)¹²

Desarrollo Social	15,538,048,145
Educación Pública	16,550,000,000
Salud	3,928,804,439
Total	36,016,852,584

Sumado el nuevo Apoyo Alimentario Vivir Mejor, Oportunidades ejerce durante 2008 un presupuesto de más de 42 mil millones de pesos, lo cual equivale a una derrama diaria de más de 115 millones de pesos.

1.3 Esquema bajo el que opera el Programa Oportunidades

Las condiciones de alimentación, educación y salud en que se encuentra la población más pobre del país, así como las pocas oportunidades que tienen para contar con una mejor vivienda, posibilidad de crédito, emprender actividades productivas, recibir capacitación o contar con un empleo estable, no permiten el adecuado desarrollo de las personas e impiden que participen plenamente en la sociedad.

Para contribuir a que la población con mayores carencias económicas y sociales pueda mejorar sus condiciones de vida, fue creado el Programa de Desarrollo Humano Oportunidades, que es un programa para mejorar la alimentación, la educación y la salud de las familias que viven en pobreza extrema, y que los jóvenes becarios, hijos de estas familias, puedan acceder a una educación superior o inicien un patrimonio.

¹² Anexo 18 del Presupuesto de egresos de la federación para el ejercicio fiscal 2007. *Nuevo Presupuesto DOF 28-12-2006*

Oportunidades coordina acciones con las instituciones de educación y salud para contribuir a la superación de la pobreza, mediante el desarrollo de las capacidades físicas y de aprendizaje de las personas, que facilite su acceso a mejores oportunidades de desarrollo económico y social, así como contribuir a la vinculación con otros servicios y programas de desarrollo. Para alcanzar su objetivo, Oportunidades cuenta con once orientaciones:

1. Contribuye al desarrollo humano, al apoyar de manera conjunta la alimentación, la educación y la salud de las familias con mayores carencias económicas.
2. Impulsa el desarrollo de capacidades físicas y de aprendizaje, al facilitar que cuenten con servicios de educación y salud, así como la adquisición de alimentos, que les permita, con su esfuerzo, alcanzar una vida plena y autosuficiente.
3. Se dirige a las familias de pobreza extrema para disminuir las causas que generan las condiciones en las que viven estas familias.
4. Se centra en la familia y fomenta el tejido social y comunitario, al considerar a la familia como el mejor lugar para enfrentar los problemas económicos y sociales; y mediante sus acciones, el Programa busca romper el círculo de la pobreza, y contribuye para que los ciudadanos lleven a cabo acciones conjuntas a favor de la superación de la misma.
5. Opera con transparencia y rinde cuentas. La identificación de las familias beneficiarias se lleva a cabo a partir de procedimientos rigurosos, objetivos e imparciales, iguales para todo el país que son verificables y auditables. Los apoyos monetarios se entregan en efectivo, directamente a la titular de la familia para evitar intermediarios. Se reportan y difunden los logros alcanzados, la utilización de los recursos asignados y la evaluación de las acciones del Programa. Informa a los beneficiarios sobre sus derechos y responsabilidades y atiende oportunamente las peticiones, dudas, quejas, denuncias y reconocimientos que realiza la ciudadanía.

6. Tiene enfoque de género y promueve que las mujeres tengan las mismas oportunidades que los hombres. Oportunidades canaliza los apoyos a través de las madres de familia e impulsa, entre las mujeres, becas educativas más altas para disminuir la desventaja de las niñas para seguir estudiando.
7. Fomenta la corresponsabilidad. El que las familias estén convencidas de su participación activa en el Programa al aprovechar los apoyos y cumplir sus compromisos es primordial para la superación de sus condiciones de vida.
8. Involucra a la sociedad en la superación de la pobreza. Al esfuerzo del Gobierno Federal es necesario sumar a la sociedad. La superación de la pobreza exige la incorporación de las instituciones y organismos públicos, sociales y privados.
9. Promueve la articulación con otros programas sociales, al buscar que las acciones de Oportunidades y otros programas sociales se desarrollen en conjunto, para mejorar el nivel de vida de las familias y las comunidades.
10. Se basa en la coordinación entre diferentes instituciones. La operación del Programa se realiza a través de Oportunidades en Internet
11. Realiza seguimiento permanente a su operación y evalúa sus impactos, al realizar un análisis de los beneficios de corto, mediano y largo plazo del Programa sobre la población que atiende, tanto en el ámbito personal, familiar y comunitario.

Para el seguimiento de la operación, cuenta con un Comité Técnico donde participan Subsecretarios de esas Dependencias, así como el Director General del IMSS y un Delegado de la Secretaría de la Función Pública. La coordinación en los estados se realiza a través de Comités Técnicos Estatales donde se involucran las responsables federales y estatales vinculados a la operación del Programa.

Oportunidades ha sido objeto de un diseño preciso, que focaliza a la población a la que va dirigido, considerando la condición de residencia rural-urbana de las

familias y sus condiciones socioeconómicas. Su aplicación evita la discrecionalidad en la identificación de las familias beneficiarias. Las Reglas de Operación del Programa establecen los criterios generales para la selección de la población a atender, reglamentando:

- Los criterios para atender la demanda de incorporación de las familias
- Esquema bajo el que opera el Programa
- La mecánica de operación
- La evaluación y el seguimiento operativo
- Mecanismos e instancias para presentar quejas y denuncias.

Identificación e incorporación de familias

Las familias beneficiarias se identifican mediante criterios y procedimientos que se establecen en las reglas de Operación. Estos procedimientos constan de dos etapas: 1) Selección de localidades y 2) Identificación de familias. La selección de localidades es el proceso mediante el cual se define qué localidades rurales y urbanas pueden ser atendidas por el Programa, para identificar a los hogares que reúnen las características para recibir los apoyos de Oportunidades. Se realiza con base en el índice de marginación establecido por el Consejo Nacional de Población (CONAPO, 2005), así como la información estadística disponible a nivel de localidades, AGEB's, colonia y/o manzanas, generada por el INEGI, dando prioridad a la selección y atención de aquellas donde la concentración de hogares en pobreza extrema es mayor.

Una vez seleccionado el universo de atención, conformado por localidades, AGEB's, colonias y/o manzanas, se procede a validar las condiciones de accesibilidad y capacidad de atención de los servicios de salud y educación es decir, las localidades seleccionadas deben contar con:

- Servicios de salud y educación en su misma localidad o en una cercana.

- La certeza de que el sector salud tenga la capacidad de atender a los posibles beneficiarios del Programa.

Las localidades definidas mediante este proceso son las consideradas para llevar a cabo la identificación de las familias a incorporar por Oportunidades.

Para identificar a las familias a incorporar al Programa se recolecta información socioeconómica de los hogares a través de una encuesta de características socioeconómicas de los hogares (ENCASEH). Se analiza la información recolectada en la encuesta, cuyo resultado indicará si el hogar, por sus condiciones de vida socioeconómicas, recibirá o no los beneficios del Programa. Este análisis se realiza a través de una metodología de puntajes y una técnica estadística llamada análisis discriminante, que permite reconocer a las familias que viven en condiciones de pobreza.

La titular beneficiaria es quien recibe los apoyos monetarios. En cada familia beneficiaria se identifica una titular, que generalmente es la madre de familia o la persona encargada del cuidado de los niños y de la preparación de los alimentos. Oportunidades tiene como prioridad fortalecer la posición de las mujeres en la familia y dentro de la comunidad. Por tal razón, son las madres de familia las titulares del Programa y quienes reciben las transferencias monetarias correspondientes.

La incorporación al Programa es el momento en que las familias que se identificaron como beneficiarias por medio de la encuesta de características socioeconómicas aceptan entrar a Oportunidades. En caso de que las familias decidan recibir los apoyos de Oportunidades, se les entrega su identificación provisional y su formato con el que deben acudir a la clínica que les fue asignada, para registrar a los miembros de la familia en la unidad de salud y se les programan sus citas médicas a cada uno de ellos. En caso de que esté por iniciar

el año escolar, también se le entregarán a la titular beneficiaria los formatos en los cuales reportará a Oportunidades, la inscripción de sus hijos que tengan entre 18 y 21 años cumplidos y estén estudiando o vayan a estudiar en escuelas de nivel medio superior.

Así pues, concretamente el esquema de operación del programa está conformado por cuatro componentes fundamentales para el desarrollo humano:

- Educativo, que incluye otorgar becas educativas y apoyos para la adquisición de útiles escolares o un paquete de éstos a los becarios. A partir del primero de secundaria, el monto de las becas escolares es mayor para las mujeres, dado que su índice de deserción aumenta a partir de ese grado.
- Salud, que opera bajo cuatro estrategias específicas: proporcionar gratuitamente el “Paquete Básico de Servicios de Salud”, promover una mejor nutrición, especialmente en los niños y mujeres embarazadas, fomentar el autocuidado mediante la comunicación educativa y reforzar la oferta de los servicios de salud. En el cuidado de la salud, las mujeres embarazadas reciben una atención especial, con un protocolo de consultas para cuidar el desarrollo del embarazo, la salud de la madre y prevenir partos con riesgo.
- Alimentario, que consiste en otorgar apoyos monetarios directos de manera bimestral a las familias beneficiarias y complementos alimenticios, así como acciones de educación alimentario-nutricional.
- Patrimonial o más conocido como Jóvenes con Oportunidades, consiste en un beneficio económico diferido que se acumula gradualmente en forma de puntos a partir del tercer grado de secundaria y hasta el último grado de educación media superior. Los puntos se convierten en pesos que se depositan en una cuenta de ahorros (a través de un factor de conversión de puntos a pesos) administrada por una institución financiera. Este

componente los vincula con programas para que continúen estudiando, inicien un negocio propio, mejoren su vivienda o se incorporen al Sistema de Protección Social en Salud o adquieran el Seguro de Salud para la Familia del IMSS.

Estos aspectos constituyen, para el Programa, la base para que las familias puedan mejorar sus condiciones de vida. Superar la situación de pobreza extrema que les impide valerse por sus propios medios requiere del esfuerzo conjunto de los gobiernos Federal, Estatal y Municipal, así como de la sociedad.

Mecánica de operación

Oportunidades requiere la corresponsabilidad de las familias atendidas, para lo cual asocia las transferencias a que las familias lleven a cabo acciones que tienden a mejorar sus condiciones de vida. El esquema de corresponsabilidad implica un conjunto de compromisos que los hogares deben cumplir a cambio de seguir beneficiándose de las intervenciones y transferencias del Programa. Las acciones de corresponsabilidad que se les piden a las familias se ubican en las áreas de educación y salud que son esenciales en el proceso de quebrar la reproducción intergeneracional de la pobreza.

Para cumplir con las corresponsabilidades del componente Salud, es necesario que las familias se registren en la unidad de salud que les corresponda. Además, todos los integrantes de la familia deberán asistir a sus citas programadas en los servicios de salud; la corresponsabilidad en educación consiste básicamente en inscribir a los niños en edad escolar a la escuela y que estos asistan regularmente.

Los montos de los apoyos del Programa siguen determinados criterios. Los montos de dichos apoyos y el monto máximo mensual que una familia puede recibir se actualizan semestralmente tomando en consideración el incremento

acumulado del Índice Nacional de Precios de la Canasta Básica a partir de julio de 1997 hasta el mes inmediato anterior al de la actualización, que publica el Banco de México en el Diario Oficial de la Federación.

MONTOS MENSUALES CORRESPONDIENTES AL SEGUNDO SEMESTRE DEL 2008

Apoyo alimentario	195
Adultos mayores	275
Apoyo energético	55
Vivir mejor	120

Componente Educativo

Primaria	
Tercero	130
Cuarto	155
Quinto	195
Sexto	165

Secundaria	Hombres	Mujeres
Primero	385	405
Segundo	405	450
Tercero	430	495

EMS	Hombres	Mujeres
Primero	645	740
Segundo	695	790
Tercero	735	840

Tope apoyos educativos

Becas de educación básica	1010
Becas de educación media superior	840

Apoyos para útiles

Primaria	
Primera entrega	175
Segunda entrega	90

Secundaria	330
EMS	330

Monto máximo mensual

Familia con becarios en primaria y secundaria	1380
Familias con becarios en primaria, secundaria y EMS	2220

Fuente: http://www.oportunidades.gob.mx/informacion_general

Recertificación de familias y Esquema Diferenciado de Apoyos (EDA)

Dado que el objetivo de Oportunidades es el desarrollo de capacidades, a través de la educación, el cuidado a la salud y una mejor alimentación, la temporalidad es muy importante. Por ello a tres años de que las familias se incorporan al Programa se realiza una nueva evaluación de las condiciones económicas, sociales y

demográficas de las familias a través de la aplicación de una encuesta llamada de recertificación (ENCRESEH).

Al proceso anterior también se le denomina Recertificación y constituye un paso importante para garantizar continuidad de apoyos a los hogares que aún lo requieren. Por tanto la Recertificación es un mecanismo que permite que Oportunidades sea un programa de apoyo temporal, evitando generar la permanencia indefinida y la dependencia en las familias. Con base en los resultados de la encuesta de Recertificación se detecta:

- Familias que presentan un mínimo de bienestar y mejoras sostenibles en su condición y que, por tanto, pueden transitar al Esquema Diferenciado de Apoyos (EDA).
- Familias que de acuerdo a sus condiciones económicas y sociales no aseguran la continuidad de sus hijos e hijas en la escuela y por tanto requieren mantener los mismos apoyos del Programa (esquema tradicional).
- Familias que se consideran errores de inclusión para darlas de baja.

El EDA consiste en:

- Becas educativas y apoyo para útiles escolares para nivel secundaria y Educación Media Superior.
- Paquete Esencial (Básico) de Servicios de Salud.
- Suplementos alimenticios para niños y mujeres embarazadas y en lactancia.
- Las familias continúan cumpliendo su corresponsabilidad.

1.4 Proceso de evaluación

El *Programa Nacional de Desarrollo Social 2001-2006* señala la necesidad de construir una “cultura de la evaluación”, con el objeto de lograr un mejor desempeño de las políticas públicas, buscar el uso eficiente de los recursos y el cumplimiento de las metas y compromisos establecidos, especialmente tratándose de acciones dirigidas a la población más pobre del país: “la evaluación de la gestión, de los resultados e impactos de las políticas públicas debe tener una concepción integral que permita valorar correctamente lo que se está haciendo, cómo se está haciendo, para quién se está haciendo, qué resultados se están obteniendo y cuál es el impacto real en las condiciones de vida de la población beneficiaria”.

Oportunidades se ha caracterizado por llevar a cabo evaluaciones de impacto para el manejo y rediseño del programa. Algunas de esas evaluaciones destacan por su fortaleza metodológica, al haber utilizado procedimientos de asignación al azar, control estadístico y doble diferencia para asegurar la validez interna de las estimaciones. El seguimiento se orienta al análisis de los componentes de la gestión interna, mientras que la evaluación atiende la relación entre los productos y los objetivos de impacto perseguidos. En el seguimiento el análisis se centra, entre otras, en la eficacia y en la eficiencia de la focalización.

Evaluación de impacto

Desde el inicio de operaciones del Programa, se ha considerado un componente de evaluación con la finalidad de identificar y cuantificar la magnitud de los impactos del Programa a través de metodologías rigurosas con enfoques de corte cuantitativo y cualitativo. De acuerdo con las Reglas de Operación del Programa, la evaluación de impacto ha sido encomendada a instituciones académicas y de investigación con prestigio nacional e internacional, como el Instituto Nacional de Salud Pública (INSP), el Centro de investigaciones y Estudios Superiores en antropología Social (CIESAS), el Instituto Internacional de Investigaciones sobre

Políticas Alimentarias (IFPRI, por sus siglas en inglés) y el Centro de Investigación y Docencia Económicas A.C. (CIDE).

Esta evaluación consiste en un proceso continuo cuyas actividades están dirigidas al análisis de los resultados e impactos del Programa. Asimismo, constituye la base de un sistema integral de valoración que permite instrumentar ajustes en el diseño y ejecución del Programa y contribuye al cumplimiento de sus objetivos y metas.

A lo largo de los años, los esfuerzos de la evaluación se han centrado en la relación entre las estrategias y acciones del Programa y los efectos que producen en la población beneficiaria, tanto en las áreas rurales como en las urbanas. La evaluación de impacto de Oportunidades ha sido reconocida a nivel internacional por su rigor académico, confiabilidad y constancia, por lo que se ha tomado como un modelo a replicar. Los trabajos cuantitativos y cualitativos han generado un cúmulo de información que incluye bases de datos, cuestionarios, notas técnicas y metodológicas, estudios de hogares y documentos de evaluación. Los principales resultados obtenidos de la evaluación de impacto del Programa son los presentados en el cuadro 1.

El seguimiento operativo

Para apoyar la detección y resolución oportuna de los problemas operativos del Programa, se cuenta con un Modelo de Seguimiento Operativo convenido y aplicado conjuntamente con los Sectores de Salud y Educación. Este modelo permite obtener información detallada sobre el estado de la operación, con el propósito de la toma de decisiones, la detección oportuna de desviaciones operativas e implementar, en su caso, acciones de mejora continua en los procesos operativos, a través de la generación y análisis de indicadores.

Cuadro 1. Principales resultados de Oportunidades

EDUCACIÓN	Oportunidades incrementa la inscripción para las transiciones de primaria a secundaria, y de secundaria a media superior, particularmente para las mujeres. También reduce la reprobación y la deserción escolar, aumenta el logro educativo y genera una mayor disposición de los padres a promover la continuidad educativa de sus hijos y el cumplimiento de sus obligaciones escolares
SALUD	El Programa reduce las tasas de mortalidad materno-infantil, disminuye la morbilidad e incapacidad, aumenta la utilización de servicios públicos ambulatorios de salud y reduce la de los servicios privados. Esto último representa un menor gasto de bolsillo por este concepto para los beneficiarios
ALIMENTACIÓN	Oportunidades reduce la prevalencia de anemia y permite incrementar la estatura y peso de los niños en edades tempranas. El Programa mejora la dieta de los hogares al permitir que se compren alimentos de origen animal y al proporcionar complemento alimenticio a una gran parte de la población infantil
ASPECTOS DEMOGRÁFICOS, ECONÓMICOS, SOCIALES Y DE GÉNERO	El Programa no incrementa significativamente la migración, no incrementa la violencia hacia las mujeres, ni parece deteriorar los vínculos comunitarios. Los beneficiarios dedican más recursos a necesidades esenciales, vivienda e inversión productiva
METAS, COSTOS UNITARIOS Y REGLAS DE OPERACIÓN	Oportunidades ha cumplido la meta de familias atendidas al dedicar sólo 2.5% de su presupuesto total a gasto administrativo. El 93.6% de los beneficiarios considera bueno el trato recibido por el personal del Programa.

Fuente: Cruz, C. De la Torre R, Velázquez C. Informe compilatorio. Evaluación externa de impacto Del Programa Oportunidades 2001-2006. Cuernavaca, México: Instituto Nacional de Salud Pública, 2006, p. 12

El Modelo de Seguimiento Operativo, que es adicional a los propios mecanismos internos de supervisión de cada sector, se integra al sistema de gestión operativa del Programa que ejecuta los procesos en ciclos bimestrales, integrando información proveniente de los resultados operativos de cada entidad federativa. Oportunidades cuenta con un sistema de monitoreo y seguimiento operativo que procesa información sobre áreas temáticas fundamentales para la operación del Programa (Cuadro 2) y para ello integra distintos instrumentos que permiten caracterizar su estado de funcionamiento e incluye los informes nacionales y los reportes de seguimiento operativo elaborados por los estados.

El sistema de monitoreo se caracteriza por el empleo combinado de múltiples fuentes y métodos de recolección de datos. Entre ellas destacan las encuestas de Puntos Centinela (tema de esta investigación), el padrón de beneficiarios, los indicadores de atención ciudadana, los indicadores bimestrales de seguimiento, evaluación, gestión y resultados, los de resultados de supervisión operativa (encuestas y cédulas de observación), los indicadores de desempeño de las coordinaciones estatales y el Sistema de Información Georeferenciada.

La operación y alcance del sistema de monitoreo es considerable en términos de magnitud. Según datos del levantamiento de información de enero a junio de 2004 (metodología vigente), se efectuaron más de 16000 encuestas a titulares en 3515 localidades a lo largo de todo el país; se realizaron 3180 encuestas a profesores de escuelas de educación básica; se llevaron a cabo 1475 encuestas a los encargados de la certificación electrónica de planteles de educación media superior, y se aplicaron 4480 cédulas a unidades de salud. Asimismo, se procedió al diseño y la estimación de los indicadores sobre la base de las muestras obtenidas en los Puntos Centinela y de la información censal integrada a las bases de datos del Programa, estableciéndose estándares para cada indicador. Los productos del sistema consistieron en 32 informes estatales y uno federal para cada bimestre.

Cuadro 2. Temas cubiertos por el sistema de monitoreo

Tema	Descripción
Permanencia de los beneficiarios en el padrón	Bajas de familias y becarios
Componente salud	Calidad del servicio del sector salud y estado nutricional de los niños menores de cinco años
Componente Educación	Servicio proporcionado por el sector educativo
Entrega de apoyos	Eficiencia y eficacia en la entrega de apoyos
Transparencia y contraloría social	Prestación de servicios de manera honesta, sin solicitar a cambio bienes, servicios o adhesión política
Orientación a beneficiarios	Avance y resultados de las actividades de la orientación a los beneficiarios
Capacitación a vocales, enlaces y personal de los sectores	Avance y resultados de las actividades de capacitación
Atención ciudadana	Los tipos de quejas y denuncias más frecuentes planteadas por la ciudadanía

Fuente: Cohen, Ernesto y Rolando Franco, Transferencias con Corresponsabilidad, una Mirada Latinoamericana, FLACSO, México, 2006

Los distintos sistemas de información institucionales son la fuente para la conformación de indicadores con temas relacionados con la entrega de servicios y productos, así como la percepción de la calidad de la atención. Entre los aspectos críticos que son objeto de seguimiento periódico destacan el tiempo y los costos de traslado a los módulos de entrega de apoyos; la entrega del complemento alimenticio; la orientación a las titulares sobre sus derechos y obligaciones; la recuperación oportuna de la certificación del cumplimiento de las corresponsabilidades y la capacitación a las vocales y al personal de salud y educación. El modelo de seguimiento operativo utiliza sus fuentes de información para:

- Proponer alternativas de solución
- Dar seguimiento a la implementación de acciones
- Medir la mejora resultante de la implementación
- Identificar situaciones críticas en la operación

El Modelo de Seguimiento Operativo está delimitado por la Evaluación de Impacto y el seguimiento diario a la ejecución de los procesos y cumplimiento de metas. Su principal fuente de información son las cédulas (encuestas) que se aplican en el operativo Puntos Centinela, con lo que se genera información oportuna sobre el estado de la operación del Programa en las Unidades de Servicio.

Puntos Centinela

Puntos Centinela es uno de los instrumentos del Modelo de Seguimiento Operativo del programa, de medición cuantitativa que se realiza dos veces al año en cada Entidad Federativa (marzo-abril y septiembre-octubre). A través de este instrumento se hace una medición interna del Programa con un enfoque para la mejora continua a través de un esfuerzo concertado interinstitucional. Su objetivo es producir información sobre la operación del Programa en las unidades de servicio (mesas de atención y servicio, escuelas y unidades de salud). Las unidades de análisis del sistema están constituidas por las titulares, las escuelas y los servicios de salud.

Este operativo está normado en las Reglas de Operación y el contenido de las encuestas se desprende principalmente de lo establecido en las mismas reglas de Operación y ha sido acordado con los representantes de los sectores Educación y Salud a nivel central. Se aplica a muestras probabilísticas, estadísticamente representativas de titulares, escuelas y unidades de médicas que atienden a becarios y familias del programa.

Los puntos centinela cuentan con cinco instrumentos de medición:

- a) Encuesta a Titulares. Tiene por objetivo verificar la calidad del servicio y la atención que reciben las beneficiarias del Programa Oportunidades; se le aplica a titulares beneficiarias del Programa en los módulos de entrega de apoyos y lo aplica el personal de la Coordinación Estatal de Oportunidades.

- b) Encuesta de Educación Básica
- c) Encuesta de Educación Media Superior. El objetivo de estas dos encuestas es conocer si el personal docente recibe y maneja información oportuna y suficiente sobre Oportunidades para hacer una correcta certificación y brindar un servicio adecuado a los becarios. Se aplica a profesores, directores y responsables de la certificación, este instrumento se aplica en cada escuela seleccionada en muestra y también es administrado por personal de la Coordinación Estatal de Oportunidades.
- d) Encuesta a Becarios de Educación Media Superior. Su objetivo es conocer el nivel de información que manejan los becarios que cursan el 5º semestre o más de Educación Media Superior, sobre el componente patrimonial Jóvenes con Oportunidades y está dirigido a becarios que cursan estos semestres.
- e) Cédula de supervisión de unidades médicas. El objetivo de este instrumento es verificar que las unidades médicas cuenten con los recursos suficientes para garantizar el acceso y la capacidad de atención de las familias beneficiarias a los servicios de salud y apoyos para mejorar su nutrición. Se aplica a los responsables de las unidades médicas en la unidad médica participante y lo aplica el personal del sector salud estatal.

El operativo Puntos Centinela pretende arrojar resultados representativos por Entidad Federativa y a nivel nacional. Los aspectos que se evalúan a través de las percepciones y opiniones de las titulares son: el funcionamiento de los componentes del Programa, la entrega de las transferencias, las actividades de capacitación y orientación, las acciones de contraloría social y la calidad de los servicios. En cuanto a las escuelas, se recogen datos sobre las condiciones físicas del plantel, la asistencia de los profesores, las prácticas de certificación de la corresponsabilidad y los medios por los cuales el personal docente obtiene información sobre Oportunidades. En las unidades de salud se obtiene información respecto a la suficiencia de los recursos humanos y materiales, las

actividades de orientación a los beneficiarios y la verificación de las corresponsabilidades. Sobre esta base, se descubren las fortalezas y aspectos que deben mejorarse para los sectores de educación y salud y para la operación del Programa. Los resultados recogen las percepciones de los servicios sectoriales y de la acción del Programa que se da entre las beneficiarias y los funcionarios de las escuelas.

Análisis de bajas del padrón de beneficiarios

Mediante este procedimiento se encuentran los errores operativos que provocan la mayor incidencia de bajas entre las familias beneficiarias. Proporciona información de base para desarrollar estrategias que permitan subsanar los errores operativos que inciden en las bajas innecesarias del padrón; en rigor, no obstante que las bajas se explican en parte por el incumplimiento de las corresponsabilidades, los errores de inclusión y las duplicaciones, se han descubierto otros factores, que pueden ser vinculados a errores de operación.

El Programa requiere desarrollar trabajo de campo para corroborar las situaciones que aparecen reflejadas en los indicadores, y de comprobarse la existencia de bajas, por errores operativos, donde se procura reintegrar a las familias perdidas. Es plausible que el descubrimiento de las bajas no justificadas requiera de procedimientos menos estandarizados, basados en el monitoreo en la comunidad, que deberían incluir también la aclaración a las familias de los motivos de la baja.

2. Técnicas de Muestreo

2.1 El por qué del muestreo

Para recabar información perteneciente a un grupo más o menos numeroso de personas o cosas, una forma de proceder puede ser estudiar a todos y cada uno de los componentes del conjunto. Sin embargo, resulta obvio que cuando las poblaciones que se quieren estudiar tienen un tamaño grande, obtener información de cada uno de los miembros del grupo resulta inviable, tanto por los costos que conlleva un estudio de tales características como por el tiempo que requiere recabar una cantidad importante de información. Por ello, se recaba información de un subgrupo representativo del colectivo que sea el objeto de estudio. Es decir, extraemos la información de una parte de este grupo, lo que se denomina muestra, y tales resultados los extrapolamos al resto de la población.

La clave de un procedimiento de muestreo es garantizar que la muestra sea “buena”. Por tanto, cualquier información respecto a las diferencias entre sus elementos debe tenerse en cuenta para seleccionar la muestra. Las técnicas de muestreo se utilizan cuando se quiere conocer cuáles son las características generales de una población y no se tiene la capacidad, el tiempo o el presupuesto suficiente para estudiar a toda la población y se estudia sólo una parte de ésta.

En todos los estudios que usan muestras, se requiere el enfoque científico, éste consiste, básicamente en usar los conocimientos previos que se tienen sobre el problema y diseñar una metodología de investigación que minimice la ocurrencia y magnitud de los errores. Para el estudio se debe determinar, primero, el proceso de captación de la información asociada a cada elemento, esto es, la forma de aplicación y el tipo de instrumentos de medición a utilizar. Entre los instrumentos se encuentra: la observación directa, los cuestionarios, las entrevistas directas o telefónicas, entre otros. Conocer y aplicar correctamente los instrumentos de investigación permitirá evitar errores en el proceso de captación de información.

Todo **muestreo** requiere de un **diseño** que contenga lo que se va hacer y cómo se va a hacer. En ese sentido se necesitan especificar los procedimientos de selección de la muestra y los de estimación.

Antes de hacerlo es indispensable especificar la población o “universo” de la que se extraerá la muestra. Dicha población quizá no sea la misma a la que inicialmente se refiere la investigación por realizar puesto que requiere adecuaciones. Un estudio que proponga investigar las tendencias de opinión pública con respecto al clima electoral (preferencias por candidatos o partidos, participación esperada, etc.) y que si pudiere, se referiría a todos los votantes registrados (población en estudio) puede modificarse, si depende de un directorio telefónico abierto, a todos los mayores de 18 años que habitan en una vivienda que tiene registrado un número telefónico no privado (población objetivo).

En el caso anterior –como muchos otros- la población en estudio es reemplazada por la población objetivo. Las inferencias plausibles de la muestra se refieren a la segunda y no a la primera. En la historia de los estudios mediante encuesta hay un gran número de errores debidos a la falta de reconocimiento de la idea anterior. De hecho tanto la investigación de mercados como la de la opinión pública producen diariamente casos en esa dirección.

Marco muestral

La selección de una muestra de una población objetivo diferirá en diversos aspectos relevantes si se lleva a cabo utilizando un marco muestral o si no se hace así. En el primer caso (sea un marco lista –o directorio- o un marco área) es posible asignar probabilidades específicas a todos y cada uno de los miembros de la población objetivo que sean seleccionados. Por lo tanto, es posible aplicar un diseño muestral probabilístico.

En el segundo caso, al seleccionar una muestra sin marco muestral, es imposible cumplir con los requisitos que un diseño muestral exige. Esto es relevante en tanto

se busquen aplicar las inferencias estadísticas de las muestras probabilísticas a una muestra no probabilística. Se podrán hacer muchas cosas e incluso recurrir a inferencias no estadísticas, pero se tiene certeza de que serán válidas las inferencias estadísticas que se derivan de un diseño muestral probabilístico.

El marco muestral es, por lo tanto, un componente relevante de la aplicación del muestreo, particularmente de la teoría del muestreo en poblaciones finitas. Por ello es importante constituir, mantener y actualizar marcos muestrales y con permanente empeño. Dada una caracterización de la población en estudio, y de las condiciones y características de un marco muestral disponible, se pueden presentar las siguientes situaciones:

- Existen elementos de la población en estudio que no se localizan en el marco muestral (omisiones). Es decir, el marco muestral empleado no cubre totalmente a la población en estudio. Existe, por lo tanto, un problema de subcobertura.
- Existen elementos del marco muestral que no forman parte de la población en estudio. Es decir, hay elementos ilegibles en el marco muestral. En consecuencia, existe un problema de sobrecobertura.

Los dos problemas citados son relevantes: los tamaños de la subcobertura (el número de omisiones) y de la sobrecobertura (inelegibles) son sólo parte de la materia a considerar. Es necesario por lo menos reflexionar sobre las diferencias posibles entre la subcobertura y el marco muestral para identificar la mejor opción entre las siguientes:

- a) Aceptar el marco muestral (depurándolo y actualizándolo) como la población objetivo redefiniendo así los alcances de la inferencia del estudio.
- b) Completar el marco muestral para hacerlo más idóneo a la Investigación. Otros marcos muestrales disponibles pueden servir para cumplir al menos parcialmente con el objetivo propuesto.

- c) Rechazar el marco muestral y construir un marco muestral *ex profeso* para la investigación o, en su defecto, llevar a cabo un diseño muestral no probabilístico sin marco muestral.
- d) No llevar a cabo la investigación mediante muestreo.

Todo marco muestral puede contener repeticiones o duplicaciones. Errores de captura, procesamiento y validación pueden ser los culpables. Otra fuente de error corresponde al ensamble de distintos directorios o bases de datos en los que los registros de identificación no son enteramente coincidentes. Este problema, salvo en casos extraordinarios, es afortunadamente el menor en relación a las problemáticas apuntadas.

Muestreo no probabilístico

En el método de muestreo no probabilístico no se puede asignar objetivamente probabilidades a las unidades, y en consecuencia, no se puede determinar la confiabilidad de los resultados de la muestra en términos de probabilidad. Por lo general existen cuatro grandes tipos de diseños muestrales no probabilísticos:

- 1) Muestreo fortuito o casual. Ocurre cuando se define la selección de un participante en la muestra por una condición externa, visible e identificable. Por ejemplo, incluir en muestra a todas las mujeres mayores de 25 años que pasen por el cruce o intersección de dos calles X y Y entre las 12:00 y las 12:15 hrs.; entrevistar a todas las personas de sexo masculino que usen traje, corbata y portafolios y que se identifiquen circulando en la esquina de López y González, etcétera.
- 2) Muestreo de “voluntarios”. Se realiza cuando los participantes en muestra se autoseleccionan. Esto es frecuente en la investigación médica cuando, si se satisfacen condiciones de elegibilidad, algunas personas intercambian cuidados médicos (o incentivos monetarios) por someterse a tratamiento o a drogas específicas. También ocurre en la investigación de Mercados

cuando se colocan buzones para registrar sugerencias y comentarios: los que contestan son “voluntarios” de la muestra y, por diversas razones, son distintos de los que no contestan.

- 3) Muestreo “a juicio” (“a juicio del entrevistador”) ocurre cuando el entrevistador, siguiendo el protocolo de levantamiento de campo, selecciona a los entrevistados. Esto sucede en vía pública como incluso en entrevistas en vivienda.
- 4) Muestreo por cuotas. Este tipo de muestreo preserva el juicio del entrevistador pero la condiciona a “cuotas” (o tamaños de muestra) generalmente fijados por sexo, grupo de edad y, en ciertas investigaciones de mercado, por nivel socio económico apreciado. Comúnmente las “cuotas” buscan replicar las condiciones demográficas pero esto, aunque importante, no es garantía alguna de que la muestra sea válida.

En ocasiones se mezclan dos o más tipos de muestreo. Por ejemplo, se conjuga el carácter de “voluntario” en un muestreo por cuotas puesto que el entrevistador y el entrevistado mutuamente se seleccionan. En ninguno de los cuatro tipos de muestreo anteriores se conoce *ex ante* la probabilidad de selección de cada elemento de la población objetivo. Por si esto fuera poco siempre habrá elementos de la población objetivo cuya probabilidad de inclusión en muestra es cero. Es el caso de las encuestas por muestreo en puntos de afluencia que nunca pasan por los sitios designados, los que en encuestas de vivienda no aceptan entrevistas personalizadas o por teléfono, etcétera.

Muestreo probabilístico

El muestreo probabilístico (un termino debido a Deming (1950)) es aquel donde cada unidad se extrae con una probabilidad conocida y distinta de cero. Este método se ha desarrollado desde 1940 y, para todos los propósitos prácticos, ha reemplazado al muestreo no probabilístico. Si se tienen probabilidades de selección iguales se llama muestra autoponderada, y si es de tamaño “grande”, la

muestra resultará, con elevada probabilidad, representativa. Si no se toma con probabilidades iguales, se hacen ajustes en la forma de estimar promedios o proporciones para recobrar la representatividad.

2.2 Conceptos básicos del muestreo

Como ya se ha apuntado, una muestra busca replicar en forma condensada a la población en estudio y a las características, atributos o variables que son relevantes para la investigación. En la medida en que la replicación o “copia” de la población en muestra cumpla con los requisitos de ser satisfactoria se garantiza que el proceso de estimación puede ser válido y consistente.

Los términos “atributo” y “variable” son ampliamente usados en la literatura. En sentido estricto son diferentes. Si se indaga si un individuo en particular posee o no una cierta característica, dicha característica puede referirse como atributo. La *cuantificación* reside en contar cuántos individuos poseen dicho atributo y cuántos no. La proporción (o porcentaje) que tienen el atributo proporciona una descripción útil de la población. En contraste, se puede estar interesado en la magnitud actual de alguna variable característica de cada elemento de la muestra –su edad, ingreso, consumo, etcétera. Entonces la cuantificación implica medir la magnitud de la característica en cada caso. Un “resumen” elemental y útil para describir la población en términos de una variable es, por ejemplo, un promedio o media aritmética.

Por cierto, que la distinción apuntada entre “atributos” y “variables” no es tan fundamental como algunas veces se dice. Una variable puede siempre transformarse en un atributo mediante un agrupamiento: la variable “edad” puede tomar en un estudio valores de 18 años o más para convertirse en un atributo al agrupar, por ejemplo, las edades de 18 a 25 años, de 26 a 35 años.

Recíprocamente cualquier atributo puede transformarse en una variable al asignar el valor “1” a quienes poseen el atributo y el valor “0” a quienes no lo tiene. Por lo

anterior, la distinción entre la proporción de la población no es básica ni en consecuencia, la lógica del muestreo difiere en los dos casos.

Otra pareja de conceptos que es conveniente precisar es la de “**estadística**” y de “**parámetro**”. “Estadística” se refiere, en este contexto, a un valor resumen o condensado de una variable (o atributo) calculado de una muestra. “Parámetro” se relaciona al valor resumen de una variable (o atributo) en la población. Al parámetro, en la teoría del Muestreo en Poblaciones Finitas, comúnmente se le designa por “valor poblacional”.

Un estimador es también una función de las observaciones de la muestra que busca asemejarse al parámetro (o valor poblacional) para “representarlo”. En el caso del muestreo probabilístico las observaciones de la muestra son variables aleatorias y, en consecuencia, los resultados relevantes al caso de la Teoría de Probabilidad pueden aplicarse.

Hay distintos enfoques o teorías del muestreo: (1) el muestreo con base en modelos, propio de los textos en estadística matemática en los que, por lo general, se dispone de información acerca de la población y ésta recibe el nombre de distribución; (2) el muestreo de “superpoblaciones” en los que se presupone que toda la población es, a su vez, una muestra; concepto que tiene un uso amplio en procesos biológicos; y (3) *el muestreo de poblaciones finitas* que es el típico de las poblaciones sociales.

El modelo de poblaciones finitas por lo general carece de modelos específicos de la población en estudio y ésta contiene un número finito de elementos. Este último aspecto no debe ignorarse puesto que existen casos en los que las diferencias son drásticas.

Los procedimientos muestrales pueden resumirse en los siguientes pasos:

1. Se identifica(n), el(los) valor(es) poblacional(es) que se estimarán.

2. Se identifican, de acuerdo con el diseño muestral por aplicar, el estimador más apropiado del valor poblacional. Este se calcula con las observaciones de la muestra.
3. Se define la varianza del estimador. Esta, generalmente desconocida y requiere, a su vez estimarse. Por lo tanto se propone y calcula el estimador de la varianza del estimador.
4. Con base en el estimador de la varianza y con el valor estimado más apropiado del valor poblacional se obtienen los intervalos de confianza de la estimación.

En una muestra ideal la población muestreada será idéntica a la población objetivo, pero este ideal se cumple muy rara vez. Una buena muestra estará a salvo de presentar un sesgo de selección; esto ocurre cuando alguna parte de la población objetivo no está en la población muestreada.

Si la población es un conjunto de N unidades, $P = \{U_1, U_2, \dots, U_N\}$ y a cada unidad se le asocia una variable $Y(U_i) = Y_i$ que sea de interés. Se pueden considerar como vectores $(Y_1, Y_2, Y_3, \dots, Y_N)$, entonces un parámetro es, por ejemplo, $\mu = \bar{Y} = \sum_{i=1}^N \frac{Y_i}{N}$ la media de la población (en muestreo se usa \bar{Y} para la media de la población finita y \bar{y} para la media de la muestra).

La varianza de la población es:

$$\sigma_y^2 = \sum_{i=1}^N \frac{(Y_i - \bar{Y})^2}{N}, \text{ aunque se usa también,}$$

$$S_y^2 = \frac{N}{N-1} \sigma_y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}, \text{ en ocasiones el total de la población es de interés y}$$

$$\text{se estima con } Y = \sum_{i=1}^N Y_i = N\bar{Y}$$

La muestra es un subconjunto de n unidades de la población, éste se obtiene con probabilidades conocidas para todos y cada uno de los elementos de la población. El tamaño de la muestra es n , En la muestra de n unidades se determinan los n valores de la variable Y que se denotan por y_1, y_2, \dots, y_n (y_i minúscula es el valor observado, Y_j es un valor no observado en la población, $i=1, \dots, n, j=1, \dots, N$)

Con los valores y_1, y_2, \dots, y_n de la muestra se construyen estimadores de los parámetros. Para evaluar los estimadores se considera que el proceso de obtención de la muestra se puede repetir muchas veces con los mismos tamaños de muestra y probabilidades de selección, y en cada repetición tener un valor para el estimador. Como las muestras varían en su constitución, los valores de los estimadores también variarán, es decir, son variables aleatorias. De hecho, siempre que se realice un muestreo se debe especificar la forma de tomar la muestra y la manera de calcular los estimadores $f(y_1, y_2, \dots, y_n) = \hat{\theta}$, a estos dos aspectos les llamamos estrategia de muestreo.

Al tener muchos valores de un estimador, se puede estudiar su regularidad estadística con su distribución de frecuencias. Estamos hablando de una nueva población que son las diferentes estimaciones del parámetro de interés. Estas distribuciones de frecuencia se estudian teóricamente. Si la media de los posibles valores de un estimador es el parámetro en estudio, se dice que el estimador es insesgado con esta forma de tomar la muestra $E(\hat{\theta}) = \theta$. En realidad es la estrategia la que produce estimadores insesgados. Para un tamaño de muestra fijo, el mejor estimador es el que tienen menor variabilidad alrededor de su media (o sea el parámetro, si el estimador es insesgado). La medida de variabilidad es la varianza del estimador.

$$V(\hat{\theta}) = E[\hat{\theta} - E(\theta)]^2$$

Si el estimador no es insesgado es conveniente tomar como medida de variabilidad el error cuadrático media (ECM)

$$ECM(\hat{\theta}) = E[\hat{\theta} - \theta]^2 \Rightarrow ECM(\hat{\theta}) = V(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2$$

donde el sesgo $B(\hat{\theta})$ se obtiene a través de esta diferencia $E(\hat{\theta}) - \theta$

Una vez seleccionado un buen estimador, el tamaño de muestra n se determina con base en el grado de precisión deseada. Esta precisión es el grado de variabilidad que presenta el estimador alrededor del parámetro al tomar muchas muestras de tamaño n . Así, la precisión se expresa por un número δ tal que:

$$P(|\hat{\theta} - \theta| < \delta) = 1 - \alpha$$

A δ se le conoce como error absoluto.

$$P[\hat{\theta} - \delta \leq \theta \leq \hat{\theta} + \delta] = 1 - \alpha$$

Donde $1 - \alpha$ será cercano a 1 y $(1 - \alpha) * 100$ se denomina por ciento de confianza.

Frecuentemente se recurre al Teorema del límite central que especifica que

$$\bar{y} \sim N\left(\bar{Y}, \frac{\sigma_y^2}{n}\right). \text{ Sin embargo, en algunos casos el Teorema central del límite no}$$

es válido, entonces la determinación del tamaño de muestra n se basa únicamente en el criterio de lograr que la varianza del estimador sea pequeña, menor o igual que un valor dado, asegurando así el grado de variabilidad que presenta el estimador alrededor de su media (el parámetro si el estimador es insesgado). Una presentación alternativa del error de estimación es:

$$P\left|\frac{|\hat{\theta} - \theta|}{\theta} < \delta^*\right| = 1 - \alpha \text{ donde } \delta^* = \frac{\delta}{\theta}$$

Aunque no conocemos θ , al emplear $\delta^* = \frac{\delta}{\theta}$ estamos pensando en un error máximo permisible que esté en función de la cantidad que vamos a estimar. A δ^* se le conoce como error relativo. También el error de muestreo se usa con porcentajes, $\frac{\delta}{\theta} 100$.

2.3 Muestreo simple aleatorio (msa)

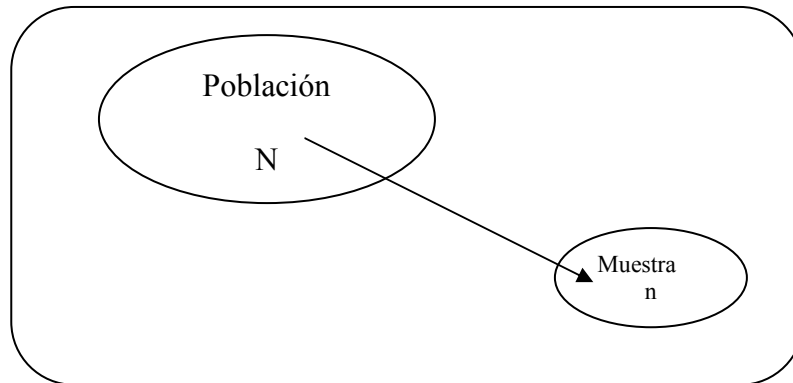
Es la forma más fácil de realizar un muestreo probabilístico y proporciona la base teórica de las demás formas complejas. Se obtiene una muestra aleatoria simple de tamaño n cuando cualquier subconjunto posible de n unidades en la población tiene la misma probabilidad de ser seleccionada para componer la muestra.

Existen dos formas de extraer una muestra aleatoria simple: con reemplazo, donde la misma unidad se puede incluir más de una vez en la muestra, y sin reemplazo, donde todas las unidades de la muestra son distintas. Nos enfocaremos en este esquema de muestreo ya que siempre se prefiere un muestreo que no tenga duplicados y es el más usado cuando se tiene un marco de muestreo que especifique directamente la manera de identificar cada unidad en la población. Además no se tiene conocimiento *a priori* sobre los posibles valores de Y_i ni otras mediciones asociadas a Y_i . En este caso cada unidad se extrae con igual probabilidad, por etapas y sin reemplazo, hasta tener las n unidades de la muestra. En la primera extracción, la probabilidad de que se seleccione específicamente una de las unidades de la población es $\frac{1}{N}$. En la segunda extracción la probabilidad de que se seleccione una de las $N-1$ unidades es: $\frac{1}{N-1}$ y así de manera sucesiva; en la selección k , la probabilidad de una de las $n-k+1$ restantes es $\frac{1}{N-k+1}$.

Entonces existen $\binom{N}{n}$ muestras posibles y cada una es igualmente probable, de modo que la probabilidad de elegir cualquier muestra individual S de n unidades es: $P(S) = \frac{1}{\binom{N}{n}} = \frac{n!(N-n)!}{N!}$.

Entonces, la probabilidad de que cualquier unidad dada aparezca en la muestra es de $\frac{n}{N}$.

La siguiente figura ilustra este tipo de muestreo



Para extraer una muestra aleatoria simple se necesita un marco de muestreo. En este tipo de muestra, la unidad de muestreo y la unidad de observación coinciden. Cada unidad tiene asignado un número, y se elige una muestra de modo que: (1) cada unidad tenga la misma posibilidad de aparecer en la muestra y (2) la selección de una unidad no tenga influencia de las demás unidades ya elegidas.

Para estimar la media de la población se utiliza como estimador la media de la muestra, es decir:

$$\bar{y} = \bar{Y} = \sum_{i=1}^n \frac{y_i}{n}$$

Este es un estimador insesgado¹ y su varianza es:

$$V(\bar{y}) = E(\bar{y} - \bar{Y})^2 = \frac{S_y^2}{n} \left(1 - \frac{n}{N}\right),$$

¹ $E(\bar{y}) = \bar{Y}$, el promedio de los posibles valores \bar{y} al tomar muchas muestras es \bar{Y}

donde $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$.

Si N es infinito, $V(\bar{y}) = \frac{S_y^2}{n}$, es el resultado que se obtiene para poblaciones infinitas. En este caso $\frac{n}{N}$ es la fracción de muestra o proporción de la población que se muestra, y $1 - \frac{n}{N}$ es el factor de corrección por finitud (*fcf*).

Para poblaciones pequeñas, la fracción de muestreo $\frac{n}{N}$ es mayor, la información que tenemos de la población también es mayor y, por lo tanto la varianza es menor. Para un censo, la corrección para poblaciones finitas, al igual que $V(\bar{y})$, se anulan. Cuando la fracción de muestreo $\frac{n}{N}$ es grande en una muestra aleatoria simple sin reemplazo, la muestra se parece mucho a un censo, el cual no tiene variabilidad de muestreo.

Para la mayor parte de las muestras extraídas de poblaciones que poseen tamaños muy grandes, la corrección es casi 1. Para las poblaciones grandes, el tamaño de la muestra extraída es el que determina la precisión del estimador (y no el porcentaje de población muestreada).

La varianza de la población S_y^2 , que depende de los valores para toda la población, es desconocida. La estimamos mediante la varianza de la muestra:

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

Un estimador insesgado de la varianza de \bar{y} es:

$$\hat{V}(\bar{y}) = \frac{s_y^2}{n} \left(1 - \frac{n}{N}\right)$$

Por lo general, no reportamos la varianza estimada de \bar{y} , sino su raíz cuadrada, el **error estándar (EE)**:

$$EE(\bar{y}) = \sqrt{\frac{s^2}{n} \left(1 - \frac{n}{N}\right)}.$$

El coeficiente de variación de una estimación proporciona una medida de la variabilidad relativa de una estimación. Es igual al error estándar dividido entre la media (definido sólo cuando la media no se anula);

$$CV(\bar{y}) = \frac{EE(\bar{y})}{\bar{y}}$$

Estos estimadores se aplican a la estimación del total de la población, pues para estimar el total $Y = N\bar{Y} = \sum_{i=1}^N Y_i$ utilizamos el estimado insesgado:

$$\hat{Y} = N\hat{\bar{Y}} = N\bar{y}$$

Además, en general, si $\hat{\theta} \sim N[\theta, V(\hat{\theta})]$, entonces:

$$P[\hat{\theta} - 1.96\sqrt{V(\hat{\theta})} \leq \theta \leq \hat{\theta} + 1.96\sqrt{V(\hat{\theta})}] = 0.95$$

Si no conocemos $V(\hat{\theta})$ tenemos que estimarla:

$$P[\hat{\theta} - 1.96\sqrt{\hat{V}(\hat{\theta})} \leq \theta \leq \hat{\theta} + 1.96\sqrt{\hat{V}(\hat{\theta})}] \cong 0.95$$

En el caso particular del *msa* se tiene:

$$\theta = \bar{Y}, \quad \hat{\theta} = \bar{y} \text{ y } V(\hat{\theta}) = V(\bar{y}) = \frac{S_y^2}{n} \left(1 - \frac{n}{N}\right)$$

$$P\left[\bar{y} - 1.96 \underbrace{\sqrt{\frac{S_y^2}{n} \left(1 - \frac{n}{N}\right)}}_{\delta} \leq \bar{Y} \leq \bar{y} + 1.96 \sqrt{\frac{S_y^2}{n} \left(1 - \frac{n}{N}\right)}\right] = 0.95$$

Entonces $P[|\bar{y} - \bar{Y}| < \delta] = 0.95$, $\delta =$ error absoluto.

Despejando n de la expresión $\delta = 1.96\sqrt{V(\bar{y})}$ se tiene:

$$n = \frac{1}{\frac{\delta^2}{(1.96)^2 S_y^2} + \frac{1}{N}} \cong \frac{(1.96)^2 S_y^2}{\delta^2}$$

Se usa S_y^2 por ser más sencilla, pero recordemos que:

$$E(y_i - E(y_i))^2 = E(y_i - \bar{Y})^2 = \sigma_y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N}$$

$$S_y^2 = \frac{N}{N-1} \sigma_y^2, \quad S_y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}$$

Tamaño de muestra

Como se mencionó anteriormente, en una muestra no basta con informar sobre la media, proporción o total de lo muestreado, también se debe proporcionar una indicación de la exactitud de sus estimaciones, los intervalos de confianza se utilizan para indicar esta exactitud.

Con frecuencia, un investigador mide distintas variables y tiene varios objetivos en un estudio. Para diseñar una muestra aleatoria simple se debe decidir la cantidad de error de muestreo en las estimaciones que sea tolerable y debe equilibrar la precisión de las estimaciones con el costo del estudio. Aunque pueden medir muchas variables, con frecuencia se debe centrar en una o dos respuestas que sean de interés fundamental en el estudio y utilizarlas para estimar el tamaño de la muestra. La precisión se logra mediante el tamaño absoluto de la muestra, no con la proporción de la población cubierta.

El valor de S_y^2 o σ_y^2 se estima con experiencias previas, una prueba piloto o bien se obtiene una primera aproximación sobre su valor, usando tablas y el conocimiento previo sobre la población.

Si se considera que \bar{y} no se ajusta a la distribución normal, se usa el criterio de fijar la magnitud de la varianza o del coeficiente de variación de \bar{y} . Se determina n para que produzca un coeficiente de variación dado (CV_0) usando estimaciones "gruesas" de \bar{Y} y de S_y^2 .

Así,

$$CV_0 = \frac{[V(\bar{y})]^{1/2}}{E(\bar{y})} = \frac{\left[\left(1 - \frac{n}{N}\right) \frac{S_y^2}{n} \right]^{1/2}}{\bar{Y}}$$

Despejando n , se obtiene:

$$n = \frac{S_y^2}{(CV_0)^2 \bar{Y}^2 + \frac{S_y^2}{N}}$$

Si n es grande se espera que el Teorema central del límite de una buena aproximación de la distribución de \bar{y} . Así:

$$\bar{y} \sim N[\bar{Y}, V(\bar{y})]$$

$$P\left[\bar{y} - z_{\alpha/2} \sqrt{v(\bar{y})} \leq \bar{Y} \leq \bar{y} + z_{\alpha/2} \sqrt{v(\bar{y})}\right] = 1 - \alpha$$

Si $1 - \alpha = 0.95$

$$P\left[\bar{y} - 1.96 \sqrt{\frac{S_y^2}{n} \left(1 - \frac{n}{N}\right)} \leq \bar{Y} \leq \bar{y} + 1.96 \sqrt{\frac{S_y^2}{n} \left(1 - \frac{n}{N}\right)}\right] = 0.95,$$

Entonces $\frac{\bar{y} - \bar{Y}}{[V(\bar{y})]^{1/2}}$ se distribuye aproximadamente como una normal

estandarizada (media cero y varianza uno), donde $V(\bar{y}) = \frac{S_y^2}{n} \left(1 - \frac{n}{N}\right)$.

Si se desea un tamaño de muestra tal que el error de estimación sea inferior a δ con una probabilidad de $1 - \alpha$, esto es:

$$P[|\bar{y} - \bar{Y}| < \delta] = 1 - \alpha, \quad \delta = z_{\alpha/2} \sqrt{V(\bar{y})}$$

Dividiendo entre $\sqrt{V(\bar{y})}$

$$P\left[\frac{|\bar{y} - \bar{Y}|}{\sqrt{V(\bar{y})}} < \frac{\delta}{\sqrt{V(\bar{y})}}\right] = 1 - \alpha$$

De las tablas de la normal estándar, $Z \sim N(0, 1)$, se obtiene un valor $z_{\alpha/2}$ tal que $P[Z < z_{\alpha/2}] = 1 - \alpha$ ($z_{\alpha/2}$ es el valor de Z obtenido en las tablas que deja un área

de $\alpha/2$ a la derecha de él). Como $\frac{|\bar{y} - \bar{Y}|}{\sqrt{V(\bar{y})}} \sim N(0, 1)$, entonces:

$$z_{\alpha/2} = \frac{\delta}{\sqrt{V(\bar{y})}} = \frac{\delta}{\sqrt{\frac{S_y^2}{n} \left(1 - \frac{n}{N}\right)}}$$

De aquí se despeja n:

$$n = \frac{1}{\frac{\delta^2}{z_{\alpha/2}^2 S_y^2} + \frac{1}{N}} \cong \frac{z_{\alpha/2}^2 S_y^2}{\delta^2} \quad (1)$$

Se puede usar $n' = \frac{z_{\alpha/2}^2 S_y^2}{\delta^2}$ como una primera aproximación y luego corregir

usando

$$n = \frac{n'}{1 + \frac{n'}{N}}$$

Si no se puede suponer normalidad de la distribución del estimador, se recurre a la desigualdad de Tchebycheff.

Desigualdad de Tchebycheff

Sea U una variable aleatoria con cualquier distribución y $E(U)=\mu_U$, $V(U)=\sigma_U^2$

Entonces

$$P[|U - \mu_U| \geq \lambda \sigma_U] \leq \frac{1}{\lambda^2}$$

$$\Rightarrow P[|U - \mu_U| \geq \lambda \sigma_U] \geq 1 - \frac{1}{\lambda^2}$$

$$\Rightarrow P[U - \lambda \sigma_U \leq \mu_U \leq U + \lambda \sigma_U] \geq 1 - \frac{1}{\lambda^2}$$

$$\Rightarrow P[\bar{y} - \lambda \sqrt{V(\bar{y})} \leq \bar{Y} \leq \bar{y} + \lambda \sqrt{V(\bar{y})}] \geq 1 - \frac{1}{\lambda^2}$$

$$\lambda = 2 \quad 1 - \frac{1}{\lambda^2} = 0.75$$

$$\lambda = 3 \quad 1 - \frac{1}{\lambda^2} = 0.889$$

$$\lambda = 4.4 \quad 1 - \frac{1}{\lambda^2} = 0.95, \quad \delta = 4.4 \sqrt{V(\bar{y})}$$

En la expresión $n = \frac{1}{\frac{\delta^2}{z_{\alpha/2}^2 S_y^2} + \frac{1}{N}}$ si tanto δ como S se expresan como

porcentajes de la media, $\delta' = \frac{\delta}{y} 100$, $CV = \frac{S}{y} 100$ y suponiendo normalidad, la

expresión se transforma en:

$$n = \frac{1}{\frac{(\delta')^2}{z_{\alpha/2}^2 (CV)^2} + \frac{1}{N}} \cong \frac{z_{\alpha/2}^2 (CV)}{(\delta')^2}$$

Estimación de proporciones

Todo lo anterior es aplicable al caso en que $Y(U_i)$ es una medida o indicador de la presencia o ausencia de una característica en la unidad U_i con valor 1 si la característica está presente y 0 si no es así. En este caso $\bar{Y} = P$ representa la proporción de unidades en la población que tienen la característica:

$$\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N} = P$$

$p = \bar{y}$ es la proporción de unidades en la muestra con la característica. El valor de S_y^2 en términos de P resulta:

$$S_y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1} = NP(1-P) \frac{1}{N-1} = \frac{N}{N-1} P(1-P), \quad \sigma^2 = P(1-P)$$

Con estimador:

$$\hat{S}_y^2 = s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{n\hat{P}}{n-1} = \frac{N}{N-1} (1-\hat{P})$$

Con este nuevo valor, resulta:

$$n = \frac{\frac{N}{N-1} (1-P)}{(CV_0)^2 P + \left(\frac{1-P}{N-1}\right)} \cong \frac{1-P}{P(CV_0)^2}$$

Para usar esta expresión, se estima a priori o con una prueba piloto el valor de P y se fija el CV_0 que se desea.

Si utilizamos la desigualdad de Tchebycheff tenemos:

$$n = \frac{1}{\frac{\delta^2}{Z_{\alpha/2}^2 \frac{N}{N-1} P(1-P)} + \frac{1}{N}} \cong \frac{Z_{\alpha/2}^2 \frac{N}{N-1} P(1-P)}{\delta^2}$$

Nótese que si P está cercano a cero, el valor de n aumenta. Esto indica que para estimar la proporción de unidades con una característica rara se requieren muchas unidades en la muestra. Esto es lo contrario de lo que sucede si se usa la aproximación a la normal, en cuyo caso se usa la expresión

$$n = \frac{1}{\frac{\delta^2}{z_{\alpha/2}^2 S_y^2} + \frac{1}{N}} \cong \frac{z_{\alpha/2}^2 S_y^2}{\delta^2} \text{ con } S_y^2 = \frac{N}{N-1} P(1-P)$$

Si se quiere conocer P , las Y_i son 0 o 1, entonces:

$$S_y^2 = \frac{N}{N-1} P(1-P) \cong P(1-P)$$

$$\Rightarrow n \cong \frac{z_{\alpha/2}^2 P(1-P)}{\delta^2} \quad (2)$$

2.4 Muestreo sistemático

Uno de los diseños muestrales más socorridos es el muestreo sistemático. Corresponde a una opción alterna del muestreo simple aleatorio.

El caso $N=nk$

En forma breve, este muestreo consiste en tomar a los elementos de la población en grupos (“clusters”) de tamaño conveniente, n , de tal forma que $N=nk$. Cada “cluster” tendrá así n elementos y habrá k clusters para que uno de ellos se elija.

El arreglo siguiente resume la información respectiva:

	1	2	...	k-1	k		1	2	...	k-1	k
$\left[\begin{array}{ccccc} Y_1 & Y_2 & \cdots & Y_{k-1} & Y_k \\ Y_{1+k} & Y_{2+k} & \cdots & Y_{k-1+k} & Y_{k+k} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ Y_{1+(n-1)k} & Y_{2+(n-1)k} & \cdots & Y_{k-1+(n-1)k} & Y_{k+(n-1)k} \end{array} \right]$	\equiv	$\left[\begin{array}{ccccc} Y_1 & Y_2 & \cdots & Y_{k-1} & Y_k \\ Y_{1+k} & Y_{2+k} & \cdots & Y_{2k-1} & Y_{2k} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ Y_{1+(n-1)k} & Y_{2+(n-1)k} & \cdots & Y_{nk-1} & Y_{nk} \end{array} \right]$									
Media:	\bar{Y}_1	\bar{Y}_2	...	\bar{Y}_{k-1}	\bar{Y}_k		\bar{Y}_1	\bar{Y}_2	...	\bar{Y}_{k-1}	\bar{Y}_k

En la notación usual (se ha preferido aplicar la notación que maneja Cochran) las observaciones de Y de cada elemento de la población dentro de cada *cluster* se debiesen caracterizar como Y_i ; sin embargo, en muchas ocasiones (y Cochran así lo hace) al momento de seleccionar una muestra sistemática (un cluster) se redefinen como y_i (aunque no exista submuestreo).

La correspondencia entre el **muestreo sistemático** y el **muestreo simple aleatorio** se genera, a partir del arreglo anterior, como si el **muestreo sistemático** fuese un **muestreo de medias o promedios**.

De aquí que si existen k clusters (con k medias: $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_k$) la elección aleatoria de un número entre 1 y k es equivalente a la elección aleatoria de un **cluster** de n elementos etiquetado de 1 a k .

Por lo tanto, si se desea estimar \bar{Y} se obtiene \bar{Y}_1 o \bar{Y}_2 o ... o \bar{Y}_k ; la media que sea es \bar{Y}_{sys}

$$E[\bar{Y}_{sys}] = \sum \bar{Y}_k P[\bar{Y}_k \in S] = \sum \bar{Y}_k \frac{1}{k} = \frac{1}{k} \sum \bar{Y}_k$$

como $\bar{Y}_k = \frac{\sum Y_{ik}}{n} \Rightarrow E[\bar{Y}_{sys}] = \frac{1}{k} \sum_k \sum_i \frac{Y_{ik}}{n} = \frac{1}{nk} \sum \sum Y_{ik}$ que es una manera de

expresar $\frac{\sum_{j=1}^N Y_j}{N} = \bar{Y}$ y por lo tanto es un estimador insesgado.

La varianza de \bar{Y}_{sys} se calcula así:

Si $Var(\bar{y}) = (1-f) \frac{S^2}{n};$ $f = \frac{n}{N}$

Entonces $Var(\bar{y}_{sys}) = \left(1 - \frac{1}{k}\right) \frac{S_{sys}^2}{1};$ $f = \frac{1}{k}$

$$Var(\bar{y}_{sys}) = \frac{(k-1)}{k} \frac{\sum (\bar{y}_i - \bar{Y})^2}{k-1} = \frac{\sum (\bar{y}_i - \bar{Y})^2}{k};$$

$$\bar{Y} = \frac{\sum \sum Y_{ik}}{nk}$$

Es obvio que si \bar{Y} se desconoce tendrá que ser estimada por \bar{Y}_c i. e., por la media del único cluster sistemático en muestra.

Por lo tanto, $\hat{Var}(\bar{y}_{sys}) = Var(\bar{y}_{sys})$ no está determinada. Se tiene un estimador cuya varianza no es, por lo general, estimable. Y tampoco lo es el estimador de la varianza. Este problema se ha tratado de resolver usando distintos atajos en los que se ha destacado notablemente Leslie Kish.

Si $Var(\bar{y}_{sys}) = \frac{1}{k} \sum_{i=1}^k (\bar{y}_i - \bar{Y})^2$, existen otras formas de expresión de esta varianza que separan tanto los componentes intra cluster y los componentes entre cluster.

Para tal efecto conviene recordar diversas igualdades importantes. Algunas de ellas son las siguientes:

$$1) (N-1)S^2 = \sum_i (Y_i - \bar{Y})^2 = \sum_i \sum_j (Y_{ji} - \bar{Y})^2;$$

donde Y_{ji} es el valor Y asociado a la j -ésima unidad de la población del i -ésimo conglomerado

$$2) (N-1)S^2 = \sum_i \sum_j (Y_{ji} - \bar{Y}_i)^2 + n \sum_i (\bar{Y}_i - \bar{Y})^2; \text{ puesto que } \sum_j (Y_{ji} - \bar{Y}_i) = 0$$

$$3) n \sum_i (Y_i - \bar{Y})^2 = nk Var(\bar{y}_{sys}) = N Var(\bar{y}_{sys}); \quad Var(\bar{y}_{sys}) = \frac{\sum (\bar{Y}_i - \bar{Y})^2}{k}$$

$$4) N Var(\bar{y}_{sys}) = (N-1)S^2 - \sum_i \sum_j (Y_{ji} - \bar{Y}_i)^2 = (N-1)S^2 - k(n-1)S_{wsy}^2; \text{ definición de cuasivarianza (intra)}$$

$$5) Var(\bar{y}_{sys}) = \frac{(N-1)S^2}{N} - \frac{k(n-1)S_{wsy}^2}{N}$$

lo que implica que $\frac{(N-1)S^2}{N} \geq \frac{k(n-1)S_{wsy}^2}{N}$

ie. $(N-1)S^2 \geq k(n-1)S_{wsy}^2; N=nk$

Todo lo anteriormente apuntado se refiere al caso del muestreo sistemático cuando N (el tamaño de la población) es divisible por n (el tamaño muestral). Si esto no es así hay varias opciones de resolución. Algunos se describen a continuación.

Si N no es divisible por n pero N no es número primo

Se inicia con un caso ilustrativo muy sencillo. Supóngase que para una población de tamaño $N=16$ se desea obtener una muestra sistemática de tamaño $n=3$. Como se advierte, N no es divisible por n pero N no es un número primo.

Si se formasen “clusters” de tamaño tres se encontraría un arreglo desbalanceado. Por ejemplo como éste:

1	2	3	4	5
Y_1	Y_2	Y_3	Y_4	Y_5
Y_6	Y_7	Y_8	Y_9	Y_{10}
Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}
Y_{16}				

se advierte aquí que $P[Y_i \in S] = \frac{1}{5}$ si Y_i está en los clusters etiquetados por 2, 3, 4 y

5. En cambio si saliera el cluster etiquetado por el “1” habría que descartar a algún elemento

$$\therefore P[Y_i \in S] = P[1 \in S]P[Y_i \text{ no se descarte}]$$

$$P[Y_i \in S] = \frac{1}{5} \cdot \frac{3}{4} = \frac{3}{20} \text{ que es distinta de la probabilidad anterior. Por lo tanto}$$

este tipo de arreglo no tiene probabilidades iguales para toda la población. En consecuencia, no es autoponderado.

Aunque para muchos practicantes este es un problema irrelevante, no lo es tanto cuando es producto del descuido o de la falta de aplicación de otros métodos que

ayudan a resolverlo. En último caso, se puede decidir que es “irrelevante” si se conocen otras opciones, no antes.

Una manera de resolverlo sería la de obtener clusters sistemáticos de igual tamaño y, posteriormente, eliminar elementos o unidades (no completar el cluster). Para $n = 4$ y $k = 4$, $N = nk = 16$ se forman cuatro clusters con un arranque aleatorio entre 1 y 4 y con un tamaño inicial de la muestra sistemática de 4, así:

1	2	3	4
Y_1	Y_2	Y_3	Y_4
Y_5	Y_6	Y_7	Y_8
Y_9	Y_{10}	Y_{11}	Y_{12}
Y_{13}	Y_{14}	Y_{15}	Y_{16}

La probabilidad de inclusión en muestra de cualquier cluster es: $P[i \in S] = \frac{1}{4}$; con $i = 1, 2, 3, 4$

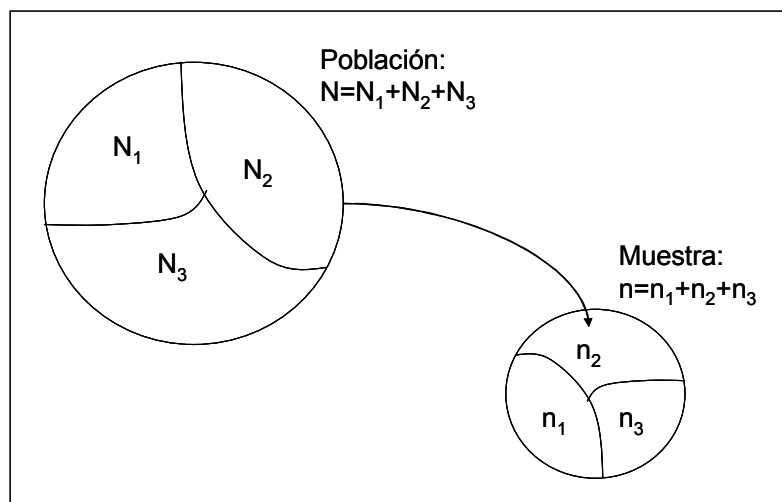
2.5 Muestreo estratificado

La esencia de la estratificación es la clasificación de la población en subpoblaciones o estratos basados en información disponible y entonces se selecciona muestras separadas por cada estrato, es decir, si la variable que nos interesa asume distintos valores promedio en diferentes subpoblaciones, podríamos obtener estimaciones más precisas de las cantidades de la población al tomar una muestra aleatoria estratificada. La palabra estratificar proviene de la palabra latina que significa “formar capas”, así pues dividimos a la población en L subpoblaciones, llamadas estratos. Los estratos no se traslapan y conforman la población completa, de modo que cada unidad de muestreo pertenece exactamente a un estrato. Extraemos una muestra independiente de cada estrato (y por lo tanto, se pueden utilizar distintas formas de muestreo dentro de cada estrato) y después reunimos la información para obtener las estimaciones globales de la población.

Otra razón para formar estratos es la disponibilidad de marcos. Si para una parte de la población se tiene un buen marco, éste se usa para el muestreo de esa parte y la o las otras partes de la población se muestrean utilizando otros marcos más imprecisos y posiblemente distintos esquemas de muestra. Otra razón puede ser el costo de localizar y levantar la información de las unidades.

Lo más frecuente es que los tres criterios para la formación de estratos coincidan, de modo que los estratos formen unidades homogéneas con un mismo tipo de marco y con costos de localización y captación de información semejantes.

El siguiente dibujo explica mejor el muestreo estratificado:



Como ya se mencionó, la población original de tamaño N es fragmentada en L estratos de los cuales el h -ésimo es de tamaño N_h , $h = 1, 2, \dots, L$.

Considérese la siguiente notación:

Valores poblacionales:

Y_{hi} Valor de la medición en el elemento i del estrato h

$$\bar{Y}_h = \frac{\sum_{i=1}^{N_h} Y_{hi}}{N_h} \quad \text{Media poblacional del estrato } h$$

$$Y_h = N\bar{Y}_h = \sum_{i=1}^{N_h} Y_{hi} \quad \text{Total poblacional del estrato h}$$

$$S_h^2 = \frac{\sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2}{N_h - 1} \quad \text{Varianza poblacional del estrato h}$$

$$Y = \sum_{h=1}^L Y_h = \sum_{h=1}^L \sum_{i=1}^{N_h} Y_{hi} = \sum_{i=1}^L N_h \bar{Y}_h \quad \text{Total de la población}$$

$$\bar{Y} = \frac{Y}{\sum_{h=1}^L N_h} \quad \text{Media de los valores } Y_{hi} \text{ en toda la población}$$

$$W_h = \frac{N_h}{N}, \quad \sum_{h=1}^L W_h = 1 \quad \text{Proporción del tamaño del estrato h (peso del estrato)}$$

Consideremos que tenemos un *msa* en cada estrato.

A nivel muestral tenemos lo siguiente:

$$n_h \quad \text{Tamaño de la muestra del estrato h}$$

$$n = \sum_{h=1}^L n_h \quad \text{Tamaño total de la muestra}$$

$$\hat{Y}_h = \bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} \quad \text{Estimador de la media del estrato h}$$

$$\hat{Y}_h = N_h \bar{y}_h = \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} \quad \text{Estimador del total del estrato h}$$

$$s_h^2 = \frac{\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2}{n_h - 1} \quad \text{Estimador de la varianza del estrato h}$$

El estimador del total poblacional es:

$$\hat{Y} = \sum_{h=1}^L \hat{Y}_h = \sum_{h=1}^L N_h \bar{y}_h = \sum_{h=1}^L N_h \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h} = \sum_{h=1}^L \sum_{i=1}^{n_h} \frac{N_h}{n_h} y_{hi}$$

donde $\frac{N_h}{n_h}$ es el factor de expansión.

La varianza del estimador del total es:

$$V(\hat{Y}) = \sum_{h=1}^L V(\hat{Y}_h) \text{ Como son muestras independientes en cada estrato}$$

$$V(\hat{Y}) = \sum_{h=1}^L V(N_h \bar{y}_h) = \sum_{h=1}^L N_h^2 V(\bar{y}_h)$$

Como tenemos un *msa* en cada estrato,

$$V(\hat{Y}) = \sum_{h=1}^L N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

El estimador de la varianza del estimador del total es:

$$\hat{V}(\hat{Y}) = \sum_{h=1}^L N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{S}_h^2}{n_h}$$

donde,

$$\hat{S}_h^2 = \frac{\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2}{n_h - 1}$$

Si el tamaño de muestra en cada estrato es grande y podemos hacer la aproximación a la normal del estimador del total, el intervalo aproximado del $(1 - \alpha) \cdot 100\%$ de confianza para el total poblacional es:

$$\hat{Y} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y})}$$

El estimador de la media poblacional es:

$$\hat{Y} = \frac{\hat{Y}}{N} = \frac{\sum_{h=1}^L N_h \bar{y}_h}{N} = \sum_{h=1}^L \frac{N_h}{N} \bar{y}_h = \sum_{h=1}^L W_h \bar{y}_h$$

\hat{Y} es la suma ponderada de los promedios muestrales en cada estrato.

La varianza del estimador de la media es:

$$V(\hat{Y}) = V\left(\sum_{h=1}^L W_h \bar{y}_h\right) = \sum_{h=1}^L W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

El estimador de la varianza del estimador de la media es:

$$\hat{V}(\hat{Y}) = \sum_{h=1}^L W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{S}_h^2}{n_h}$$

Si el tamaño de muestra en cada estrato es grande y podemos hacer la aproximación a la normal del estimador de la media, el intervalo aproximado del $(1 - \alpha) \cdot 100\%$ de confianza para la media poblacional es:

$$\hat{Y} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y})}$$

Estimador de una proporción

Como ya se sabe, una proporción es una media de una variable que asume los valores 0 y 1. Para hacer inferencias sobre proporciones utilizaremos las ecuaciones anteriores pero haciendo $\bar{y}_h = \hat{p}_h$ y $S_h^2 = \frac{n_h}{n_h - 1} \hat{p}_h (1 - \hat{p}_h)$. Entonces,

El estimador de la proporción P de unidades que tienen cierta característica es:

$$\hat{P} = \sum_{h=1}^L W_h \hat{p}_h \quad \text{con} \quad \hat{p}_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h}$$

La varianza de este estimador:

$$V(\hat{P}) = \sum_{h=1}^L W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{P_h (1 - P_h)}{n_h}$$

con estimador:

$$\hat{V}(\hat{P}) = \sum_{h=1}^L W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{p}_h (1 - \hat{p}_h)}{n_h - 1}$$

Si el tamaño de muestra en cada estrato es grande y podemos hacer la aproximación a la normal del estimador de la proporción, el intervalo aproximado del $(1 - \alpha) \cdot 100\%$ de confianza para la proporción poblacional es:

$$\hat{P} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{P})}$$

Distribución de la muestra a los estratos

Distribución proporcional

Un criterio es lo que se le llama distribución (afijación) proporcional, donde la muestra se divide de manera proporcional a los tamaños de los estratos N_h . Se busca que se cumpla la relación:

$$\frac{n_h}{n} = \frac{N_h}{N} = W_h$$

De esta relación se tiene:

$$n_h = n \frac{N_h}{N} = n W_h$$

Esta distribución de la muestra total se usa cuando no se tiene información sobre la magnitud de las S_h^2 , o cuando las S_h^2 sean semejantes; se usa además cuando los costos de muestrear las unidades en los diferentes estratos son semejantes. También se emplea cuando el muestreo o encuesta va a determinar varias características (varias mediciones) en cada unidad de la población. También cuando se quiere que la muestra sea autoponderada, es decir, que todos los elementos de la muestra tienen un mismo factor de expansión $\frac{N_h}{n_h} = \frac{N}{n}$. Con esta

distribución proporcional se tiene:

$$\hat{Y} = \sum_{h=1}^L \hat{Y}_h = \sum_{h=1}^L N_h \bar{y}_h = \sum_{h=1}^L N_h \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h} = \sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} = k \sum_{h=1}^L \sum_{i=1}^{n_h} y_{hi}$$

$$\text{donde } k = \frac{N_h}{n_h} = \frac{N_h}{n \frac{N_h}{N}} = \frac{N}{n}$$

Distribución óptima

Cuando se tienen costos muy diferentes para el muestreo de unidades en los diferentes estratos, se usa la distribución óptima. Si el costo para obtener información de una unidad en el estrato h es C_h , el costo total será:

$$C = C_0 + \sum_{h=1}^L C_h n_h$$

C_0 es el costo adicional y C_h el costo de una observación en el estrato h .

La varianza del estimador \hat{Y} y el costo se minimizan cuando:

$$n_h = n \frac{N_h S_h}{\sqrt{C_h}} \left[\sum_{h=1}^L \frac{N_h S_h}{\sqrt{C_h}} \right]^{-1}$$

Observe que,

$$n_h \propto \frac{N_h S_h}{\sqrt{C_h}}$$

Esto quiere decir que en un estrato dado, se toma más muestra si:

- ✓ Es estrato es más grande.
- ✓ El estrato es más variable.
- ✓ El costo es menor.

Distribución de Neyman

Si se considera que los costos C_h son constantes en todos los estratos:

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^L N_h S_h}$$

Tamaño de muestra

1. Consideremos la distribución óptima:

$$n_h = n \frac{N_h S_h}{\sqrt{C_h}} \left[\sum_{h=1}^L \frac{N_h S_h}{\sqrt{C_h}} \right]^{-1}$$

A) Valor de n que produce varianza mínima para un costo total fijo.

$$C = C_0 + \sum_{h=1}^L n_h C_h$$

Sustituyendo la expresión para n_h y despejando n :

$$\begin{aligned}
C - C_0 &= \sum_{h=1}^L n_h C_h \\
C - C_0 &= \sum_{h=1}^L C_h \left[n \frac{N_h S_h}{\sqrt{C_h}} \left(\sum_{i=1}^L \frac{N_i S_i}{\sqrt{C_i}} \right)^{-1} \right] \\
&= n \sum_{h=1}^L \frac{C_h N_h S_h}{\sqrt{C_h}} \left(\sum_{i=1}^L \frac{N_i S_i}{\sqrt{C_i}} \right)^{-1} \\
n &= \frac{(C - C_0) \sum_{h=1}^L \frac{N_h S_h}{\sqrt{C_h}}}{\sum_{h=1}^L N_h S_h \sqrt{C_h}}
\end{aligned}$$

B) Valor de n que produce costo mínimo para una varianza fija, o equivalentemente para un error de estimación fijo

$$\delta = z_{\alpha/2} \sqrt{V(\hat{Y})}$$

a) Para estimar la media

$$\hat{V}(\hat{Y}) = \sum_{h=1}^L W_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_h^2}{n_h} = \sum_{h=1}^L \frac{N_h^2}{N^2} \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_h^2$$

La asignación óptima es:

$$\begin{aligned}
n_h &= n \frac{N_h S_h}{\sqrt{C_h}} \left[\sum_{i=1}^L \frac{N_i S_i}{\sqrt{C_i}} \right]^{-1} \\
\delta &= z_{\alpha/2} \sqrt{V(\hat{Y})} \Rightarrow V(\hat{Y}) = \frac{\delta^2}{z_{\alpha/2}^2}
\end{aligned}$$

Sustituyendo en la primera ecuación y despejando n:

$$n = \frac{\sum_{h=1}^L N_h S_h \sqrt{C_h} \left[\sum_{i=1}^L \frac{N_i S_i}{\sqrt{C_i}} \right]}{N^2 \frac{\delta^2}{z_{\alpha/2}^2} + \sum_{h=1}^L N_h S_h^2}$$

b) Para estimar el total

$$\hat{V}(\hat{Y}) = \sum_{h=1}^L N_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_h^2}{n_h}$$

$$\delta = z_{\alpha/2} \sqrt{V(\hat{Y})} \Rightarrow V(\hat{Y}) = \frac{\delta^2}{z_{\alpha/2}^2}$$

$$n_h = n \frac{N_h S_h}{\sqrt{C_h}} \left[\sum_{i=1}^L \frac{N_i S_i}{\sqrt{C_i}} \right]^{-1}$$

Sustituyendo y despejando n :

$$n = \frac{\sum_{h=1}^L N_h S_h \sqrt{C_h} \left[\sum_{i=1}^L N_i S_i / \sqrt{C_i} \right]}{\frac{\delta^2}{z_{\alpha/2}^2} + \sum_{h=1}^L N_h S_h^2}$$

2. Considerando la asignación de Neyman (Costos C_h constantes)

$$n_h = n \frac{N_h S_h}{\sum_{i=1}^L N_i S_i}$$

a) Para estimar la media:

$$n = \frac{\left[\sum_{h=1}^L N_h S_h \right]^2}{N^2 \frac{\delta^2}{z_{\alpha/2}^2} + \sum_{h=1}^L N_h S_h^2}$$

b) Para estimar el total:

$$n = \frac{\left[\sum_{h=1}^L N_h S_h \right]^2}{\frac{\delta^2}{z_{\alpha/2}^2} + \sum_{h=1}^L N_h S_h^2}$$

3. Si consideramos la distribución proporcional:

$$n_h = n \frac{N_h}{N}$$

a) Para estimar la media:

$$\hat{V}(\hat{Y}) = \sum_{h=1}^L \frac{N_h^2}{N^2} \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_h^2$$

$$\delta = z_{\alpha/2} \sqrt{V(\hat{Y})} \Rightarrow V(\hat{Y}) = \frac{\delta^2}{z_{\alpha/2}^2}$$

$$n = \frac{N \sum_{h=1}^L N_h S_h}{N^2 \frac{\delta^2}{z_{\alpha/2}^2} + \sum_{h=1}^L N_h S_h^2}$$

b) Para estimar el total

$$n = \frac{N \sum_{h=1}^L N_h S_h^2}{\frac{\delta^2}{z_{\alpha/2}^2} + \sum_{h=1}^L N_h S_h^2}$$

Nota: Para estimar proporciones se utiliza las expresiones de muestra para estimar la media con $S_h^2 = P_h Q_h$

Todas las expresiones anteriores determinan el tamaño de muestra para estimadores globales de toda la población. **Las inferencias no son para cada estrato con esas muestras.** Si lo que se desea es estimar medias, totales o proporciones en cada estrato, las expresiones anteriores no se deben usar, lo que se debe emplear son las fórmulas dadas para el msa (1) y (2) **para cada estrato por separado** y así determinar las n_h a usarse en cada uno de ellos. Por supuesto que en este último caso la muestra total n es mucho más grande. Esto es de esperarse, puesto que ahora se están haciendo inferencias por separado para L poblaciones.

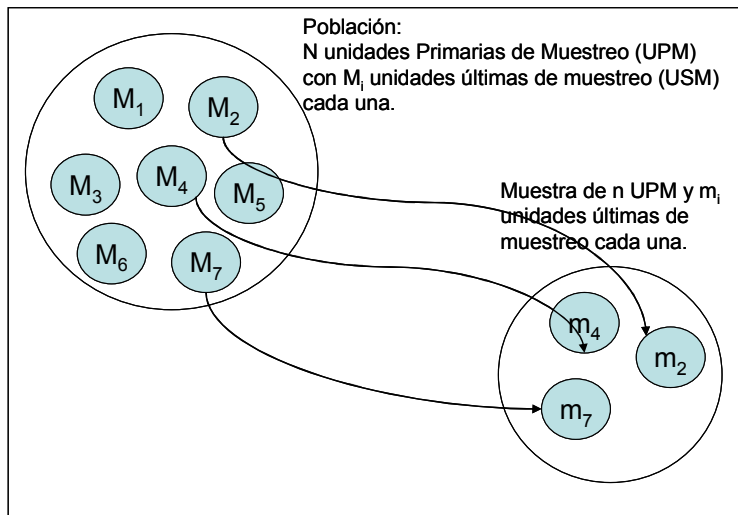
2.6 Muestreo por conglomerados y muestreo por etapas

Después de haber revisado tanto el muestreo simple aleatorio (en una sola etapa de muestreo y sin etapas de diseño) como el muestro estratificado simple toca el turno al muestreo por conglomerados. En contraste al muestro estratificado simple que considera una etapa de diseño (una estratificación) que involucra una partición de la población objetivo, el muestreo por conglomerados da la oportunidad al especialista de escoger sólo partes o raciones de la población objetivo. Ésta en sí misma es una característica que produce ventajas (de costo, de control de las operaciones de campo, de aprovechamiento de las condiciones

del marco área, principalmente) y desventajas (las relativas a la disminución de la precisión y a merma de posibles pruebas estadísticas sobre subconjuntos que cruzan toda la población).

Por dichas razones, muchas veces se ha propuesto la inserción del muestreo por conglomerados (sea en una sola etapa de muestreo o con dos etapas de muestreo, esto es con submuestreo en los conglomerados) al compararlo con el estratificado simple. También se usará aquí así en primera instancia.

La siguiente figura esquematiza al Muestreo Bietápico



En el muestreo estratificado simple, todos y cada uno de los elementos de la población objetivo están en un y sólo un conjunto de la partición (en un estrato). Todos los estratos participan en muestra.

En el muestreo estratificado, lo más deseable (desde el punto de vista de la precisión) es que los estratos sean internamente homogéneos. Esto es que los elementos de un estrato entre sí se asemejen en lo que respecta a la(s) característica(s) en estudio que en relación a elementos de otro estrato.

Esto se puede advertir rápidamente a partir de la Varianza de un estimador como $\bar{y} = \bar{Y}$ en el muestreo estratificado simple.

$$\text{Así } \text{Var}(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 (1 - f_h) \frac{S_h^2}{n_h}$$

Donde $W_h = \frac{N_h}{N}$ es el “peso” del Estrato;

$f_h = \frac{n_h}{N_h}$ es la fracción de muestreo;

S_h^2 La cuasivarianza por estrato; y

n_h el tamaño de la muestra en el estrato h , $h=1, \dots, L$

En otros términos $\text{Var}(\bar{y}_{st}) = \sum \theta_h S_h^2$; $\theta_h = \frac{W_h^2 (1 - f_h)}{n_h}$

En contraste, para \bar{y}_c como estimador de \bar{Y}_c en el muestreo por conglomerados

(con $\bar{y}_c = \hat{Y}_c = \frac{\sum y_i}{\sum m_i}$, un estimador de razón)

$$\text{Var}(\bar{y}_c) = \frac{(K - k)}{K \cdot k \bar{N}^2} \sum_{i=1}^k \frac{(y_i - \bar{y}_c m_i)^2}{k - 1} = \frac{(K - k)}{K \cdot k \bar{N}^2 (k - 1)} \sum_{i=1}^k (y_i - \bar{y}_c m_i)^2$$

Una varianza que depende de las diferencias entre observaciones de conglomerados y la media poblacional.

Con K : número de conglomerados en la población

k : número de conglomerados en muestra

M_i : el número de elementos en el conglomerado i

Es decir, se compara:

$$\text{Var}(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 (1 - f_h) \frac{S_h^2}{n_h} \qquad \text{Var}(\bar{y}_c) = \frac{(K - k)}{K \cdot k \bar{N}^2} \sum_{i=1}^k \frac{(y_i - \bar{y}_c m_i)^2}{k - 1}$$

$$\text{o } \text{Var}(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^L N_h (N_h - n_h) \frac{S_h^2}{n_h}$$

Se advierte que los elementos básicos a comparar son:

- (a) las cuasivarianzas S_h^2 ($h = 1, \dots, L$) que sólo reportan las diferencias internas en los Estratos

$$\hat{S}_h^2 = \frac{\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2}{n_h - 1}; y$$

- (b) $\sum_{i=1}^k \frac{(y_i - \bar{y}_c m_i)^2}{k-1}$ que reporta diferencias entre cada conglomerado y la media de todos los conglomerados.

De aquí que la “recomendación general” que los Estratos sean internamente homogéneos (o lo más homogéneos posibles”) y los conglomerados sean internamente “lo más heterogéneos posibles” tiene sentido y significado. Esta proposición tiene, en la teoría estadística, validez.

El siguiente cuadro resume lo anterior

Cuadro 3. Comparativo entre muestreo estratificado y muestreo por conglomerados

	Internamente (intra-)	Externamente (inter-)	Precisión
Estratos	Homogéneos	Heterogéneos	$Var(\bar{y})_{opt} < Var(\bar{y})_{porp} < Var(\bar{y})_{msa}$
Conglomerados	Heterogéneos	Homogéneos	$Var(\bar{y})_{msa} < Var(\bar{y}_c)$

Este ejercicio de comparación advierte que para comparar $Var(\bar{y}_c)$ con $Var(\bar{y}_{st})$ es necesario pedagógicamente introducir a los estimadores de razón y “jugar” con las diferentes formas de reconocer conglomerados.

Al igual que en el muestreo sistemático

$$V(\bar{y}_{sys}) = \frac{(N-1)}{N} S^2 - \frac{k(n-1)}{N} S_{wsys}^2$$

Se tienen dos componentes de variabilidad. El primero corresponde a la variabilidad total (la cuasivarianza). El segundo se asocia a la variabilidad dentro de cada cluster sistemático (within) o sea S_{wsys}^2 . Así habrá que “descomponer” la varianza de cada estimador dentro de cada conglomerado (intra) y entre conglomerados.

La expresión general es:

$$S^2 = \frac{(N-1)}{NM-1} S_b^2 + \frac{N(M-1)}{NM-1} S_w^2$$

Donde S_b^2 es la varianza entre conglomerados y S_w^2 es la varianza dentro de cada conglomerado.

En la fórmula N es el número de clusters o conglomerados, M es el número de elementos de cada conglomerado (su tamaño) y cada conglomerado se supone de tamaño igual. Por lo tanto, el tamaño de la población es NM

S_b^2 está corregida por $\frac{(N-1)}{NM-1}$

Donde el numerador es igual a los grados de libertad entre conglomerados y el denominador es igual al total de grados de libertad

S_w^2 está corregida por $\frac{N(M-1)}{NM-1}$

Donde el numerador es igual a los grados de libertad dentro de los conglomerados y el denominador es igual al total de grados de libertad

Nótese que en el caso del muestreo sistemático hay k clusters ($N = k$) cada uno de tamaño n (por lo que $n = M$) lo que llevaría a que:

$$S^2 = \frac{(N-1)}{NM-1} S_b^2 + \frac{N(M-1)}{NM-1} S_w^2$$

Sustituyendo

$$S^2 = \frac{(k-1)}{nk-1} S_b^2 + \frac{k(n-1)}{nk-1} S_w^2$$

$$S^2 = \frac{1}{nk-1} [(k-1)S_b^2 + k(n-1)S_w^2]$$

Los cuadros de análisis de Varianza en un diseño en el que se seleccionan n conglomerados con probabilidades iguales (de N posibles) y cada uno de tamaño M , así como el cuadro de análisis de Varianza para toda la población se reportan a continuación.

Cuadro 4. Análisis de Varianza para toda la población

	Grados de libertad	Cuadrado medio
Entre conglomerados	$N - 1$	S_b^2
Intra conglomerados	$N(M - 1)$	S_w^2
Total (entre todas las unidades de la pob.)	$NM - 1$	$S^2 = \frac{(N-1)S_b^2 + N(M-1)S_w^2}{NM-1}$

Cuadro 5. Análisis de Varianza para la Muestra (una sola etapa de muestreo)

	Grados de libertad	Cuadrado medio
Entre conglomerados	$n - 1$	$s_b^2 = \hat{S}_b^2$
Intra conglomerados	$n(M - 1)$	$s_w^2 = \hat{S}_w^2$
Total (entre todas las unidades en muestra)	$nM - 1$	$s^2 = \frac{(n-1)s_b^2 + n(M-1)s_w^2}{nM-1}$

s^2 es estimador sesgado de S^2 debido al efecto de conglomeración.

En cambio, $\hat{s}^2 = \frac{(N-1)s_b^2 + n(M-1)s_w^2}{NM-1}$, sí es estimador insesgado de S^2

La que se aproxima por $s^2 \equiv \frac{s_b^2 + (M-1)s_w^2}{M}$

Con este pequeño arsenal se posibilita la presentación de la varianza del estimador $\bar{\bar{y}}$ de \bar{y}_c cuando se muestrean n clusters de N con probabilidades iguales y sin reemplazo teniendo todos los clusters igual tamaño: M

$$V(\bar{\bar{y}}) = \frac{(1-f)}{n} \cdot \frac{NM-1}{M^2(N-1)} \cdot S^2 [1 + (M-1)Roh]$$

Roh sigue siendo el coeficiente de correlación intraclase (o intra-cluster)

Nótese que la probabilidad de selección de cada cluster es $\frac{n}{N}$ y, por lo tanto, la probabilidad de selección de un elemento en un cluster determinado es también $\frac{n}{N}$; así es que $f = \frac{n}{N}$

Igual que en el caso sistemático este tipo de diseño muestral es equivalente a un *msa* de medias por cluster. Así que $\bar{y} = \frac{\sum y_i}{n}$ donde y_i es el total de la variable y para el i -ésimo cluster en muestra. \bar{y} es un estimador “natural”.

Las propiedades de \bar{y} son las siguientes:

1. \bar{y} es insesgado con respecto a \bar{Y} $\left(\bar{Y} = \frac{\sum Y_i}{N} \right)$

$$E(\sum y_i) = \sum Y_i \Rightarrow E(\sum y_i) = k(\sum Y_i)$$

k es un factor de ajuste (de tamaño) = $\frac{n}{N}$

2. $V(\bar{y}) = \frac{(1-f)}{n} \cdot \frac{\sum (y_i - \bar{Y})^2}{(N-1)}$ (aplicación directa del *msa*)

3. Ahora se reconoce que $\bar{Y}_c = \bar{\bar{Y}} = \frac{\bar{Y}}{M}$ (la media por unidad)

$$\left(\bar{Y}_c = \frac{\sum Y_i}{M} = \frac{\sum Y_i}{NM} = \sum_i \sum_j \frac{Y_{ij}}{NM} \right) \text{ por lo tanto si } \bar{y} \text{ es estimador insesgado de}$$

\bar{Y}_c (dentro del cluster)

$\bar{y}_c = \bar{y} = \frac{\bar{y}}{M}$ es estimador insesgado de \bar{Y}_c

$$4. \text{Var}(\bar{y}_c) = \text{Var}(\bar{y}) = \frac{1}{M^2} \text{Var}(\bar{y}) = \frac{1}{M^2} \left[\frac{(1-f)}{n(N-1)} \sum (y_i - \bar{Y})^2 \right]$$

5. Ya se ha visto que

$$(y_i - \bar{y}) = \left(\sum_j y_{ij} \right) - \bar{y} = \left(\sum_j y_{ij} \right) - \frac{M \bar{y}}{M} = \left(\sum_j y_{ij} \right) - M \bar{y}_c = \sum_j (y_{ij} - \bar{y}_c)$$

$$6. (y_i - \bar{Y})^2 = \left(\sum_j (y_{ij} - \bar{Y}_c) \right)^2 = \sum_j (y_{ij} - \bar{Y}_c)^2 + 2 \sum_{j \neq k} \sum (y_{ij} - \bar{Y}_c)(y_{ik} - \bar{Y}_c)$$

$$7. \sum_i (y_i - \bar{Y})^2 = \sum_i \sum_j (y_{ij} - \bar{Y}_c)^2 + 2 \sum_i \sum_{j \neq k} \sum (y_{ij} - \bar{Y}_c)(y_{ik} - \bar{Y}_c)$$

$$8. \therefore \sum (y_i - \bar{Y})^2 = (NM - 1)s^2 + (M - 1)(NM - 1)Roh s^2$$

$$9. \sum_i (y_i - \bar{Y})^2 = (NM - 1)s^2 [1 + (M - 1)Roh]$$

$$10. \therefore \text{Var}(\bar{y}_c) = \frac{1}{M^2} \frac{(1-f)}{n(N-1)} (NM - 1)s^2 [1 + (M - 1)Roh]$$

11. Una versión aproximada de $\text{Var}(\bar{y}_c)$ es:

$$\text{Var}(\bar{y}_c) \cong \frac{(1-f)}{nM} s^2 [1 + (M - 1)Roh] \cong \text{Var}(\bar{y}) [1 + (M - 1)Roh]$$

Es decir el *Deff* de este diseño se aproxima por $1 + (M - 1)Roh$

Así es que si, por ejemplo, una población de tamaño $NM = 2000$ esta "dividida" en $N=20$ clusters de $M=100$ elementos cada uno y se desea obtener una muestra de tamaño $N=400$ bajo un diseño por conglomerados (sin submuestreo) bastaría con seleccionar 4 clusters con probabilidades iguales y sin reemplazo.

La probabilidad de selección de cada cluster $= \frac{4}{20} = \frac{1}{5} = f$ y ésta es la misma probabilidad que tiene cada elemento en el cluster. En este caso:

$$Deff \cong [1 + (100 - 1)Roh] = [1 + (99)Roh]$$

Si la variable en estudio es indicadora o característica $y_i = 0,1$; $Var(p)$ según el *msa* es:

$$Var(p) = \frac{(1-f)NPQ}{n(N-1)} \cong \frac{(1-f)PQ}{n} \langle Var^*(p) = \frac{\left(1 - \frac{n}{N}\right)}{n} \left(\frac{1}{4}\right)$$

$Var^*(p)$ es el valor máximo para $Var(p)$

$$\therefore Var^*(p) \cong \frac{\left(1 - \frac{1}{5}\right)}{400} \cdot \frac{1}{4} = \frac{4}{5} \cdot \frac{1}{4} \cdot \frac{1}{400} = \frac{1}{2000} = 0.0005$$

(Equivalente a un margen de error en la estimación de ± 0.0438)

De aquí que $Var(p)$ en el cluster se estima por:

$$Var(p) \cong Var^*(p)(1 + 99Roh) = \frac{1}{2000}(1 + 99Roh)$$

(a) Un Valor posible de Roh es 0.001

$$\therefore Var(p) \cong \frac{1}{2000}(1 + 99(0.001)) = \frac{1}{2000}(1.099)$$

$Var(p) \cong 0.0005495$ (Equivalente a un margen de error en la estimación de ± 0.0459)

(b) Otro valor posible de Roh es 0.01

$$\therefore Var(p) \cong \frac{1}{2000}(1.99) = 0.000995$$
 (Equivalente a un margen de error

en la estimación de ± 0.0618)

Este ejemplo abre la puerta para apuntar las expresiones correspondientes a la varianza de un estimador de P en un muestreo por conglomerados.

Una primera instancia es la sustitución directa de $\hat{p} = \bar{\bar{p}}$ en la expresión general o en la aproximada. Esto es sustituir en:

$$V(\bar{y}) = \frac{(1-f)}{n} \frac{NM-1}{M^2(N-1)} s^2 [1 + (M-1)Roh] \quad (3)$$

$$V(\bar{p}) = \frac{(1-f)}{n} \frac{(NM-1)}{M^2(N-1)} \left(\frac{NPQ}{N-1} \right) [1 + (M-1)Roh] \quad (4)$$

Si el especialista se satisface con (4), aún desconociendo el valor de Roh, con esa expresión concluye la tarea. Si no es así, se requiere encontrar otra manera que describa pormenorizadamente a $Var(\bar{p})$, Un método es el siguiente:

Paso 1. A partir de $V(\bar{y})$, donde $\bar{y} = \sum y_i/n$ es la media por cluster

$$\text{De aquí que } V(\bar{y}) = \frac{(1-f)}{n} \frac{\sum (y_i - \bar{Y})^2}{N-1} = \frac{(1-f)}{n} \left(\frac{1}{N-1} \right) \sum (y_i^2 - 2y_i\bar{y} + \bar{y}^2)$$

$$V(\bar{y}) = V(p) = \frac{(1-f)}{n(N-1)} \left(\sum y_i - 2\bar{Y} \sum y_i + N\bar{Y}^2 \right) \quad (5)$$

$$= \frac{(1-f)}{n(N-1)} \left(\sum np_i - 2\bar{y} \sum np_i + NP^2 \right)$$

$$= \frac{(1-f)}{n(N-1)} \left[n \sum p_i - 2Pn \sum p_i + NP^2 \right]$$

$$Var(p) = \frac{(1-f)}{n(N-1)} \left[(n \sum P_i)(1-2P) + NP^2 \right]$$

$$\text{Paso 2. } Var(p) \cong \frac{(1-f)}{n(N-1)} [nP(1-2P) + NP^2] \cong \frac{(1-f)}{n(N-1)} [nP - 2nP^2 + NP^2]$$

Paso 3. A partir de $Var(p)$ en forma aproximada se puede obtener el valor máximo de $Var(p)$ y acotar la función.

Cochran propone un método más simple a partir de (5)

$$V(p) = \frac{(1-f)}{n} \sum \frac{(y_i - \bar{Y})^2}{N-1} \quad y_i = p_i; \bar{Y} = P$$

$$V(p) = \frac{(1-f)}{n} \frac{\sum (p_i - P)^2}{N-1}$$

$$V(p) = \frac{(N-n)}{nN} \frac{\sum (p_i - P)^2}{N-1}$$

Alternativamente, si se usa una muestra simple aleatoria (*msa*) de nM elementos, la varianza de p se obtendría usando los resultados binomiales.

$$\text{En un } msa \quad V(p) = V(\bar{y}) = (1-f) \frac{s^2}{n} = \frac{(1-f)}{n} \frac{NPQ}{N-1} = \frac{(N-n)PQ}{n(N-1)}$$

En un *msa* de nM elementos

$$V(p) = V(\bar{y}) = (1-f') \frac{s'^2}{n'} = \frac{(1-f')}{nM} \frac{NMPQ}{NM-1} = \frac{(NM-nM)PQ}{nM(NM-1)}$$

$$\therefore V(p) = \frac{(NM-nM)PQ}{nM(NM-1)} = \frac{(N-n)PQ}{n(NM-1)} \cong \frac{(N-n)PQ}{N(nM)}$$

(si N es grande)

3. El diseño muestral para el seguimiento operativo

3.1 Diseños muestrales

Los diseños muestrales probabilísticos son múltiples dependiendo (1) del número de etapas de diseño (las etapas de estratificación) y (2) de las etapas de muestreo (procedimientos de selección y estimación).

Etapas de diseño

Del diseño muestral más básico -que ya se mencionó anteriormente en estas notas- el muestreo simple aleatorio (msa) que se presenta como un diseño de una sola etapa de muestreo sin etapas de diseño, con probabilidades iguales de selección y sin reemplazo- la teoría y la práctica del muestreo fue avanzando.

En el siguiente nivel del msa está el diseño estratificado simple el que permite ilustrar los efectos de las etapas de diseño. Si una población objetivo tiene dispersión geográfica, presenta dificultades y oportunidades particulares, depende administrativa o políticamente de diferentes instancias tiene grupos más homogéneos o incluye subpoblaciones a los que se desea estudiar más profundamente o comparar, entonces la opción de estratificación parece conveniente.

En forma simple, la estratificación es una división interna o "partición" de la población objetivo en "pedazos" o "mosaicos" que tienen una definición precisa (en tiempo, espacio, extensión, etcétera.) y que cubren exhaustivamente a la población. Cada estrato o "mosaico" tiene la propiedad de que la selección de su propia muestra interna es independiente entre estrato y estrato. Todo estrato tiene que incluirse en la muestra.

La estratificación puede realizarse de acuerdo con un criterio geográfico (por ejemplo, la división de la República Mexicana en las 32 entidades federativas; la de una entidad en sus diferentes municipios; la de un municipio en localidades; la

de una delegación política del D.F. en colonias pudiese ser directamente en manzanas o colonias).

Etapas de muestreo

El procedimiento de selección se asocia a la manera cómo se elegirán las unidades muestrales de la población objetivo. Si este puede definir para todas y cada una de las unidades muestrales la probabilidad de selección y esta es distinta de cero entonces el muestreo es probabilístico. Si no es así, entonces el muestreo es no probabilístico.

Las ventajas más importantes del muestreo probabilístico radican en la capacidad de inferencia estadística que tienen los resultados muestrales. Por lo general, debido a que se requiere un marco de muestreo, acarrearán más costos y necesitan más tiempo de preparación que las encuestas basadas en el muestreo no probabilístico. En este sentido, se gana en la expansión de resultados de la muestra a la población y se “pierde” en costo.

La determinación del tipo de muestreo por realizar define también la pertinencia de la inclusión de conceptos y criterios que buscan garantizar que la muestra sea “apropiada”. Para que este concepto entre en funciones la teoría del muestreo ha incorporado los siguientes términos: (1) confiabilidad; (2) precisión; (3) replicabilidad; (4) reproducibilidad y (5) validez.

Los tres primeros son propios del muestreo probabilístico tanto por la forma como se definen como por los supuestos que requieren.

Nivel de confiabilidad

La confiabilidad se define en términos de un juicio probabilístico y se relaciona a la probabilidad de que un estimador $\hat{\theta}$ generado por la muestra esté "próximo" a un valor θ (el valor real). Es decir, $\Pr[|\theta - \hat{\theta}| \leq d] = 1 - \varepsilon$

"La probabilidad de que (a lo más) $\hat{\theta}$ difiera de θ en d es de $1-\varepsilon$ ". A $1-\varepsilon$ se le llama nivel de confiabilidad o nivel de confianza. Este juicio no se puede completar si no se sabe la probabilidad de selección de cada unidad de muestreo puesto que, si no es así, la probabilidad del estimador $\hat{\theta}$ (una variable aleatoria) no se podría especificar.

Precisión

La precisión y la confiabilidad están íntimamente ligadas aunque en un principio se partirá de una descripción de la primera en forma separada.

Supóngase que en un juego de tiro al blanco con dardos se presentan tres distintos participantes. El *primero*, *A*, exhibe una pericia notable y coloca todos sus tiros al centro o en sus proximidades. El *segundo*, *B*, también lanza sus tiros de tal forma que los "pega" en una misma área aunque alejados del centro. El *tercero*, *C*, es un jugador más bien errático: colocó sus tiros a lo largo y ancho del tablero con gran variabilidad.

Un observador estadístico comenta que *A* y *B* parecen ser "igualmente precisos" porque la probabilidad de sus lanzamientos es la misma: tienen igual concentración. Se le comenta que no puede ser esto así puesto que *A* sí se acerca al centro y *B* no lo hace. El observador replica que aunque esto es cierto para efecto de la precisión no cuenta. Los dos tiradores en condiciones semejantes son igualmente *predecibles*.

Esto es igual, en la teoría del muestreo, a los estimadores, que tienen una variabilidad igual puesto que el "valor verdadero" (salvo contadas excepciones) no se conoce. En ese sentido se diría que el tirador *A* es *exacto* y el tirador *B* es *preciso* y que para la teoría muestral los dos términos son coincidentes.

Por su parte el tirador *C* es el menos preciso de los tres por la gran dispersión de sus tiradas. Es, también, el menos predecible de los tres. Si este fuese equivalente

a un estimador los resultados que este generaría producirían "sorpresas" notables por la variabilidad y dispersión que exhibe. No son estos, en consecuencia, los "estimadores" que se recomienda usar.

La precisión de un estimador, por lo tanto, está directamente asociada a su variabilidad (medida por la varianza). La relación que existe entre precisión y varianza es inversamente proporcional. A medida que esta última es mayor la precisión es menor y viceversa.

Ahora bien, también la precisión está asociada, como condición de diseño, al llamado "margen de error" de la estimación. Este "margen de error" puede ser absoluto como es el caso de la estimación de proporciones o porcentajes o relativo como en el caso de totales, diferencias, y otros valores poblacionales continuos.

Brevemente se ilustra el margen de error en términos de estimaciones de proporciones. En encuestas electorales (de intención de voto) o en los Conteos Rápidos o en las Encuestas de Salida de Casilla interesa conocer con cierta exigencia si la proporción de votantes por distintos candidatos permite aseverar (dada una muestra) si hay una ventaja significativa de uno de ellos o si la muestra no puede mostrar confiablemente resultados.

En ese sentido, por ejemplo en una elección estos dos candidatos importantes, se busca estimar la proporción P que está en favor del candidato A (Con $Q=1-P$, la proporción que está en favor del candidato B). Para tal efecto se propone usar un estimador $\hat{P} = p$ que a lo más tenga un margen de error de ± 0.02 , esto es, dos puntos de error porcentual hacia arriba o hacia abajo, y con un nivel de confianza de 95%. De acuerdo con la muestra obtenida el estimador p se calcula en 0.485. Esto indicaría que (si el margen de error es de ± 0.02) la proporción P de votantes en la población que se inclina por A está entre 0.465 y 0.505 con un 95% de confianza (una probabilidad de 0.95).

Concepto de representatividad

Desde el punto de vista estadístico ortodoxo el concepto de representatividad no existe porque no tiene sentido o significado operativo.

En la práctica -y por influencia de los especialistas en los sondeos no probabilísticos- el concepto de representatividad se ha hecho equivalente al de "Copia Fidedigna" de la Población.

Este concepto es, por lo tanto, elusivo puesto que depende del criterio o juicio con el que se evalúe que tan "fidedigna" es la copia. Requiere precisarse.

Es por ello que toda muestra debe ser analizada según diferentes criterios objetivos que, en conjunto, pueden constituir una base repetible y replicable de examen.

Cuando la encuesta es no probabilística se considera que la representatividad sólo se refiere a los resultados. En ocasiones se llega a afirmar que al ser los resultados obtenidos semejantes a otras encuestas o a otras fuentes de información la encuesta en cuestión es "representativa".

Como al afirmar que determinada encuesta es "no representativa" (juicios de valor disfrazado de evaluación teórica) implica que la encuesta "no es buena" o que "no sirve" se entra en otro terreno no deseable que es preferible evadir dejando que las encuestas puedan ser evaluadas.

Toda encuesta tiene alcances y limitaciones con respecto a la validez, consistencia y confiabilidad de sus resultados. Los resultados que genera una encuesta son examinados a la luz de la confiabilidad. Este concepto, se desprende de un juicio probabilístico: "la probabilidad de que un estimador esté cercano al valor real estimado". A medida de que dicha probabilidad se aproxime a

uno, los resultados son confiables. Si, en cambio, la probabilidad respectiva es “baja” los resultados no son confiables.

Como este juicio probabilístico no tiene sentido en encuesta probabilísticas, se requiere operacionalizar cuando ya se tienen los resultados del levantamiento de campo. La confiabilidad ex post se evalúa mediante el concepto de reproducibilidad o variabilidad interna. Dicho concepto es aplicable a los dos grandes tipos de encuestas y tiene un desarrollo sistemático a partir de fechas muy reciente.

Es sentido estricto, el concepto de reproducibilidad está más vinculado con la precisión que con la confiabilidad aunque se pueda conectar con la última. La teoría sobre reproducibilidad partió de las encuestas probabilísticas y de ahí cruzó al campo no probabilístico.

Existen varios métodos para evaluar la reproducibilidad. Entre ellos está el método de “grupos aleatorios” o “replicaciones”. Este es el caso que se describe a continuación.

El método consiste en: (a) seleccionar dos o más submuestras aleatorias de la misma muestra empleando el mismo diseño muestral en cada caso; (b) construir un estimador apropiado y calcularlo separadamente de cada submuestras y de la combinación de todas las submuestras; (c) calcular la varianza muestra; (d) determinar el margen de error; y (e) si es posible, obtener el intervalo de confianza.

Existen dos variantes del método: (i) cuando las submuestras son independientes y (ii) cuando no lo son. El caso más simple es el primero puesto que bastaría aplicar un muestreo con reemplazo para asegurar la independencia estadística.

Fuentes de errores

Como se ha advertido anteriormente, en las encuestas existen dos grandes tipos de errores: los errores muestrales (propios de los procedimientos de selección y estimación) y los errores no muestrales (propios del proceso de observación).

Los primeros corresponden a la asignación equivocada de probabilidades de selección o a que las reglas de estimación son erróneas. Este tipo de errores pueden evaluarse propiamente en las encuestas probabilísticas. En las no probabilísticas se esconden u ocultan.

Tienen un peso o contribución entre 10 y 25% con respecto al error total. Kish, hablando de la práctica en EEUU opta por el 10%, la práctica mexicana se acerca al 25%.

Los errores no muestrales tienen un peso mayor y pocas veces se reportan en forma explícita. Entre los errores no muestrales se tienen los siguientes:

- a) errores producidos por los instrumentos
- b) errores generados por la aplicación
- c) errores de codificación y captura
- d) errores de procesamiento
- e) errores de análisis

Cada error no muestral tiene fuentes específicas de las que se genera:

- (a) los errores producidos por los instrumentos se derivan del diseño de éstos y de los manuales o guías de aplicación
- (b) Los errores generados en la aplicación surgen en el modo o forma como los cuestionarios se aplican así como en la forma como se reporta la información y en el control del envío, entrega y recepción.
- (c) Los errores de codificación y captura se derivan en muchas ocasiones desde el diseño conceptual, en la revisión de campo, en la validación y en la estructura y contenido de las bases de datos.

- (d) Los errores de procesamiento suceden tanto por el mal uso de paqueterías comerciales como por la aplicación de programación específica.
- (e) Los errores de análisis estadístico (cuando se aplican técnicas específicas a conjuntos de datos que no satisfacen los supuestos básicos). Sería el caso de aplicar técnicas estadísticas que exigen un muestreo irrestricto aleatorio a datos producidos por un diseño muestral diferente.

Si bien estos errores no son los únicos sí son los principales. Tienen remedio ex ante y solución posterior prefiriendo siempre lo primero a lo segundo.

Errores y sesgos de muestreo

En cuanto a los errores de muestreo estos se pueden especificar en forma panorámica como sigue:

Errores de selección

- a) Cuando el procedimiento de selección (físico o en la computadora) difiere de la definición del procedimiento señalado en forma lógico-matemática. Este es el caso cuando se define un procedimiento de selección de tipo *ppt* (con muestreo proporcional al tamaño) y la selección se lleva a cabo con probabilidades iguales en cada etapa de muestreo.
- b) Cuando las probabilidades de selección son desconocidas y se "ajustan" a un esquema distinto.

Errores muestrales en los procedimientos de estimación

El error muestral corresponde a la variación entre el valor muestral y el valor poblacional. Esta variación o discrepancia existe siempre. La razón fundamental se deriva de la diferencia entre los contenidos de información de la población y de la muestra.

El error muestral en las encuestas probabilísticas es medible. Esto indica que la diferencia entre el estimador $\hat{\theta}$ y el valor poblacional θ (desconocido) puede

aproximarse probabilísticamente.

Un juicio del tipo $\Pr[|\hat{\theta} - \theta| \leq e] = 1 - \alpha$ es viable porque se aceptan hipótesis acerca de la distribución de $\hat{\theta}$ que es una referencia de la regla de estimación empleada. El intervalo $|\hat{\theta} - \theta| \leq e$ es equivalente a: $\hat{\theta} - e \leq \theta \leq \hat{\theta} + e$ por lo que el valor θ que es desconocido se aproxima por $\hat{\theta} - e \leq \theta \leq \hat{\theta} + e$ en “ $(1 - \alpha) \cdot 100$ veces por cada 100”.

El error muestral se puede ilustrar también de la siguiente forma: si \bar{y} es estimador de Y en un muestreo simple aleatorio, ya se ha comentado que

$$\text{Var}(\bar{y}) = \frac{(1-f)S^2}{n}, \quad \text{con } f = \frac{n}{N}$$

La varianza refleja el error producido por tomar un estimador y no el valor poblacional. Si en vez de una muestra de tamaño n se optase por la enumeración completa o censo, y pasaría a ser Y y $\text{Var}(\bar{y}) = \text{Var}(\bar{Y}) = \left(1 - \frac{N}{N}\right) \frac{S^2}{N} = (1-1) \frac{S^2}{N} = 0$

Por lo tanto, no se tendría error muestral.

Cuando se opera en un censo -por enumeración completa- no deja de haber no respuesta (no observación) y, en consecuencia, está oculto un error pero este se atribuye a los errores de observación (es decir, a los errores no muestrales).

Sesgos en los Procedimientos de Estimación

Cuando por construcción de $\hat{\theta}$ el intervalo $\hat{\theta} - e, \hat{\theta} + e$ no rodea simétricamente a θ se confirma que existe un sesgo.

Esto ocurre cuando el estimador no es insesgado (la esperanza matemática no coincide con el valor poblacional). Esto ocurre con frecuencia por falta de elaboración técnica en diseños no autoponderados y con estimadores de razón (del tipo $R=Y/X$)

En las encuestas probabilísticas el sesgo muestral es también medible. El efecto de sesgo muestral está determinado por $B(\hat{\theta}) = E(\hat{\theta}) - \theta$. En muchos casos a medida que el tamaño de la muestra se incrementa el sesgo se reduce sustancialmente. En otros casos no es así.

En el ejemplo del tirador preciso que (casi) siempre tira sin variación pero alejado del centro del "tiro al blanco", el sesgo no se reduciría aunque aumentara el número de sus lanzamientos si no corrige ese error. Es decir, si no aprende.

Teóricamente es posible ajustar estimadores sesgados para reducir el sesgo. Esto en la práctica implica complejidad adicional en el procesamiento y no siempre es recomendable.

3.2 Diseño utilizado en Oportunidades

El Programa de Desarrollo Humano Oportunidades, hasta el primer semestre de 2004, ocupaba el siguiente diseño muestral para la aplicación de sus encuestas de seguimiento operativo a familias beneficiarias, llamada puntos centinela. A este diseño se le atribuían las siguientes características¹:

1. Se considera un muestreo simple aleatorio
2. Nivel de confianza del 95%
3. Se fija una precisión del 2.5%
4. La muestra se reparte proporcionalmente por entidad federativa de acuerdo a la cantidad de familias beneficiarias
5. La selección se hace de forma aleatoria por entidad federativa

¹ El anexo se conforma por tres "notas técnicas" utilizadas por Oportunidades para el cálculo del tamaño de muestra, se incluyen tres notas porque son complementarias, de la 2004 se toma el ejemplo, y de las anteriores se obtiene la forma de estimación y detalles de la metodología empleada para el cálculo de la muestra.

Marco muestral

El marco de muestreo de la encuesta de puntos centinela es el listado de localidades beneficiarias del Programa un semestre antes del diseño, por ejemplo, para el bimestre julio-agosto se toma el listado de localidades beneficiarias del Programa en enero-febrero del mismo año, que cuentan con cinco o más hogares beneficiarios. Por ejemplo, 70,283² localidades para el bimestre julio-agosto de 2004.

Esquema de muestreo

En la nota metodológica de la encuesta de puntos centinela se dice que la muestra fue diseñada con un esquema de muestreo simple aleatorio.

Tamaño de la muestra

El tamaño de muestra está calculado para dar estimaciones a nivel nacional y por entidad federativa.

La expresión empleada para el cálculo del tamaño de muestra es la siguiente:

$$n' = pq \left(\frac{Z}{d} \right)^2 \quad (6)$$

donde:

p = Proporción en que aparece la característica a estudiar

q = Proporción de no aparición

Z = Valor correspondiente al nivel de confianza

d = Precisión deseada del estimador

De acuerdo al tamaño de la población se aplica la siguiente fórmula:

$$n_1 = \frac{n'}{\left(1 + \frac{n'}{N} \right)} \quad (7)$$

² Dato tomado de la "Nota técnica" para la Muestra de Localidades para Puntos Centinela 2004 que se solicitó a través del Sistema de Solicitudes de Información (SISI) mediante número de folio 2000100010207, de fecha 02 de febrero de 2007

donde:

n' = Tamaño de muestra preliminar

N = Tamaño de la población

Si se considera la no respuesta se aplica la siguiente fórmula:

$$n = \frac{n_1}{1 - TNR} \quad (8)$$

donde:

n_1 = Tamaño de muestra preliminar

TNR = Tasa anticipada de no respuesta

Para ilustrar las fórmulas anteriores tomaremos como ejemplo el cálculo del tamaño de muestra para una población de 10,000 elementos, si sabemos que la tasa anticipada de respuesta es del 90%, se requiere como máximo un margen de error del 5%, un nivel de confianza del 95% y para obtener el mayor tamaño de muestra consideramos $p=0.5$, entonces sustituyendo valores en la fórmula (1) tenemos:

$$\begin{aligned} n' &= pq \left(\frac{Z}{d} \right)^2 = 0.5 * (1 - 0.5) \left(\frac{1.96}{0.05} \right)^2 \\ &= 0.25(39.2)^2 = 0.25 * 1536.64 = 384.16 \end{aligned} \quad (9)$$

Sustituyendo (4) en (2) tenemos:

$$n_1 = \frac{n'}{\left(1 + \frac{n'}{N}\right)} = \frac{384.16}{\left(1 + \frac{384.16}{10000}\right)} = \frac{384.16}{(1 + 0.038416)} = 369.95 \quad (10)$$

Sustituyendo (5) en (3) nos queda el tamaño de muestra

$$n = \frac{n_1}{1 - TNR} = \frac{369.95}{1 - 0.1} = \frac{369.95}{0.9} = 411.05 \cong 412$$

Afijación de la muestra

Para fijar el tamaño de muestra se tomaron las siguientes consideraciones:

- Se consideró un nivel de confianza del 95% con una precisión del 2.5% y una tasa anticipada de no respuesta del 10%

- Aplicando las fórmulas antes mencionadas tenemos una muestra de 1,671 localidades a nivel nacional. El cálculo que se realizó fue el siguiente. Sustituyendo valores en la fórmula (1) tenemos:

$$n' = pq \left(\frac{Z}{d} \right)^2 = 0.5 * (1 - 0.5) \left(\frac{1.96}{0.025} \right)^2$$

$$= 0.25(78.4)^2 = 0.25 * 6146.56 = 1536.64 \quad (11)$$

Sustituyendo (6) en (2) tenemos:

$$n_1 = \frac{n'}{\left(1 + \frac{n'}{N}\right)} = \frac{1536.64}{\left(1 + \frac{1536.64}{70283}\right)} = \frac{1536.64}{(1 + 0.02186361)} = 1503.76 \quad (12)$$

Sustituyendo (7) en (3) nos queda el tamaño de muestra

$$n = \frac{n_1}{1 - TNR} = \frac{1503.76}{1 - 0.1} = \frac{1503.76}{0.9} = 1670.85 \cong 1671$$

- La muestra se repartió proporcionalmente en las entidades federativas de acuerdo a la cantidad de familias beneficiarias de acuerdo a la siguiente fórmula:

Entero más próximo a: $n_i = \frac{nF_i}{FT}$, donde F_i es el número de familias beneficiarias en la entidad federativa i y FT es el número total de familias beneficiarias, n_i es el tamaño de muestra en la entidad federativa i y n es el tamaño de muestra total calculado por la fórmula (3).

Por ejemplo: para calcular el número de familias en muestra para Aguascalientes, dado que el número de familias beneficiarias en esa entidad federativa es de 14,726 y aplicando la fórmula anterior tenemos:

$$n_i = \frac{nF_i}{FT} = \frac{1671 * 14726}{4173741} = \frac{24607146}{4173741} = 5.9 \approx 6$$

En el cuadro siguiente se presenta la distribución proporcional de familias en muestra por entidad federativa.

**Cuadro 6. Distribución de localidades en muestra
(proporcional a familias beneficiarias por Entidad Federativa)**

Entidad Federativa	Total de localidades	Familias beneficiarias	Localidades en muestra proporcional al número de familias beneficiarias
1 Aguascalientes	355	14,726	6
2 Baja California	230	7,438	3
3 Baja California Sur	175	5,802	2
4 Campeche	505	47,118	19
5 Coahuila	825	34,360	14
6 Colima	170	12,404	5
7 Chiapas	6,087	491,384	197
8 Chihuahua	1,666	39,749	16
10 Durango	1,860	57,085	23
11 Guanajuato	4,178	189,341	76
12 Guerrero	4,085	269,614	108
13 Hidalgo	3,405	175,535	70
14 Jalisco	4,039	76,167	30
15 México	3,017	278,823	112
16 Michoacán	4,877	261,249	105
17 Morelos	624	68,228	27
18 Nayarit	865	43,483	17
19 Nuevo León	1,130	24,781	10
20 Oaxaca	6,230	397,051	159
21 Puebla	3,739	319,205	128
22 Querétaro	1,276	67,500	27
23 Quintana Roo	333	40,819	16
24 San Luis Potosí	3,617	150,348	60
25 Sinaloa	2,402	101,679	41
26 Sonora	888	54,574	22
27 Tabasco	1,414	120,982	48
28 Tamaulipas	1,555	69,543	28
29 Tlaxcala	400	35,537	14
30 Veracruz	7,434	490,061	196
31 Yucatán	753	116,225	47
32 Zacatecas	2,149	112,930	45
TOTAL	70,283	4,173,741	1,671

Fuente: Cálculos propios tomando como referencia la "Nota técnica" para la Muestra de Localidades para Puntos Centinela 2004 que se solicitó a través del Sistema de Solicitudes de Información (SISI) mediante número de folio 2000100010207, de fecha 02 de febrero de 2007

Como se requería dar estimaciones a nivel estatal y se observó que con este tamaño de muestra para algunas entidades federativas la muestra no era suficiente, se procedió a obtener el tamaño de muestra por el siguiente método:

- Se clasifican las entidades federativas en tres grupos de acuerdo a la cantidad de familias beneficiarias y se asignó un tamaño de muestra por grupo. Los grupos se clasificaron de la siguiente forma:

Grupo	Hogares beneficiarios en la Entidad Federativa	Tamaño de muestra fijado (localidades)
1	Menos de 45 mil	40
2	De 45 mil a 149,999	100
3	Con 150 mil o más	200

- Se consideró para el tamaño de muestra definitivo por entidad federativa el mínimo entre el tamaño determinado por el grupo al que pertenece y el 20% del total de localidades en la entidad federativa.
- Se determinó entrevistar a cinco familias por localidad en muestra.

Esto dio como resultado una muestra de 3,489 localidades, lo que representa el 4.96% del marco muestral y se distribuye de la siguiente forma:

Cuadro 7. Distribución de localidades a encuestar

Entidad Federativa	Total de localidades	Familias	Localidades en muestra proporcional	20% de las localidades	Muestra final de localidades	Muestra final de familias beneficiarias	Grupo
1 Aguascalientes	355	14,726	6	71	40	200	1
2 Baja California	230	7,438	3	46	40	200	1
3 Baja California Sur	175	5,802	2	35	35	175	1
5 Coahuila	825	34,360	14	165	40	200	1
6 Colima	170	12,404	5	34	34	170	1
8 Chihuahua	1,666	39,749	16	333	40	200	1
18 Nayarit	865	43,483	17	173	40	200	1
19 Nuevo León	1,130	24,781	10	226	40	200	1
23 Quintana Roo	333	40,819	16	67	40	200	1
29 Tlaxcala	400	35,537	14	80	40	200	1
4 Campeche	505	47,118	19	101	100	500	2
10 Durango	1,860	57,085	23	372	100	500	2
14 Jalisco	4,039	76,167	30	808	100	500	2
17 Morelos	624	68,228	27	125	100	500	2
22 Querétaro	1,276	67,500	27	255	100	500	2

Entidad Federativa	Total de localidades	Familias	Localidades en muestra proporcional	20% de las localidades	Muestra final de localidades	Muestra final de familias beneficiarias	Grupo
25 Sinaloa	2,402	101,679	41	480	100	500	2
26 Sonora	888	54,574	22	178	100	500	2
27 Tabasco	1,414	120,982	48	283	100	500	2
28 Tamaulipas	1,555	69,543	28	311	100	500	2
31 Yucatán	753	116,225	47	151	100	500	2
32 Zacatecas	2,149	112,930	45	430	100	500	2
7 Chiapas	6,087	491,384	197	1,217	200	1,000	3
11 Guanajuato	4,178	189,341	76	836	200	1,000	3
12 Guerrero	4,085	269,614	108	817	200	1,000	3
13 Hidalgo	3,405	175,535	70	681	200	1,000	3
15 México	3,017	278,823	112	603	200	1,000	3
16 Michoacán	4,877	261,249	105	975	200	1,000	3
20 Oaxaca	6,230	397,051	159	1,246	200	1,000	3
21 Puebla	3,739	319,205	128	748	200	1,000	3
24 San Luis Potosí	3,617	150,348	60	723	200	1,000	3
30 Veracruz	7,434	490,061	196	1,487	200	1,000	3
TOTAL	70,283	4,173,741	1,671	14,057	3,489	17,445	

Fuente: Cálculos propios tomando como referencia la "Nota técnica" para la Muestra de Localidades para Puntos Centinela 2004 que se solicitó a través del Sistema de Solicitudes de Información (SISI) mediante número de folio 2000100010207, de fecha 02 de febrero de 2007.

Selección de la muestra

Se eligieron las localidades de la muestra por estado y por tipo de localidad de manera aleatoria.

Estimación de resultados

Debido que la selección de localidades dentro de cada estado se realizó de forma aleatoria, la muestra es autoponderada³ a nivel estatal.

Los inconvenientes de este diseño

En este apartado se tomarán las distintas fases del esquema de muestreo utilizado para describir los inconvenientes que presenta dicho diseño.

³ Para aclarar este término ver capítulo 2.

Esquema de muestreo

En la nota metodológica de la encuesta de Puntos Centinela se dice que la muestra fue diseñada con un esquema de muestreo simple aleatorio; sin embargo, no es un muestreo simple aleatorio, ya que como también lo señala la misma nota, previo a seleccionar a las familias se hace una selección de localidades, es decir, se están realizando dos etapas de selección, en la que en la primera etapa se selecciona a las localidades y en la segunda a las familias beneficiarias, por lo cual es un muestreo por conglomerados en dos etapas. Por otro lado, si el muestreo fuera un msa, se tendría que hacer el cálculo del tamaño de muestra tomando en cuenta el número de familias beneficiarias en el periodo tratado y no al número de localidades que cumplen con las características, es decir a nivel nacional con los parámetros utilizados en el diseño se tendrían 1,671 familias y no localidades.

Tamaño y afijación de la muestra

Como se señaló en párrafos anteriores, para calcular el tamaño de muestra el diseño se basa en el número de localidades y no en el número de familias beneficiarias, que es la última unidad de muestreo debido a que el objetivo de la encuesta es conocer algunas percepciones y características que tienen las familias beneficiarias y no las localidades.

Por otra parte, el tamaño de muestra no se calculó conforme las formulas descritas, debido a que, como ya se había señalado, si se reparte proporcionalmente, para algunos Estados la muestra no es suficiente aún con parámetros con amplios márgenes de error y poco nivel de confianza esto es debido a que el cálculo se está realizando tomando en cuenta a las localidades y no a las familias. Por ejemplo, como puede notarse en el cuadro anterior, para Baja California y Baja California Sur, sólo se encuestaría a familias beneficiarias pertenecientes a 3 y 2 localidades respectivamente, es decir a 15 y 10 familias, lo cual es insuficiente para dar estimaciones.

Después de afirmar que es un muestreo simple aleatorio, en la nota metodológica, y dar las fórmulas para el cálculo del tamaño de la muestra, se recurre al muestreo estratificado, pero no hay alguna justificación estadística para el tamaño muestral en cada estrato ya que divide a los Estados en tres grupos de acuerdo al número de familias beneficiarias y se les asigna un tamaño de muestra a cada grupo.

Lo anterior implica una contradicción en la nota metodológica debido a que al inicio se afirma que para fijar el tamaño de muestra se utilizó un diseño simple aleatorio y después afirma que se tomó el mínimo entre el tamaño determinado para cada grupo y el 20% del total de localidades, es decir realmente se está recurriendo a un muestreo estratificado.

En el muestreo probabilístico cuando se desea dar estimaciones por Estado, comúnmente se debe realizar un muestreo estratificado, en este caso cada Estado es un estrato y si se tuviera un muestreo simple aleatorio en cada uno de ellos con los parámetros antes señalados, a pesar de que el muestreo planteado es estratificado, para algunos Estados el tamaño de muestra no es suficiente para dar estimaciones a nivel estatal como veremos en los siguientes ejemplos. Tomemos a un estado por grupo: Aguascalientes, Campeche y Chiapas.

Cuadro 8. Ejemplo de la distribución muestral

Entidad Federativa	Total de locs.	Familias	Mtra. Locs. Propor.	20% de las locs.	Mtra. Locs. Final	Mtra. Fams. Final	Grupo
1 Aguascalientes	355	14,726	6	71	40	200	1
4 Campeche	505	47,118	19	101	100	500	2
7 Chiapas	6,087	491,384	197	1,217	200	1,000	3

Fuente: Extracto del cuadro 7.

Ahora calcularemos el tamaño de muestra con un muestreo simple aleatorio, nivel de confianza del 95%, margen de error máximo de 2.5%, una proporción para la característica buscada de 0.5 y una tasa de no respuesta del 10%, entonces tenemos:

Para Aguascalientes:

$$n' = pq \left(\frac{Z}{d} \right)^2 = 0.5 * (1 - 0.5) \left(\frac{1.96}{0.025} \right)^2$$

$$n' = pq \left(\frac{Z}{d} \right)^2 = 0.25(78.4)^2 = 0.25 * 6146.56 = 1536.64$$

$$n_1 = \frac{n'}{\left(1 + \frac{n'}{N} \right)} = \frac{1536.64}{\left(1 + \frac{1536.64}{14726} \right)} = \frac{1536.64}{(1 + 0.10434877)} = 1391.44$$

$$n = \frac{n_1}{1 - TNR} = \frac{1391.44}{1 - 0.1} = \frac{1391.44}{0.9} = 1546.05 \cong 1546 \quad (13)$$

Para Campeche:

$$n' = pq \left(\frac{Z}{d} \right)^2 = 0.5 * (1 - 0.5) \left(\frac{1.96}{0.025} \right)^2$$

$$n' = pq \left(\frac{Z}{d} \right)^2 = 0.25(78.4)^2 = 0.25 * 6146.56 = 1536.64$$

$$n_1 = \frac{n'}{\left(1 + \frac{n'}{N} \right)} = \frac{1536.64}{\left(1 + \frac{1536.64}{47118} \right)} = \frac{1536.64}{(1 + 0.03261259)} = 1488.11$$

$$n = \frac{n_1}{1 - TNR} = \frac{1488.11}{1 - 0.1} = \frac{1488.11}{0.9} = 1653.45 \cong 1653 \quad (14)$$

Para Chiapas:

$$n' = pq \left(\frac{Z}{d} \right)^2 = 0.5 * (1 - 0.5) \left(\frac{1.96}{0.025} \right)^2$$

$$n' = pq \left(\frac{Z}{d} \right)^2 = 0.25(78.4)^2 = 0.25 * 6146.56 = 1536.64$$

$$n_1 = \frac{n'}{\left(1 + \frac{n'}{N} \right)} = \frac{1536.64}{\left(1 + \frac{1536.64}{491384} \right)} = \frac{1536.64}{(1 + 0.00312717)} = 1531.85$$

$$n = \frac{n_1}{1-TNR} = \frac{1531.85}{1-0.1} = \frac{1531.85}{0.9} = 1702.06 \cong 1702 \quad (15)$$

Como se observa en los tres ejemplos anteriores, para ninguno de los tres grupos se cumple con el tamaño mínimo de muestra para hacer inferencias con los parámetros dados, ya que el máximo tamaño muestral con lo descrito es de 1000 familias, mientras que aplicando la fórmula para un msa para cualquiera de los tres grupos es mayor, lo cual implica que la muestra no es “representativa” a nivel estatal con las características especificadas, lo puede ser pero con distintas características, por ejemplo mayor margen de error.

Selección de la muestra.

En la nota metodológica se dice, que se eligen a las localidades de manera aleatoria por entidad federativa y tipo de localidad. Aquí no se especifica a qué se refiere con tipo de localidad ni cómo se selecciona a las familias que entrarán en muestra.

Estimación de resultados.

Otra inconsistencia es el decir que es un muestreo autoponderado⁴ a nivel estatal por ser una muestra aleatoria lo cual es erróneo ya que el muestreo con un diseño autoponderado, cada familia representa al mismo número de familias en la población y para que cada familia de la muestra represente al mismo número de familias de la población, según lo apuntado en el capítulo dos de este documento, debería cumplir con:

- a) el tamaño de familias en muestra debe ser proporcional al tamaño de familias en la localidad, así en localidades grandes se extraen más familias para la muestra que en las localidades pequeñas; o
- b) la probabilidad de seleccionar a la localidad debe ser proporcional al número de familias en ésta; o
- c) El diseño debe ser verdaderamente simple aleatorio a nivel estatal

⁴ Para aclarar este término ver capítulo 2

Por los inconvenientes señalados en este subapartado, se propone en este documento realizar un diseño muestral distinto al ya expuesto y vigente, pues con el diseño empleado no se pueden realizar inferencias con las características especificadas, ya que los resultados no se están obteniendo de forma adecuada y por lo tanto la información se podría mal interpretar y podría llevar a toma de decisiones erróneas lo que conduciría a no mejorar la calidad de los servicios o a una mala percepción del Programa por las familias beneficiarias y población en general.

3.3 Diseño sugerido

El diseño de la muestra sugerido se caracteriza por ser probabilístico, por lo cual los resultados obtenidos de la encuesta se generalizan a toda la población. A su vez es bietápico, estratificado y por conglomerados, donde la unidad primaria de selección es la localidad y la segunda y última unidad de selección es una titular (familia) beneficiaria de Oportunidades. Cada Estado es un subconjunto de la población sobre los que también hay que dar estimaciones, por lo tanto es un dominio de estudio independiente (estrato).

Marco muestral

El marco muestral sugerido para la encuesta de puntos centinela es el listado de familias activas⁵ dos bimestres anteriores al que se realizará el cálculo de tamaño de muestra y que vivan en localidades con cinco o más familias beneficiarias. Pondremos el mismo ejemplo que el señalado en el apartado anterior, donde se

⁵ Se dice que son familias activas a un bimestre cuando la familia está incorporada al programa y ha cumplido con sus corresponsabilidades de dicho bimestre para poder recibir sus apoyos.

explica el diseño utilizado, esto es, 70,283⁶ localidades para el bimestre del ejemplo señalado.

Formación de unidades de muestreo

Unidad Primaria de Muestreo (UPM).

Las unidades primarias de muestreo están constituidas por cada una de las localidades en las que viven por lo menos cinco familias beneficiarias del Programa y activas al bimestre en que se están tomando los datos.

Unidad Secundaria de Muestreo (USM)

Las unidades secundarias y últimas de muestreo están constituidas por cada una de las familias beneficiarias.

Estratificación

Tomando como referencia que se desea una muestra representativa a nivel nacional y estatal, a nivel nacional se sometió a una estratificación geográfica utilizando las entidades federativas como estratos.

Esquema de muestreo

Se sugiere que la encuesta de puntos centinela sea diseñada con un esquema de muestreo probabilístico, estratificado y bietápico por conglomerados.

- Probabilístico. Esto significa que las unidades de selección tienen una probabilidad conocida y distinta de cero de entrar en la muestra para cada miembro de la población, lo cual permite conocer la precisión de los resultados muestrales.
- Estratificado. Porque las unidades de selección son agrupadas por características de tipo geográficas
- Bietápico. Porque la unidad última de selección (familia) es seleccionada después de una etapa.

⁶ Dato tomado de la "Nota técnica" para la Muestra de Localidades para Puntos Centinela 2004 que se solicitó a través del Sistema de Solicitudes de Información (SISI) mediante número de folio 2000100010207, de fecha 02 de febrero de 2007

- Por conglomerados. Porque previamente se conforman conjuntos de unidades muestrales de los cuales se obtiene la muestra.

Tamaño de la muestra

El tamaño de muestra está calculado para dar estimaciones en los siguientes niveles de desagregación⁷:

- Nacional
- Entidad federativa

El tamaño de muestra para estos dominios se calcula para las siguientes tres variables:

- Conocimiento del número de inasistencias a la unidad médica para dar de baja a una familia
- Conocimiento del número de faltas a la escuela para suspender la beca a un becario
- Buena participación de las vocales

Consideradas las principales variable en esta encuesta, tomando la proporción que diera mayor tamaño de muestra dichas variables lo que garantiza que las estimaciones de todas las variables de interés queden cubiertas con ese tamaño.

La expresión empleada para el cálculo es independiente para cada entidad-estrato y es la siguiente:

$$n = \frac{z^2 pq N DEFF}{(1 - TNR)(r^2 N + z^2 pq)} \quad (16)$$

donde:

n = Tamaño de la muestra

p = Estimación de la proporción

$q = (1-p)$

r = Error relativo máximo aceptable

⁷ Se consideró esta desagregación debido a que se desea dar estimaciones referentes al total nacional y a nivel estatal

z = Valor asentado en las tablas estadísticas que garantiza realizar las estimaciones con una confianza prefijada

$DEFF$ = Cociente de la varianza del diseño utilizado entre la varianza del muestreo aleatorio simple para un mismo tamaño de muestra

TNR = Tasa de no respuesta esperada

N = Tamaño poblacional (total de familias del estrato)

Tomando como base, para cada estrato, un nivel de confianza del 95% y un efecto de diseño de 1.44⁸ observado en las experiencias anteriores, así como un error relativo máximo esperado de 5% y una tasa de no respuesta anticipada del 10%, se determinó una muestra a nivel nacional de: 17,128 familias beneficiarias del programa.

Tomando en cuenta que se seleccionan a cinco titulares por localidad, el tamaño de muestra se ajustó por Estado, quedando una muestra total de 17,130 familias beneficiarias y el de cada estrato se presenta en el cuadro 9.

Afijación de la muestra

Debido a que se desea realizar inferencias para cada estrato con la muestra dada, se realiza el cálculo de muestra para cada estrato por separado y así se determina el tamaño para cada uno de ellos utilizando la fórmula (16). Es por eso que la muestra total n es mucho más grande puesto que ahora se está haciendo inferencias por separado para 32 poblaciones.

⁸ Efecto de diseño, es la pérdida o ganancia en la eficiencia del diseño por efecto de estratificar y/o conglomerar elementos de la población para formar unidades muestrales. En muchos estudios se toma el efecto de experiencias pasadas en estudios similares. En este caso se utilizó el efecto de diseño derivado de la Encuesta del Censo de Población y Vivienda 1995 (INEGI).

Cuadro 9. Tamaños de Muestra por Estado según el diseño sugerido

	Estado	Localidades	Familias beneficiarias	Proporción de familias que manifiesta una determinada característica	Familias en muestra	Localidades en Muestra
1	Aguascalientes	355	14,726	0.71	495	99
2	Baja California	230	7,438	0.77	420	84
3	Baja California Sur	175	5,802	0.56	565	113
4	Campeche	505	47,118	0.37	570	114
5	Coahuila	825	34,360	0.56	600	120
6	Colima	170	12,404	0.73	470	94
7	Chiapas	6,087	491,384	0.61	585	117
8	Chihuahua	1,666	39,749	0.54	605	121
10	Durango	1,860	57,085	0.42	595	119
11	Guanajuato	4,178	189,341	0.56	605	121
12	Guerrero	4,085	269,614	0.56	605	121
13	Hidalgo	3,405	175,535	0.63	570	114
14	Jalisco	4,039	76,167	0.45	605	121
15	México	3,017	278,823	0.60	590	118
16	Michoacán	4,877	261,249	0.40	590	118
17	Morelos	624	68,228	0.75	460	92
18	Nayarit	865	43,483	0.79	405	81
19	Nuevo León	1,130	24,781	0.70	510	102
20	Oaxaca	6,230	397,051	0.69	525	105
21	Puebla	3,739	319,205	0.73	485	97
22	Querétaro	1,276	67,500	0.56	600	120
23	Quintana Roo	333	40,819	0.87	390	78
24	San Luis Potosí	3,617	150,348	0.54	610	122
25	Sinaloa	2,402	101,679	0.44	605	121
26	Sonora	888	54,574	0.60	585	117
27	Tabasco	1,414	120,982	0.42	595	119
28	Tamaulipas	1,555	69,543	0.38	575	115
29	Tlaxcala	400	35,537	0.63	565	113
30	Veracruz	7,434	490,061	0.60	590	118
31	Yucatán	753	116,225	0.65	555	111
32	Zacatecas	2,149	112,930	0.55	605	121
	Total	70,283	4,173,741		17,130	3,426

Fuente. Cálculos propios utilizando datos de cuadros anteriores y la fórmula (16)

Selección de la muestra

La selección de las unidades de análisis se realiza en dos etapas. En la primera se seleccionan localidades y en la segunda se selecciona directamente a la familia, donde se captan los datos de cada una de las unidades de interés.

A continuación se presenta la forma en que se construyen las unidades de selección correspondientes, así como el mecanismo a través del cual se incluyen en la muestra.

Unidades Primarias de Muestreo

En cada uno de los estratos-entidad federativa, se procedió a la construcción de las Unidades Primarias de Muestreo (UPM), que se integran por cada una de las localidades en el marco muestral.

La selección de las UPM se llevó a cabo de manera sistemática y cada una tiene la misma probabilidad de selección. Para llevar a cabo el procedimiento de selección se ordenó a las localidades (en esta disposición) tipo de área (rural, semiurbana y urbana); institución liquidadora (Telecomm, Bancomer, Bansefi abono a cuenta, Bansefi entrega directa y Bansefi caja de ahorro); institución de salud (SSA, IMSS Oportunidades e IMSS Régimen Obligatorio); y tamaño de la localidad (número de familias por localidad en forma ascendente).

Unidades Secundarias de Muestreo

Las familias que viven en cada una de las UPM seleccionadas, constituyen las Unidades Secundarias de Muestreo (USM). En cada una de las UPM se procedió a seleccionar a través de un msa a las USM (familias).

Por lo tanto, la probabilidad de seleccionar a una familia beneficiaria es:

$$p_{ki} = \frac{n_k}{N_k} \frac{m_i}{M_i}$$

donde:

N_k = Número total de localidades en el estrato k ;

n_k = Número de localidades en muestra en el estrato k ;

m_i = Número de familias beneficiarias en muestra de la localidad i ; y

M_i = Número total de familias beneficiarias en la localidad i .

Entonces, su factor de expansión será:

$$f_{ki} = \frac{N_k}{n_k} \frac{M_i}{m_i}$$

Estimadores

El estimador del total de la característica X a nivel estatal es:

$$\hat{X}_k = \sum_i \sum_l f_{ki} X_{kil}$$

F_{ki} = Factor de expansión de la i -ésima localidad en el k -ésimo estrato

X_{kil} = Valor observado de la característica de interés X en la l -ésima familia de la i -ésima localidad en el k -ésimo estrato

Mientras que a nivel nacional:

$$\hat{X}_{NAL} = \sum_k \hat{X}_k$$

Para la estimación de proporciones, tasas y promedios se utiliza el estimador de razón:

$$\hat{R}_{NAL} = \frac{\hat{X}_{NAL}}{\hat{Y}_{NAL}}$$

donde, la variable \hat{Y}_{NAL} es definida en forma similar a \hat{X}_{NAL} .

El esquema de muestreo descrito, es totalmente probabilístico y el tamaño de muestra está calculado para dar estimaciones a nivel estatal y nacional, ya que no es obtenido mediante cuotas o porcentajes de la población, esto hace que las estimaciones obtenidas con este tamaño de muestra sean válidas para las especificaciones técnicas establecidas desde el inicio del diseño siempre y cuando la tasa de recuperación de la muestra sea al menos del 90% del total de la muestra. Ahora se presenta un cuadro comparativo entre el diseño empleado y el diseño sugerido con el objeto de resumir las diferencias y contrastar las bondades de éste último en relación al primero.

Cuadro 10. Comparación de los diseños muestrales

Características	Diseño empleado	Diseño sugerido
Marco de la muestra	Listado de localidades con familias beneficiarias	Listado de familias beneficiarias. UPM son las localidades y las USM son las familias.
Esquema de muestreo	Se dice que es un msa pero en efecto no lo es por que se utilizan dos etapas de muestreo	Probabilístico, estratificado, bietápico por conglomerados
Tamaño de la muestra	$n = \frac{z^2 pq N}{(1 - TNR)(r^2 N + z^2 pq)}$ <p>n = 1671 localidades entonces quedarían 8355 familias en muestra. Fijado sin respaldo estadístico: n= 17445 familias, 3489 localidades Nivel de confianza: 95% Margen de error: 2.5% TNR: 10% Por estratos: Ags.: 200 familias Camp.: 500 familias Chis.: 1000 familias</p>	$n = \sum n_i$ $n_i = \frac{z^2 pq N DEFF}{(1 - TNR)(r^2 N + z^2 pq)}$ <p>n = 3426 localidades, 17130 familias en muestra. Nivel de confianza: 95% Margen de error: 5% TNR: 10% Deff: 1.44 Por estratos: Ags.: 495 familias Camp.: 570 familias Chis.: 585 familias</p>
Afijación de la muestra	Proporcional Fijado por el muestrista, sin respaldo estadístico, en tres grupos y por % de la población	Cálculo del tamaño de muestra independiente para cada estrato (k)
Selección de la muestra	Aleatoria	UPM sistemático USM MSA
Estimación	Autoponderada a nivel estatal	Se utiliza el factor de expansión dado por: $f_{ki} = \frac{N_k M_i}{n_k m_i}$

Fuente: Cálculos propios tomando en consideración los cuadros 7 y 9

Como se observa, aún con un margen de error mayor, para algunas entidades federativas, el tamaño muestral establecido por grupos, no es suficiente para

alcanzar el nivel de confianza del 95% como es el caso de Aguascalientes y Campeche. Para Chiapas, aún con un msa (que no lo es) el tamaño de muestra fijado por el muestrista no es suficiente para dar estimaciones con las características señaladas⁹.

También observamos que, a pesar de que el tamaño de muestra en el diseño empleado inicialmente se obtiene con la fórmula y especificaciones dadas para un msa, este tamaño no se emplea por ser una muestra muy pequeña para dar estimaciones a nivel estatal; el esquema de muestreo no es simple aleatorio al tener dos etapas de selección (localidades y familias) por lo tanto no se puede afirmar que la muestra se obtuvo con un muestreo simple aleatorio ni que tiene el 95% de confianza y un margen de error del 2.5% establecidos *ex ante* ni que es autoponderada dado que no es un msa ni la selección es realizada para que sea autoponderada, es decir, cada unidad de muestreo tiene un peso diferente, por lo cual se le tendría que aplicar ese peso al dar estimaciones. Aún con estas consideraciones, para asegurar el nivel de confianza y el margen de error para el tamaño de muestra empleado, estos se tendrían que calcular *ex profeso* (a posteriori) con métodos más complicados.

Para el diseño sugerido sí que se puede afirmar que el tamaño de la muestra utilizado es suficiente para tener una muestra “representativa” con un margen de error de 5% y un nivel de confianza de 95% establecidos *ex ante*, donde el tamaño de muestra calculado con las fórmulas se respeta, se establece desde el diseño que la selección de la muestra se hace en dos etapas y se determinan los factores de expansión para poder proporcionar estimaciones a los niveles respectivos. Así pues, los resultados de la encuesta de seguimiento en la que se emplea el diseño actual, pueden mejorarse si se emplea el diseño sugerido.

⁹ Ver fórmula No. 15

Conclusiones

La teoría del muestreo de poblaciones finitas nos insiste en que el tamaño de la muestra –en estudios probabilísticos- no depende del tamaño de la población ni de una regla de porcentaje. El cálculo del tamaño de la muestra depende, fundamentalmente, de la distribución de la característica, atributo o variable en estudio en la población objetivo (o “universo”) (mientras más homogénea la población en esos términos menor tamaño de muestra será necesario; mientras más heterogénea la población mayor tamaño de muestra será necesario) y de las condiciones fijadas (*ex ante*) relativas al nivel de confianza y a la precisión que requiere la estimación muestral (Kish, Cochran, Kalton).

Para un msa, en un mismo margen de error, la variación por tamaño de la población (“el efecto de la corrección por finitud”) ni es proporcional ni sigue “reglas de porcentaje”. Obtener muestras basadas en “reglas de porcentaje” (vgr.: el 10% o el 15%) puede ser innecesario (“porque hay más muestra de la necesaria”) que llevaría (de ser probabilística) a obtener márgenes de error menores a $\pm 1\%$ o si $N=1,000,000$ se requerirían 150,000 unidades muestrales; la misma encuesta podría ser imposible de lograr por el costo y tiempo involucrados. También puede darse el caso que los mismos porcentajes no alcancen los tamaños esperados.

Todo esto se refiere al msa, dicho diseño, no es, desgraciadamente, posible de ser empleado en la práctica porque se requiere insertar: (a) etapas de diseño (estratos, subestratos, etcétera); y (b) etapas de muestreo (conglomerados, subconglomerados, etcétera) y se forman diseños polietápicos complejos. Con base en lo anterior, podemos concluir que el diseño empleado para obtener el tamaño de muestra de la encuesta de puntos centinela, no es suficiente para afirmar que los resultados se extrapolan a toda la población beneficiaria con un nivel de confianza del 95% y un margen de error del 2.5%, ya que estos supuestos son empleados para calcular un tamaño de muestra tomando como base un diseño para un msa, y hemos notado a lo largo del documento que la selección

real de las unidades de muestreo se realiza por etapas, además de que al tratar de dar estimaciones o tener una muestra representativa por entidad federativa, recurren al muestreo estratificado, esto aunado a la selección por etapas nos llevaría a dar pesos de muestreo para las estimaciones, lo cual no se realiza al afirmar que es un diseño autoponderado.

Por otro lado, si el cálculo de la muestra se realiza tomando el diseño sugerido, las características de la muestra están dadas ex ante del cálculo, ya que en ningún momento se está recurriendo a algún tipo de muestreo no probabilístico. También se está especificando que es un muestreo probabilístico, por estratos y conglomerados bietápico y para dar estimaciones a nivel estatal, cada entidad federativa se toma como un estrato, un dominio de estudio sobre el cual hay que dar estimaciones y por lo cual se considera independiente de los demás y se realiza un cálculo de tamaño de muestra para cada uno de ellos. Se puede afirmar que ese tamaño muestral es suficiente y representativo de toda la población en estudio con un nivel de confianza del 95% y un margen de error de 5%. En este sentido, se ha tomado en cuenta el nivel de desagregación para el cual se desea dar estimaciones, se han tomado en cuenta las dos etapas de muestreo o selección (localidades y familias) para lo cual se utilizó el efecto de diseño.

Con el efecto de diseño y con la estratificación estatal, el tamaño de muestra aumenta suficiente en comparación con el diseño del msa. Además, se puede establecer desde el inicio del diseño el nivel de confianza y margen de error deseados (y estos pueden no ser los mismos para todos los estratos) y se puede afirmar que con esas características establecidas se tendrá un tamaño de muestra suficiente para dar resultados útiles que permitan inferir características sobre toda la población objetivo tanto a nivel estatal como nacional, utilizando los respectivos factores de expansión.

Con el diseño de muestreo sugerido en este documento no se modificará la forma de elegir a las familias que cumplen con las condiciones para ser beneficiarias, ni

se escogerán con mayor eficiencia, ni es para detectar errores de inclusión o exclusión, este diseño es para obtener una mejor muestra con el tamaño apropiado para disminuir los errores de muestreo y así mejorar la medición los indicadores que se obtienen a través de la encuesta de Puntos Centinela y poder dar la estimación del error que se comete.

Anexo 1. Notas metodológicas utilizadas para el cálculo de muestra en Oportunidades

MUESTRA DE LOCALIDADES PARA PUNTOS CENTINELA JUL-AGO 2004

NOTA TECNICA

Objetivo.-

Contar con un instrumento que permita evaluar diferentes aspectos de la operación del Programa de Desarrollo Humano Oportunidades.

Marco Muestral.-

Listado de localidades incorporadas al bimestre enero-febrero de 2004, esto representa 70,283 localidades.

Diseño Muestral.-

Se consideró un muestreo aleatorio simple.

Tamaño de la muestra._

Para fijar el tamaño de muestra se tomaron las siguientes consideraciones:

Se consideró un nivel de confianza del 95% con una precisión del 2.5%.

La muestra se repartió proporcionalmente en los estados de acuerdo a la cantidad de familias beneficiarias.

Se consideró para el tamaño de muestra por estado el mínimo entre el tamaño determinado por el grupo al que pertenece el estado y el 20% del total de localidades en el estado.

Esto dio como resultado una muestra de 3,489 localidades, lo que representa el 4.96% del marco muestral.

Selección de la muestra.-

Se determinaron las localidades de la muestra por estado de manera aleatoria, considerando únicamente localidades con cinco ó más titulares incorporados.

Observaciones:

Se clasificaron los estados en tres grupos de acuerdo a la cantidad de familias beneficiarias y se asignó un tamaño de muestra por grupo.

Grupo	Tamaño de muestra
1 (con menos de 45 mil hogares benef.)	40
2 (de 45 mil a 149,999 hogares benef.)	100
3 (con 150 mil o más hogares benef.)	200

MUESTRA DE LOCALIDADES PARA PUNTOS CENTINELA 2004

NOTA TECNICA

Objetivo.-

Contar con un instrumento que permita evaluar diferentes aspectos de la operación del Programa de Desarrollo Humano Oportunidades.

Marco Muestral.-

Listado de localidades incorporadas al bimestre julio-agosto de 2003, que cuentan con cinco ó más titulares. Esto representa 60,374 localidades.

Diseño Muestral.-

Se consideró un muestreo aleatorio simple.

Tamaño de la muestra._

Para fijar el tamaño de muestra se tomaron las siguientes consideraciones:

Se consideró un nivel de confianza del 95% con una precisión del 2.5%.

La muestra se repartió proporcionalmente en los estados de acuerdo a la cantidad de familias

beneficiarias y al tipo de localidades.

Se consideró para el tamaño de muestra por estado el mínimo entre el tamaño determinado por el

grupo a que pertenece el estado y el 20% del total de localidades en el estado.

Esto dio como resultado una muestra de 3,515 localidades, lo que representa el 5.82% del marco muestral, de las cuales 2,488 son localidades rurales y 1,027 son semiurbanas o urbanas.

Selección de la muestra.-

Se determinaron las localidades de la muestra por estado y por tipo de localidad de manera aleatoria.

Observaciones:

Se clasificaron los estados en tres grupos de acuerdo a la cantidad de familias beneficiarias y se asignó un tamaño de muestra por grupo.

Grupo	Tamaño de muestra
1 (con menos de 45 mil hogares benef.)	40
2 (de 45 mil a 149,999 hogares benef.)	100
3 (con 150 mil o más hogares benef.)	200

De acuerdo a la institución liquidadora, la muestra de las 3,515 localidades quedó distribuida de la siguiente manera:

- 2,718 localidades pagadas por telecom.
- 596 localidades pagadas por Bansefi
- 201 localidades pagadas por Bancomer

Estimación de resultados:

Debido que la selección de localidades dentro de cada estado se realizó de forma aleatoria, la muestra es auto ponderada a Nivel Estatal.

**MUESTRA PUNTOS CENTINELA
Periodo Julio-Diciembre 2003**

NOTA TÉCNICA

Objetivo.-

Contar con un instrumento que permita evaluar diferentes aspectos de la operación del Programa de Desarrollo Humano Oportunidades.

Marco Muestral.-

Listado de localidades incorporadas al bimestre noviembre-diciembre de 2002, que cuentan con cinco ó más titulares. Esto representa 60,222 localidades.

Diseño Muestral.-

Se consideró un muestreo aleatorio simple.

Tamaño de la muestra._

Para fijar el tamaño de muestra dentro de un contexto que fuera operativamente viable, se tomaron las siguientes consideraciones:

1. Para el cálculo de la muestra nacional, se consideró un nivel de confianza del 95% con una precisión del 2.5%.
2. La muestra se repartió proporcionalmente en los estados de acuerdo a la cantidad de familias beneficiarias.
3. Para asignar el tamaño de muestra por estado, se consideró el mínimo entre el tamaño determinado por el grupo al que pertenece el estado y el 20% del total de localidades en el estado.
4. La selección de las unidades de servicio¹ se hace de forma aleatoria por estado

¹ Escuelas de Educación Básica, Escuelas de Educación Media Superior, Centros de Salud

Esto dio como resultado una muestra de 3,457 localidades, lo que representa el 5.74% del marco muestral.

Selección de la muestra.-

1. Se determinaron las localidades de la muestra por estado de manera aleatoria.
2. Una vez identificadas las localidades de la muestra de manera sistemática y aleatoria se seleccionó la Unidad de Salud, Escuela de Educación Básica y Escuela de Educación Media Superior que daban servicio a la localidad.
3. Al seleccionar una unidad de servicio ésta se marcaba en la base de datos para que no fuera doblemente seleccionada en caso de que esa unidad de servicio atendiera a dos o mas localidades.

Observaciones:

A fin de hacer operativo el tamaño de la muestra resultante, se consideró la siguiente agrupación de estados y se asignó un tamaño de muestra por grupo.

Grupo	Tamaño de muestra
1	40
2	100
3	200

Se consideró para el tamaño de muestra por estado el mínimo entre el tamaño determinado por el grupo al que pertenece el estado y el 20% del total de localidades en el estado.

Conceptos Estadísticos para el cálculo de la muestra

1. UNIVERSO O POBLACION

Parámetros poblacionales

2. MUESTRA

Estimadores de los parámetros
Poblacionales

3. UNIVERSO O POBLACION

Marco muestral

4. MUESTRA ALEATORIA

- Selección de algunos elementos de la población
- Todos tienen la misma probabilidad de ser seleccionados
- Es una muestra representativa de la población

Siempre que se hable de muestreo, se estará hablando de probabilidades y de estimaciones

Por lo tanto con el muestreo se obtienen estimadores de los parámetros poblacionales con un cierto nivel de confianza

La distribución de las posibles muestras es una distribución normal

5. **Nivel de Confianza** – Probabilidad de que sin importar cuál sea mi muestra, el valor de mi parámetro estará en un intervalo determinado
6. **Precisión** – Grado de variabilidad que el estimador presenta alrededor del parámetro

Cálculo del tamaño de muestra

Para el caso de proporciones:

$$n' = p * q * (Z / d)^2$$

Donde

p-proporción en que aparece la característica
q-proporción de no aparición
Z-valor correspondiente al nivel de confianza
d-precisión deseada del estimador

de acuerdo al tamaño de la población se aplica la siguiente fórmula:

$$n = n' * (1 + n'/N)$$

Donde

n' – tamaño de muestra preliminar
N – tamaño de la población

Valores de Z

- Al 90% - 1.64
- Al 95% - 1.96
- Al 99% - 2.58

Características de la Muestra:

1. Se considera un muestreo simple aleatorio
2. Nivel de confianza del 95%

3. Se fija una precisión del 2.5%
4. La muestra se reparte proporcionalmente por estado de acuerdo a la cantidad de familias beneficiarias
5. La selección se hace de forma aleatoria por estado

Selección de unidades:

Para cada localidad de la muestra se seleccionó de forma aleatoria una unidad de salud que atiende a las familias beneficiarias de la localidad, de igual forma se seleccionó una escuela de educación básica y una escuela de educación media superior a las que asisten becarios de la localidad en muestra. Para la aplicación de la Cédula de Localidad se seleccionaron de forma aleatoria a cinco titulares beneficiarias de cada una de las localidades en muestra en los puntos de entrega de apoyos.

Estimación de resultados:

Debido que la selección de localidades dentro de cada estado se realizó de forma aleatoria, la muestra es auto ponderada a Nivel Estatal.

Anexo 2. Encuesta aplicada a las familias beneficiarias en las localidades en muestra

INSTRUCCIONES GENERALES

<p>Objetivo: La presente encuesta es para verificar la calidad de la atención que reciben las titulares del Programa</p> <p>Para realizar lo anterior se requiere:</p> <p>1. Entrevistar a la titular que se especifica en la carátula.</p> <p>Al momento de aplicar las encuestas ten presente que:</p> <ul style="list-style-type: none"> - Circule el número de la opción de respuesta a cada pregunta y se indicará cuando deba marcar más de una respuesta. - Dependiendo de la respuesta puede realizarse un salto a otra pregunta, tal como lo indica el cuestionario en la columna "PASE A". - Cuando en la columna PASE "A" se ha omitido el número, debes continuar con la pregunta inmediata siguiente. - En ningún caso se debe inducir la respuesta del entrevistado, se dejará que la titular conteste según lo considere. <p>Para finalizar la aplicación del instrumento:</p> <ul style="list-style-type: none"> - Agradecer el tiempo que dedicó la titular para la entrevista. <p>Material Requerido:</p> <ul style="list-style-type: none"> - Documento donde presente los datos que corroboren que es titular de Oportunidades (Identificación provisional, Contrato de la Institución liquidadora, holograma) en caso de que no sea la titular especificada en la carátula no aplique la cédula. <p>Si no puedes realizar la entrevista.</p> <ul style="list-style-type: none"> - Anota tus observaciones en el numeral 10 de la carátula y si el espacio no es suficiente en la pregunta final (13901) de la cédula, se pueden incluir las razones por las que no puedes aplicar la encuesta.

Selección al menos una opción de respuesta para cada pregunta

No.	PREGUNTA	OPCIONES		PASE A
101	¿En su familia hay niños que asistan a la escuela <u>primaria (3ro a 6to. grado) o secundaria?</u>	1	Sí	201
		2	No	
		98	No sabe	1301
		99	No contestó	

De los niños de su familia que reciben beca y asisten a la escuela Primaria o Secundaria ...

No.	PREGUNTA	OPCIONES		PASE A
201	¿Recibió información de cómo debe inscribir a sus hijos para que reciban la beca de Oportunidades?	1	Sí	301
		2	No	401
		98	No sabe	
		99	No contestó	
301	¿Quién le dio la información para inscribir a sus hijos a las becas de Oportunidades?	1	Director o Profesor	
		2	Alguna vocal	
		3	Enlace municipal	
			Personal de Oportunidades	
		4	Otra titular	
		5	Otra persona	
98	No sabe			
99	No contestó			
401	¿Sus hijos reciben beca de Oportunidades?	1	Sí	501
		2	No	
		98	No sabe	1301
		99	No contestó	
501	¿Cómo realizó la inscripción de sus hijos en la beca de Oportunidades?	1	Acudió a la escuela para llevar los formatos a firma y los entregó a Oportunidades	701
		2	Acudió a la escuela para solicitar la certificación	
		3	No ha realizado la inscripción	
		4	Otro	601
		98	No sabe	701
		99	No contestó	
601	¿Cuál?	_____		
701	¿Tuvo algún problema para inscripción sus hijos a la beca de Oportunidades?	1	Sí	801
		2	No	
		98	No sabe	901
		99	No contestó	
801	¿Qué problemas tuvo?	_____		

No.	PREGUNTA	OPCIONES		PASE A
901	En el presente ciclo escolar ¿Le pidieron en la escuela algo a cambio para la inscripción a sus hijos en la beca de Oportunidades? (formato E1)	1	Sí, asistencia a reuniones escolares	1101
		2	Sí, faenas o trabajos comunitarios en la escuela	
		3	Sí, comidas o regalos personales	
		4	Sí, trabajo personal (domésticos)	
		5	Sí, dinero	
		6	Otro	1001
98		7	No	1201
		98	No sabe	
		99	No contestó	
1001	¿Qué le pidieron?			
1101	¿Fue por acuerdo de la comunidad o de la asamblea del Comité de Promoción Comunitaria?	1	Sí	
		2	No	
		98	No sabe	
		99	No contestó	
1201	¿Recibieron sus libros de texto al principio de este ciclo escolar (2004-2005)	1	Sí	
		2	No	
		98	No sabe	




No.	PREGUNTA	OPCIONES		PASE A
1301	¿En su familia hay jóvenes que asistan a la escuela en el nivel <u>bachillerato</u> ?	1	Sí	1401
		2	No	2401
		98	No sabe	
		99	No contestó	

De los jóvenes de su familia que asisten a la escuela de Bachillerato y reciben beca...

No.	PREGUNTA	OPCIONES		PASE A
1401	¿Recibió información de cómo debe inscribir a sus hijos para que reciban la beca de Oportunidades?	1	Sí	1501
		2	No	1601
		98	No sabe	
		99	No contestó	
1501	¿Quién le dio la información para inscribir a sus hijos a las becas de Oportunidades?	1	Director o Profesor	
		2	Alguna vocal	
		3	Enlace municipal	
		4	Personal de Oportunidades	
		5	Otra titular	
		6	Otra persona	
98	No sabe	1901		
99	No contestó			
1601	¿Sus hijos reciben beca de Oportunidades?	1	Sí	1701
		2	No	2401
		98	No sabe	
		99	No contestó	
1701	¿Cómo realizó la inscripción de sus hijos en la beca de Oportunidades?	1	Acudió a la escuela para llevar los formatos a firma y los entregó a Oportunidades	1901
		2	Acudió a la escuela para solicitar la certificación	
		3	No ha realizado la inscripción	1801
		4	Otro	
		98	No sabe	
99	No contestó	1901		
1801	¿Cuál?			
1901	¿Tuvo algún problema para la inscripción de sus hijos en la beca de Oportunidades?	1	Sí	2001
		2	No	2101
		98	No sabe	
		99	No contestó	

No.	PREGUNTA	OPCIONES		PASE A
2001	¿Qué problemas tuvo?	_____		
2101	En el presente ciclo escolar ¿Le pidieron en la escuela algo a cambio para inscribir a sus hijos en la beca de Oportunidades?	1	Sí, asistencia a reuniones escolares	2301
		2	Sí, faenas o trabajos comunitarios en la escuela	
		3	Sí, comidas o regalos personales	
		4	Sí, trabajo personal (domésticos)	
		5	Sí, dinero	
		6	Otro	2201
7	No	2401		
98	No sabe			
99	No contestó			
2201	¿Qué le pidieron?	_____		
2301	¿Fue por acuerdo de la comunidad o de la asamblea del comité de promoción comunitaria?	1	Sí	
		2	No	
		98	No sabe	
		99	No contestó	

Ahora le voy a preguntar sobre los servicios de Salud...

No.	PREGUNTA	OPCIONES		PASE A
2401	¿Qué institución de salud le da la consulta médica de Oportunidades?	1	SSA 	
		2	IMSS Oportunidades  Oportunidades	
		3	IMSS Régimen-Obligatorio 	
		98	No sabe	
		99	No contestó	
2501	¿Tiene su tarjeta de citas médicas para la familia?	1	Sí	2601
		2	No	2701
		98	No sabe	
		99	No contestó	
2601	¿El personal de la unidad de salud le anota el día de su próxima consulta en su tarjeta?	1	Sí	
		2	No	
		98	No sabe	
		99	No contestó	
2701	¿Cuándo le tocó la última consulta a Usted o algún miembro de su familia?	1	Hace menos de 1 mes	
		2	De 1 a 2 meses	
		3	De 2 a 3 meses	
		4	De 3 a 6 meses	
		5	De 6 meses a 1 año	
		6	Hace más de 1 año	
		98	No sabe	
99	No contestó			
2801	¿Fue o fueron a la consulta?	1	Sí	3001
		2	No	2901
		98	No sabe	3301
		99	No contestó	
2901	¿Por qué no han asistido a las consultas?	1	Por asuntos personales	3301
		2	No quiso ir a la consulta médica	
		3	Olvido	
		98	No sabe	
		99	No contestó	
3001	¿Le dieron consulta?	1	Sí	3201
		2	No	3101
		98	No sabe	3301
		99	No contestó	

No.	PREGUNTA	OPCIONES		PASE A
3101	¿Por qué no le dieron consulta?	1	Estaban atendiendo una emergencia	3301
		2	Llegue tarde a la cita	
		3	Había mucha gente esperando atención y tuvo que irse	
		4	El personal de salud llegó tarde y tuvo que irse	
		5	El médico no asistió	
		6	La unidad móvil no llegó	
		7	El médico le negó la consulta	
		98	No sabe	
		99	No contestó	
3201	¿Dónde le dieron la consulta?	1	En la Clínica de Salud	
		2	Casa de Salud	
		3	En la Unidad Móvil	
		98	No sabe	
		99	No contestó	
3301	La última vez que fue a su consulta de Oportunidades ¿Le pidieron algo a cambio de otorgarle la atención?	1	No	3901
		2	Sí	3401
		98	No sabe	3901
		99	No contestó	
3401	¿Qué le pidieron a cambió?	1	Sí, faenas o trabajos en la unidad de salud o comunidad	3601
		2	Sí, utilizar algún método de planificación familiar	
		3	Sí, comidas o regalos personales	
		4	Sí, trabajo personal (domésticos)	
		5	Sí, dinero	
		6	Otro	
		98	No sabe	3901
		99	No contestó	
3501	¿Qué?			
3601	¿Hubo acuerdo de la comunidad o de la asamblea del comité de promoción comunitaria, para que lo pidieran?	1	Sí	3901
		2	No	3701
		98	No sabe	3901
		99	No contestó	
3701	¿Quién le pidió?	1	Médico	3901
		2	Enfermera	
		3	Promotor de Oportunidades	
		4	Otro	
		98	No sabe	3901
		99	No contestó	
3801	¿Quién?			

Vamos a hablar ahora de las pláticas de Salud

No.	PREGUNTA	OPCIONES		PASE A
3901	¿El personal de salud le anota las fechas en que debe acudir a las pláticas de educación para la salud?	1	Sí	
		2	No	
		98	No sabe	
		99	No contestó	
4001	¿El mes de _____ asistió a la plática para la salud? (El encuestador deberá nombrar el mes anterior)	1	Sí	4201
		2	No	4101
		98	No sabe	4401
		99	No contestó	
4101	¿Porqué no asistió a la plática?			
4201	¿Le dieron la plática?	1	Sí	4401
		2	No	4301
		98	No sabe	4401
		99	No contestó	

No.	PREGUNTA	OPCIONES		PASE A
4301	¿Por qué no le dieron la plática?	1	Estaban atendiendo una emergencia	
		2	El personal de salud llegó tarde y tuvo que irse	
		3	El personal de salud no asistió	
		98	No sabe	
		99	No contestó	
4401	¿En los últimos dos meses el personal de salud le pidió algo a cambio de registrar su asistencia a las pláticas?	1	No	5001
		2	Sí	4501
		98	No sabe	
		99	No contestó	5001
4501	¿Qué le pidieron a cambió?	1	Sí, faenas o trabajos en la unidad de salud o comunidad	
		2	Sí, utilizar algún método de planificación familiar	
		3	Sí, comidas o regalos personales	4701
		4	Sí, trabajo personal (domésticos)	
		5	Sí, dinero	
		6	Otro	4601
		98	No sabe	
		99	No contestó	5001
4601	¿Qué?			
4701	¿Hubo acuerdo de la comunidad o de la asamblea del comité de promoción comunitaria, para que lo pidieran?	1	Sí	5001
		2	No	4801
		98	No sabe	5001
		99	No contestó	
4801	¿Quién le pidió?	1	Médico	
		2	Enfermera	5001
		3	Promotor de Oportunidades	
		4	Otro	4901
		98	No sabe	
		99	No contestó	5001
4901	¿Quién?			

Vigilancia Nutricional a Niños menores de cinco años

No.	PREGUNTA	OPCIONES		PASE A
5001	¿En su familia hay niños menores de 5 años de edad?	1	Sí	5101
		2	No	
		98	No sabe	6301
		99	No contestó	
5101	En la última consulta médica de los niños, ¿Pesaron a su hijo?	1	Sí	
		2	No	
		98	No sabe	
		99	No contestó	

Niños de 4 meses a 2 años de edad

No.	PREGUNTA	OPCIONES		PASE A
5201	¿Cuántos niños tiene de 4 meses a 2 años de edad?	1	1	
		2	2	
		3	3	5301
		4	4	
		5	5 o más	
		6	Ninguno	
		98	No sabe	5701
		99	No contestó	

No.	PREGUNTA	OPCIONES		PASE A
5301	En la última cita ¿Cuántos sobres de suplemento alimenticio o papilla le entregaron para cada uno de estos niños?	1	De 1 a 4 sobres	5401
		2	5 sobres	
		3	De 5 a 9 sobres	
		4	10 sobres	
		5	Más de 10 sobres	
		6	Ninguno	5501
98	No sabe	5701		
99	No contestó			
5401	Según las indicaciones del médico ¿Para cuantos meses le deben de alcanzar los sobres que le dió?	1	1 mes	5701
		2	2 meses	
		3	más de 2 meses	
		98	No sabe	
		99	No contestó	

5501	¿Sabe por qué no le entregaron suplemento?	1	Porque yo no lo recibí	5701
		2	Porque no había	
		3	El personal de salud no me lo quiso dar	
		4	No fui a la entrega de suplemento (papilla) a la unidad de	
		5	Otra razón	
		98	No sabe	5701
99	No contestó			
5601	¿Cuál?			

Niños de 2 a 4 años de edad

No.	PREGUNTA	OPCIONES		PASE A
5701	¿Cuántos niños tiene de 2 a 4 años de edad?	1	1	5801
		2	2	
		3	3	
		4	4	
		5	5 o más	
		6	Ninguno	6301
98	No sabe			
99	No contestó			
5801	¿De cuántos de estos niños, el personal de salud le informó que tienen bajo peso o desnutrición?	1	1	5901
		2	2	
		3	3	
		4	4	
		5	5 ó más	
		6	Ninguno	6301
98	No sabe			
99	No contestó			
5901	En la última cita ¿Cuántos sobres de suplemento alimenticio o papilla le entregaron por cada uno de estos niños con bajo peso o desnutrición?	1	De 1 a 4 sobres	6001
		2	5 sobres	
		3	De 5 a 9 sobres	
		4	10 sobres	
		5	Más de 10 sobres	
		6	Ninguno	6101
98	No sabe	6301		
99	No contestó			
6001	Según las indicaciones del médico ¿Para cuantos meses le deben de alcanzar los sobres que le dió?	1	1 mes	6301
		2	2 meses	
		3	más de 2 meses	
		98	No sabe	
		99	No contestó	

No.	PREGUNTA	OPCIONES		PASE A
6101	¿Sabe por qué no le entregaron suplemento?	1	Porque yo no lo recibí	6301
		2	Porque no había	
		3	El personal de salud no me lo quiso dar	
		4	No fui a la entrega de suplemento (papilla) a la unidad de	
		5	Otra razón	6201
98	No sabe	6301		
99	No contestó			
6201	¿Cuál?			

Mujeres

No.	PREGUNTA	OPCIONES		PASE A
6301	¿En su familia hay mujeres embarazadas?	1	Sí	6401
		2	No	6901
		98	No sabe	
		99	No contestó	
6401	¿El personal de salud sabe que usted o en su familia hay mujeres embarazadas?	1	Sí	6501
		2	No	6901
		98	No sabe	
		99	No contestó	
6501	En el último mes ¿Cuántos sobres de suplemento alimenticio le entregaron por cada mujer embarazada?	1	De 1 a 5 sobres	6601
		2	6 sobres	
		3	De 6 a 11 sobres	
		4	12 sobres	
		5	Más de 12 sobres	
		6	Ninguno	6701
98	No sabe	6901		
99	No contestó			
6601	Según las indicaciones del médico ¿Para cuantos meses le deben de alcanzar los sobres que le dió?	1	1 mes	6901
		2	2 meses	
		3	más de 2 meses	
		98	No sabe	
		99	No contestó	
6701	¿Sabe por qué no le entregaron suplemento?	1	Porque yo no lo recibí	6901
		2	Porque no había	
		3	El personal de salud no me lo quiso dar	
		4	No fui a la entrega de suplemento (papilla) a la unidad de	
		5	Otra razón	6801
		98	No sabe	6901
99	No contestó			
6801	¿Cuál?			
6901	¿En su familia hay mujeres en periodo de lactancia (que le dan pecho a sus hijos)?	1	Sí	7001
		2	No	7501
		98	No sabe	
		99	No contestó	
7001	¿El niño al que le está dando pecho es menor a un año?	1	Sí	7101
		2	No	7501
		98	No sabe	
		99	No contestó	
7101	En la última cita ¿Cuántos sobres de suplemento alimenticio le entregaron por cada mujer en periodo de lactancia (dando pecho)? Nota: Mujeres con niños menores de 1 año	1	De 1 a 5 sobres	7201
		2	6 sobres	
		3	De 6 a 11 sobres	
		4	12 sobres	
		5	Más de 12 sobres	
		6	Ninguno	7301
98	No sabe	7501		
99	No contestó			

No.	PREGUNTA	OPCIONES		PASE A
7201	Según las indicaciones del médico ¿Para cuantos meses le deben de alcanzar los sobres que le dió?	1 2 3 98 99	1 mes 2 meses más de 2 meses No sabe No contestó	7501
7301	¿Sabe por qué no le entregaron suplemento?	1 2 3 4 5 98 99	Porque yo no lo recibí Porque no había El personal de salud no me lo quiso dar No fui a la entrega de suplemento (papilla) a la unidad de Otra razón No sabe No contestó	7501 7401 7501
7401	¿Cuál?			

No.	PREGUNTA	OPCIONES		PASE A
7501	Nota: El encuestador llena esta pregunta de acuerdo a las respuestas de las preguntas anteriores: ¿Existe en la familia de la titular niños menores de 2 años o niños de 2 a 4 años con algún grado de desnutrición o mujeres embarazadas o mujeres en lactancia?	1 2	Sí No	7601 8401

Vamos a continuar con el tema de Suplementos Alimenticios o Papilla

No.	PREGUNTA	OPCIONES		PASE A
7601	¿Quién en su familia toma el suplemento alimenticio que recibe?	1 2 3 98 99	Únicamente la persona a quien se lo recetan (niños menores de 2 años, niños con desnutrición, mujeres embarazadas o en periodo de lactancia) A la persona que se lo recetan y a otros integrantes de la familia Nadie No sabe No contestó	7701 8401
7701	Qué efecto provoca el suplemento alimenticio a quien lo consume?	1 2 3 98 98	Le hace bien a quien lo toma (están más sanos, sube de peso) Les hace daño cuando lo toman No le hace nada No sabe No contestó	7901 7801 7901
7801	Si considera que les hace daño, ¿Qué tipo de daño le hace ?			
7901	¿Le pidieron algo a cambio de entregarle los sobres de suplemento alimenticio?	1 2 3 4 5 6 7 98 99	No Sí, faenas o trabajos en la unidad de salud o comunidad Sí, utilizar algún método de planificación familiar Sí, comidas o regalos personales Sí, trabajo personal (doméstico) Sí, dinero Otro No sabe No contestó	8201 8101 8001 8201
8001	¿Qué?			
8101	¿Se aceptó en la comunidad o en asamblea del comité de promoción comunitaria que le pidieran algo a cambio de la entrega del suplemento?	1 2 98 99	Sí No No sabe No contestó	

No.	PREGUNTA	OPCIONES		PASE A
8201	¿El personal de salud le ha enseñado a preparar el suplemento alimenticio?	1	Sí	
		2	No	
		98	No sabe	
		99	No contestó	
8301	Me puede decir por favor ¿Cómo se prepara el suplemento? NO INDUZCA LA RESPUESTA Procedimiento suplemento para niños <i>* Se deposita en un recipiente cuatro</i> <i>* Poner tres cucharadas de agua hervida o</i> <i>* Se revuelven bien los ingredientes.</i> Procedimiento suplemento para mujeres <i>* Se deposita en un recipiente cinco</i> <i>cucharadas soperas copeteadas del</i> <i>* Agrégales a 1/2 vaso de agua hervida o</i> <i>clorada</i> <i>* Se revuelven bien los ingredientes.</i>	1	Describió completo un procedimiento	
		2	No describió completo un procedimiento	
		98	No sabe	
		99	No contestó	

Le voy a preguntar la opinión de los servicios de Salud

No.	PREGUNTA	OPCIONES		PASE A
8401	¿En su última consulta cómo le trataron? :	1	Bien	8601
		2	Regular	
	NOTA: Lea las opciones de respuesta	3	Mal	8501
		98	No sabe	8601
		99	No contestó	
8501	¿Por qué?			

Ahora vamos a hablar de la entrega de apoyos monetarios

No.	PREGUNTA	OPCIONES		PASE A
8601	¿Qué institución le entrega los apoyos monetarios de Oportunidades?	1	Telecomm	
		2	Bansefi ó Institución de Ahorro	
		3	Bancomer	
		98	No sabe	
		99	No contestó	
8701	¿Cuánto dinero gasta en trasladarse (ida y vuelta) para recoger sus apoyos monetarios de Oportunidades?	1	Nada	
		2	De \$1 a \$10	
		3	De \$11 a \$20	
		4	De \$21 a \$30	
		5	Más de \$30	
		98	No sabe	
		99	No contestó	
8801	¿En qué se trasporta usted al lugar donde le entregan los apoyos monetarios? Puede seleccionar más de una opción	1	Caminando	
		2	En transporte público	
		3	En bicicleta	
		4	En auto particular	
		5	En animal de carga	
		6	Otro	
		98	No sabe	
		99	No contestó	
8901	¿Cuánto tiempo tarda para llegar desde su casa al lugar donde le entregan los apoyos monetarios?	1	Menos de una hora	
		2	De 1 a 3 hrs.	
		3	Más de 3 hrs.	
		98	No sabe	
		99	No contestó	

No.	PREGUNTA	OPCIONES		PASE A
9001	En la última entrega de apoyos monetarios, ¿Le dieron completo lo correspondiente a su apoyo de alimentación?	1	Sí	9201
		2	No	9101
		3	No sabe, no le han entregado su recibo	9201
		4	No revisa su recibo	
		5	No sabe leer	
		98	No sabe	
99	No contestó			
9101	¿Por qué?			
9201	En la última entrega de apoyos monetarios, ¿Le dieron completo lo correspondiente a becas educativas?	1	Sí	9401
		2	No	9301
		3	No sabe, no le han entregado su recibo	9401
		4	No revisa su recibo	
		5	No sabe leer	
		97	No aplica (No hay becarios en su familia)	
		98	No sabe	
		99	No contestó	
9301	¿Por qué?			
9401	En la última entrega de apoyos monetarios, ¿Cómo fue la atención del personal que le entrega su dinero?	1	Buena	9601
		2	Regular	9501
		3	Mala	
		97	No aplica (Entrega de apoyos monetarios por cajero automático)	9901
		98	No sabe	9601
99	No contestó			
9501	¿Cuál es la causa principal por la cual Usted considera que la atención fue mala ?			
9601	En la última entrega de apoyos monetarios, ¿El personal que le entrega su dinero le pidió algo a cambio?	1	Sí, faenas o trabajos en la unidad de salud o comunidad	9801
		2	Sí, utilizar algún método de planificación familiar	
		3	Sí, comidas o regalos personales	
		4	Sí, trabajo personal (domésticos)	
		5	Sí, dinero	
		6	Otro	9701
		7	No	9901
		98	No sabe	
99	No contestó			
9701	¿Qué?			
9801	¿Hubo acuerdo de asamblea del comité de promoción comunitaria o de la comunidad, para que lo pidieran?	1	Sí	
		2	No	
		98	No sabe	
		99	No contestó	

Ahora le voy a preguntar sobre lo que debe saber del programa Oportunidades

No.	PREGUNTA	OPCIONES		PASE A
9901	En este año, ¿Cuántas veces recibió información o capacitación sobre Oportunidades? (no se consideran las pláticas educativas de salud en la unidad de salud)	1	2 o más	
		2	Una	
		3	Ninguna	
		98	No sabe	
		99	No contestó	
<p><i>Pueden ser:</i></p> <ul style="list-style-type: none"> * Reuniones convocadas por las vocales o enlace municipal * En los puntos de entrega de apoyos monetarios, impartidos por el personal de Oportunidades o Contraloría Social. * Reuniones convocadas por el personal de Oportunidades 				

No.	PREGUNTA	OPCIONES		PASE A
10001	¿Qué apoyos les otorga el programa Oportunidades?	1	Atención médica	
		2	Sesiones de educación para la salud	
		3	Suplemento alimenticio	
		4	Apoyo monetario para la alimentación	
		5	Apoyo monetario para educación	
		6	Apoyo para útiles escolares	
		7	Jóvenes con Oportunidades	
		98	No sabe	
		99	No contestó	
10101	¿Qué debe realizar su familia para seguir recibiendo los apoyos del programa?	1	Registrarse en la unidad médica	
		2	Asistir a la plática mensual de salud	
		3	Cumplir con la citas programadas	
		4	Destinar los apoyos al bienestar familiar en especial a la alimentación	
		98	No sabe	
		99	No contestó	
10201	¿Con cuántos meses de faltas a la unidad de salud pueden dar de baja a su familia? <i>4 meses continuos o 6 discontinuos en el curso de un año calendario</i>	1	Sabe	
		2	No sabe	
		99	No contestó	
10301	¿Con cuántas faltas injustificadas se suspende la beca a los niños que la reciben? <i>Primaria: 4 o más inasistencias injustificadas Secundaria: 28 horas de inasistencia</i>	1	Sabe	
		2	No sabe	
		97	No aplica (No hay becarios en su familia)	
		98	No contestó	
10401	¿Cuánto tiempo lleva usted como beneficiaria del Programa?			
10501	¿Sabe usted que a los tres años de estar en el Programa le van aplicar una encuesta de características socioeconómicas ? EXPLICAR	1	Sí	
		2	No sabe	
		99	No contestó	

Le voy a preguntar sobre las vocales del comité de promoción comunitaria

No.	PREGUNTA	OPCIONES		PASE A
10601	¿Es usted vocal del comité de promoción comunitaria?	1	Sí	11301
		2	No	10701
		98	No sabe	
		99	No contestó	11301
10701	En los dos últimos meses, ¿Alguna de las vocales le solicitó dinero para traslado o trámites con relación al Programa?	1	Sí, por acuerdo de asamblea	10801
		2	Sí, sin acuerdo de asamblea	
		3	No	
		98	No sabe	10901
		99	No contestó	
10801	¿Sabe que trámites?			
10901	En general, ¿Cómo ha sido la participación de las vocales del comité de promoción comunitaria con la comunidad beneficiaria de Oportunidades?	1	Buena	11101
		2	Regular	
		3	Mala	11001
		98	No sabe	11101
		99	No contestó	
11001	¿Cuál es la causa principal por la cual Usted considera que la participación fue mala ?			

No.	PREGUNTA	OPCIONES		PASE A
11101	¿Quién le avisó de la entrega de apoyos monetarios?	1	El Enlace Municipal	
		2	La vocal del Comité de Promoción	
		3	Personal de Oportunidades	
		4	Otra titular	
		5	Radio	
		6	Otro	
		98	No sabe	
		99	No contestó	
11201	Para esta entrega ¿Con cuántos días antes le avisaron de la fecha de entrega de apoyos?	1	5 o más días	
		2	3 o 4 días	
		3	2 días	
		4	Ayer	
		5	No le avisaron	
		98	No sabe	
		99	No contestó	
11301	En los dos últimos meses, ¿Alguna persona la presionó o amenazó con quitarle los apoyos del programa si no participaba o votaba a favor de algún partido político o candidato a elección popular?	1	Sí	11401
		2	No	
		98	No sabe	11501
		99	No contestó	
11401	¿Quién?			
11501	¿Conoce al Enlace Municipal?	1	Sí	11601
		2	No	
		98	No sabe	11801
		99	No contestó	
11601	En general, ¿Cómo ha sido la atención que ha recibido del Enlace Municipal?	1	Buena	11801
		2	Regular	
		3	Mala	11701
		98	No sabe	
		99	No contestó	11801
11701	¿Cuál es la causa principal por la cual Usted considera que la atención del Enlace Municipal es mala?			

Le voy a preguntar sobre el personal de la Coordinación Estatal de Oportunidades

No.	PREGUNTA	OPCIONES		PASE A
11801	En los dos últimos meses, ¿ha solicitado algún trámite a Oportunidades?	1	Sí	11901
		2	No	
		98	No sabe	12401
		99	No contestó	
	<div style="border: 1px solid black; padding: 5px;"> <p><i>Cambio de Titular</i> <i>Cambio de domicilio</i> <i>Cambio de escuela</i> <i>Reexpedición de apoyos</i> <i>Alta de integrantes</i> <i>Baja de Integrantes</i> <i>Corrección de Nombres</i> <i>Etc.</i></p> </div>			

No.	PREGUNTA	OPCIONES		PASE A
11901	¿Cuál? <div style="border: 1px solid black; padding: 2px; width: fit-content; margin: 10px auto;">Puede elegir más de una opción</div>	1 2 3 4 5 6 7 8 9 10 11 12 98 99	Cambio de titular Cambio de domicilio Cambio de escuela Reexpedición de apoyos Reactivación Reincorporación Corrección de nombres Reposición de medios de entrega de apoyos Inscripción extemporánea a escuela Alta de integrantes Baja de integrante Otro No sabe No contestó	
12001	¿A quién lo solicitó?	1 2 3 4 98 99	Personal de Oportunidades Enlace Municipal Vocal Otro No sabe No contestó	12201 ----- 12101 ----- 12201
12101	¿A quién?			
12201	¿Le dieron respuesta o atención a su trámite?	1 2 3 99	Sí No No sabe No contestó	
12301	¿Hace cuánto tiempo lo solicitó?	1 2 3 98 99	De 1 día a 1 mes De 1 mes a 2 meses Más de 2 meses No sabe No contestó	
12401	¿En los últimos dos meses, alguien del personal de la Coordinación Estatal de Oportunidades le ha pedido algo a cambio de inscribirlo o mantenerlo en el Programa?	1 2 3 4 5 6 7 98 99	Sí, asistencia a reuniones políticas Sí, faenas o trabajos comunitarios en la localidad Sí, comidas o regalos personales Sí, trabajo personal (doméstico) Sí, dinero Otro No No sabe No contestó	12601 ----- 12501 ----- 12701
12501	¿Qué le solicitaron?			
12601	¿Quién?			
12701	Cuando recibe el apoyo monetario, ¿Sabe que Bimestre recibe?	1 2 98 99	Sí No No sabe No contestó	
12801	¿Cada cuándo trata con el personal de la Coordinación Estatal de Oportunidades?	1 2 3 4 5 98 99	Cada entrega de apoyos Cada que convocan a reunión Cuando va a la Coordinación Estatal de Oportunidades Cada visita a la clínica Nunca, no los conoce No sabe No contestó	12901 ----- 13101
12901	¿Cómo es la atención del personal de Coordinación Estatal de Oportunidades?	1 2 3 98 99	Buena Regular Mala No sabe No contestó	13101 ----- 13001 ----- 13101

No.	PREGUNTA	OPCIONES		PASE A
13001	¿Cuál es la causa principal por la cual Usted considera que la atención del personal de Oportunidades es mala?			

Ahora le voy a preguntar de las quejas, denuncias o reconocimientos sobre Oportunidades...

No.	PREGUNTA	OPCIONES		PASE A
13101	¿Le han informado que usted puede reclamar si no recibe un buen trato ó servicio, por parte del personal que otorga los servicios del programa?	1 2 98 99	Sí No No sabe No contestó	
13201	¿Le dijeron que los trámites y servicios de Oportunidades son gratuitos?	1 2 98 99	Sí No No sabe No contestó	
13301	¿En el caso de tener alguna queja o reconocimiento sobre Oportunidades, me podría decir cómo denunciarla? <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;">Puede elegir más de una opción</div>	1 2 3 4 5 6 7 8 98 99	Teléfono Escrito (Buzón de Contraloría Social) Escrito (Carta vía correo) Escrito (Entrega personal) Escrito (Reunión con el Comité de Promoción Comunitaria) Escrito (Reunión con el Enlace Municipal) Correo electrónico Otro No sabe No contestó	
13401	¿Ha visto buzones de Oportunidades para presentar quejas, denuncias o reconocimientos?	1 2 98 99	Sí No No sabe No contestó	13501 13701
13501	¿Me podría decir en dónde ha visto buzones de la contraloría para presentar quejas, denuncias o reconocimientos	1 2 3 4 5 6 98 99	Módulos de pago Sucursales bancarias Unidades de salud Presidencia Municipal Otro No ha visto No sabe No contestó	13701 13601 13701
13601	¿Dónde?			
13701	¿Sabe si hay algún número telefónico en el que puede presentar sus quejas, denuncias, peticiones o reconocimientos? <i>Si la titular no conoce los teléfonos, hacer que lo visualice en su tarjeta de identificación o que los anote en algún papel.</i>	1 2 98 99	Sí No No sabe No contestó	
13801	En este año, ¿Ha presentado alguna queja, denuncia, petición o reconocimiento?	1 2 98 99	Sí No No sabe No contestó	

No.	PREGUNTA	OPCIONES		PASE A
13901	Plantear las observaciones pertinentes con respecto al levantamiento de la cédula o bien las causas de la no aplicación de la cédula			

Bibliografía y referencias

Cochran, William G., *Técnicas de Muestreo*, Décima reimpresión, CECSA, España 1993

Cohen, Ernesto y Rolando Franco, *Transferencias con corresponsabilidad, una Mirada latinoamericana*, FLACSO, México, 2006

Colegio Mexiquense. *Evaluación del cumplimiento de metas, costos unitarios de apego del Programa de Desarrollo Humano Oportunidades a las Reglas de Operación 2005*. Introducción y resumen ejecutivo. México: Colegio Mexiquense, 2006.

Cruz, C. De la Torre R, Velázquez C. *Informe compilatorio. Evaluación externa de impacto Del Programa Oportunidades 2001-2006*. Cuernavaca, México: Instituto Nacional de Salud Pública, 2006,

Deming, William Edwards, *Some Theory of Sampling*, Dover Publications, Inc, New York, 1966

González de la Rocha, Mercedes, *Procesos Domésticos y Vulnerabilidad*, Publicaciones de la Casa Chata, CIESAS, 2006

Hansen, M.H., Horwitz, W.N. y Madow, W.G., *Sample Survey Methods and Theory*, Vol. 1, Methods and Applications. John Wiley & Sons, Inc., New York, 1953.

Kalton, Graham, *Introduction to Survey Sampling*, Series: Quantitative Applications in the Social Sciences, SAGE publications, The International Professional Publishers, Newbury Park, London, 1983

Kish, Leslie, *Muestreo de encuestas*, Primera reimpresión, México, 1975, Editorial Trillas

Levy, Santiago y Rodríguez, Evelyne, *Sin Herencia de Pobreza*, El Programa PROGRESA-Oportunidades de México, Banco Interamericano de Desarrollo, Planeta, 2005.

Ley de Planeación, publicada el 13 de junio de 2003 en el Diario Oficial de la Federación

Ley General de Desarrollo Social, publicada el 20 de enero de 2004 en el Diario Oficial de la Federación

Lohr, Sharon L., *Muestreo: Diseño y Análisis*, 2000, International Thomson Editores, México

López, María de la Paz, Salles, Vania, Coordinadoras, *El Programa Oportunidades examinado desde el género*, UNIFEM, El Colegio de México

Madariaga Orozco, Camilo; Omar Sierra . 2000. "Redes Sociales y Pobreza". *Psicología desde el Caribe*, vol. , num. 005 , pp.127-156.

Manual de Procedimientos para la Operación del Programa. Programa de Desarrollo Humano Oportunidades, 2004

Mendenhall, W., Ott L. y Scheafer R.L. (1971), *Elementary Survey Sampling*, Duxbury Press

Méndez Ramírez, Ignacio, et al., *Conceptos básicos de muestreo*, Monografías, volumen 12, No. 27, mayo 2004, Instituto de investigaciones en matemáticas aplicadas y en sistemas.

Pérez López, César, *Muestreo Estadístico*, 2005, Pearson Prentice Hall

Plan Nacional de Desarrollo 2001-2006

Raj, Des (1968), *Sampling Theory*, McGraw Hill Co.

Reglas de Operación 2004, Programa de Desarrollo Humano Oportunidades.

Reglas de Operación 2005, Programa de Desarrollo Humano Oportunidades.

Román, M.F., Notas de Introducción al Muestreo, Curso Internacional de Muestreo, INEGI, 1998.

Särndal, C.E. Swensson, B., Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer-Verlag

Sudman, Seymor, *Applied Sampling*, Quantitative Studies in social Relations, Academic Press, New York, 1976.

Sukhatme, P.V. y Sukhatme, B. V. Asok, C. (1984), *Sampling Theory of Surveys with Applications*, Iowa State University Press.

Torres González, Luis Gabriel, *Sinergias con Oportunidades*, CIESAS, Papeles de la Casa Chata, 2006.

<http://www.conapo.gob.mx/publicaciones/IndiceMargLoc2005.pdf>

Página electrónica de Oportunidades en: <http://www.oportunidades.gob.mx/>

"Por una Globalización justa: crear oportunidades para todos". *Aportes* [en línea] 2004, IX (027):[fecha de consulta: 14 de noviembre de 2007] Disponible en:

<http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=37602710>> ISSN 1665-1219

Rodríguez, Pedro Gerardo. Fábulas de pobreza, desigualdad y educación. *Revista Latinoamericana de Estudios Educativos* [en línea] 2002, XXXII (004):[fecha de consulta: 14 de noviembre de 2007] Disponible en: <http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=27032402>> ISSN 0185-1284

Gangas Peiró, Pilar. Desigualdad y pobreza: América Latina y Europa desde 1950. *Política y Cultura* [en línea] 2003, (020):[fecha de consulta: 12 de diciembre de 2007] Disponible en: <http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=26702002>> ISSN 0188-7742