



| **UNR** Universidad  
Nacional de Rosario

---

Tesina

LICENCIATURA EN ESTADÍSTICA

# **“Diseño muestral para Encuesta de Satisfacción de Asociados AVALIAN”**

Alumna: Sofía Suares

Director: M. Sc. Gonzalo Pablo Marí

Codirectora: Mag. Virginia Laura Borra

Año 2024

# Agradecimientos

Encontrándome en la recta final de mi carrera universitaria, sentimientos de alegría, orgullo y emoción colman mi corazón. Nada hubiese sido igual sin todas esas personas que fueron sostén y me acompañaron en este camino. Siento un profundo agradecimiento a ellos.

En primer lugar, a mi mamá y a mi papá. Por su presencia y apoyo, por confiar siempre en mí y por mimarme y alentarme con los cafés, té y capuchinos cuando preparaba materias. A mis hermanos Franco y Agustín, a la Lela y al Lelo, a la Chiche, a Flor, a la Tata y a mis tíos y primos por el apoyo y el amor.

A Jose, por estar siempre presente de manera incondicional. No sólo por acompañarme en cada momento de la carrera, sino por haberme acompañado en todo aspecto de mi vida. Por las tardes y noches de estudio, por ser compañera de facultad, mejor amiga y hermana.

A Clari, Meli, Vicky, Cande, Miru, Martu y Abru. Amigas de mi alma, estoy eternamente agradecida por su amistad, por haber compartido los logros y haberme apoyado en cada examen y etapa de mi vida.

A mis amigos de la facultad, Juanchi, Luisi, Luli, Vale, Alfon, Eve, Mauri, Cami, Romi y Goty. Por la ayuda y el camino transitado juntos.

A Gonzalo y Virginia, por acompañarme y guiarme en esta última etapa, por su conocimiento y dedicación en esta tesina.

A todos los profesores, por transmitirme más que conocimiento. Cada uno dejó una parte suya en mí, que hoy me acompaña como profesional.

A Avalian por brindarme la oportunidad de desarrollarme profesionalmente y por la confianza que me tuvieron para realizar este gran proyecto.

A Lu, Rodri y Marce por enseñarme y hacerme sentir tan cómoda en este momento que fue tan importante para mí: mi incorporación en el ámbito laboral. Por la paciencia, dedicación y ayuda brindada en el proyecto de la Encuesta.

Gracias a todos los que fueron parte de mi vida en estos años.

# Resumen

En el año 2023, la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario firma un convenio con Avalian, una empresa de medicina prepaga de Argentina, para colaborar en el proyecto de la Encuesta de Satisfacción que realiza anualmente la empresa.

Esta encuesta se lleva a cabo con la finalidad de determinar el nivel de satisfacción general que tienen sus asociados. No sólo ofrece una valiosa oportunidad para identificar áreas de mejora, sino también para reconocer aquellos aspectos en los que Avalian se ha desempeñado exitosamente. Los resultados obtenidos sirven como base para la toma de decisiones. A partir de los mismos se desarrollan estrategias que buscan fortalecer la relación de confianza y fidelidad con sus asociados.

La confianza en los resultados es, por lo tanto, crucial para asegurar que las decisiones sean efectivas y contribuyan a mejorar la satisfacción del asociado. Sin embargo, en los últimos años, esta confianza se ha visto afectada tanto por sesgos causados por el método de recolección de datos como por el diseño muestral elegido. Frente a esta problemática y en base al interés de Avalian en obtener conclusiones válidas, en esta tesina se busca presentar una estrategia disruptiva que minimice los sesgos y maximice la confianza de los resultados.

En primer lugar, se plantea utilizar un diseño muestral que se enmarque en los muestreos de tipo probabilístico. Posterior al análisis de antecedentes y a la determinación de los objetivos de la encuesta actual, se define la población objetivo, las variables de interés y los dominios de estimación con los que se quiere trabajar. Se decide utilizar un muestreo estratificado simple al azar de grupos familiares (GF). Debido al interés en obtener estimaciones precisas de tres valores poblacionales en cuatro tipos de dominios de estimación, la adjudicación de la muestra sigue un enfoque multivariado y multidominio. La metodología utilizada es una generalización del algoritmo de Bethel.

Se obtienen tamaños de muestra para cuatro escenarios de interés, diferenciados en las categorizaciones utilizadas de los tipos de dominios, que van desde 1.173 hasta 7.218 GF. Se elige el menor tamaño de muestra, por tratarse de la primera experiencia utilizando esta metodología y por cuestiones de recursos. Además, al planear el diseño, se contempla la no respuesta y se plantea una posible solución frente a este problema. Se asume que la misma no está asociada a las características de la encuesta, es decir, quienes no responden se comportan de manera similar a quienes sí lo hacen, por lo que se utiliza la técnica de sobremuestreo. Se

supone una tasa de respuesta del 50%, siendo la muestra inflada final de 2.346 GF. Finalmente, se obtiene el listado de GF seleccionados en la encuesta.

Respecto a la forma de contacto, se busca obtener la respuesta cualquiera sea el medio de contacto. El diseño muestral se define de manera tal que la forma de contacto no genere sesgo significativo en la respuesta.

Además, se presenta una propuesta de formulario, cuya extensión consta de un máximo de catorce preguntas. Se plantea simplificar la formulación de preguntas y las escalas utilizadas, de manera que sea claro y conciso.

# Índice

<b>Agradecimientos.....</b>	<b>1</b>
<b>Resumen.....</b>	<b>2</b>
<b>Índice.....</b>	<b>3</b>
<b>1. Introducción.....</b>	<b>6</b>
<b>2. Objetivos.....</b>	<b>9</b>
2.1 Objetivo general.....	9
2.2 Objetivos específicos.....	9
<b>3. Antecedentes.....</b>	<b>10</b>
3.1 Objetivos de las encuestas anteriores.....	10
3.2 Población objetivo.....	11
3.3 Método de selección.....	12
3.4 Forma de contacto.....	12
3.5 Instrumento de recolección de datos.....	13
3.6 Tamaños de muestra.....	13
3.7 Indicadores de satisfacción.....	14
3.7.1 Satisfacción General.....	14
3.7.2 Recomendación.....	14
3.7.3 Inconvenientes.....	15
3.7.4 Recompra.....	15
3.7.5 Otros indicadores.....	15
3.8 Recomendaciones basadas en antecedentes.....	16
<b>4. Encuesta de Satisfacción 2023.....</b>	<b>17</b>
4.1 Materiales.....	17
4.2 Objetivos de la Encuesta de Satisfacción 2023.....	18
4.3 Población objetivo.....	18
4.4 Parámetros de interés.....	23
4.5 Dominios de interés.....	24
4.6 Cuestionario y forma de contacto.....	24
<b>5. Diseño muestral.....</b>	<b>26</b>
5.1 Determinación del tamaño de la muestra y su adjudicación.....	26
5.1.1 Definición de estratos.....	27
5.1.2 Adjudicación óptima.....	27
5.1.3 Adjudicación multivariada: estimaciones para la población completa.....	28
5.1.3.1 Modelo de adjudicación.....	28
5.1.3.2 Problema de adjudicación óptima.....	30
5.1.3.3 Sensibilidad de los costos de la encuesta respecto a las restricciones de variancias.....	32
5.1.3.4 Algoritmos computacionales.....	32
5.1.4 Adjudicación multivariada: estimaciones para múltiples dominios.....	33

5.1.4.1 Algoritmos resolutivos.....	35
5.1.5 Paquete R2BEAT del software R.....	36
5.1.5.1 Bases de datos requeridas.....	36
5.1.5.1.1 Base STRATUM.....	36
5.1.5.1.2 Base CV.....	39
5.1.5.2 Archivos de salida.....	40
5.2 Estimación de parámetros.....	41
5.2.1 Estimación de parámetros bajo muestreo estratificado.....	41
5.2.1.1 Estimación de los parámetros de interés.....	41
<b>6. Resultados.....</b>	<b>45</b>
6.1 Obtención del tamaño de la muestra.....	45
6.2 Tratamiento de la no respuesta.....	46
6.3 Selección de grupos familiares.....	47
<b>7. Conclusiones.....</b>	<b>49</b>
<b>8. Bibliografía.....</b>	<b>51</b>
<b>9. Anexo.....</b>	<b>53</b>
9.1 Anexo 1: Definiciones de Avalian.....	53
9.2 Anexo 2: Cuestionario.....	54
9.3 Anexo 3: Seguimiento encuesta.....	56
9.4 Anexo 4: Código de R utilizado.....	57

# 1. Introducción

Avalian es una empresa de medicina prepaga que brinda cobertura médica hace más de 45 años, en cada rincón de Argentina. Su propósito es cuidar y mantener la salud de sus asociados, garantizando el acceso a una cobertura geográfica nacional e integrada.

Bajo el nombre Aca Salud, nació en el año 1975 para cuidar la salud del productor agropecuario. De a poco fue creciendo, cubriendo luego las necesidades de las familias del productor agropecuario, la de los profesionales independientes, la de pequeños y medianos empresarios, y finalmente, la de todas las personas que buscaban una cobertura médica humana y de alta calidad.

En el año 2020, luego de un gran proceso de evolución, Aca Salud se transformó en Avalian. Manteniendo sus principios y valores, pero con un compromiso renovado: cuidar con eficiencia y calidez a cada vez más personas.

Con el objetivo de conocer el nivel de satisfacción general de sus asociados con el desempeño de la empresa, cada año se lleva a cabo en Avalian la Encuesta de Satisfacción de Asociados. Esta encuesta ha sido abordada anteriormente por colaboradores y recursos internos de la empresa y, en algunas oportunidades, por distintas consultoras. En general, se ha realizado bajo una modalidad híbrida de recolección de datos por los métodos CATI y CAWI<sup>1</sup>. Las preguntas incluidas están vinculadas a la satisfacción general del asociado, al nivel de recomendación de Avalian a familiares/amigos, a la recompra, entendida como si el asociado volvería a elegir Avalian como empresa de medicina prepaga, y a la existencia de inconvenientes, sus causas y resolución.

Los resultados obtenidos a través de la encuesta son de gran valor, ya que en base a ellos se toman decisiones que aspiran a mejorar el servicio y, por lo tanto, aumentar la satisfacción del asociado. Es por esto que resulta fundamental que estos resultados sean confiables, de manera que se garantice la validez de las conclusiones.

---

<sup>1</sup> Los métodos CATI y CAWI refieren a entrevistas telefónicas asistidas por computadora y entrevistas web asistidas por computadoras, respectivamente. En el método de entrevista CATI, los entrevistadores capacitados llaman a los números de teléfono de los encuestados y registran y guardan las respuestas utilizando un programa computacional en un dispositivo electrónico (computadora, tablet, celular inteligente, etcétera). Mientras que en el método CAWI, las encuestas son autoadministradas a través de un navegador web o aplicación móvil. En el transcurso de esta tesina se hará referencia a estos métodos utilizando los términos encuestas/abordaje telefónico y online.

Sin embargo, en los últimos años, la confianza de los resultados se ha visto afectada tanto por sesgos causados por el método de recolección de datos como por el diseño muestral elegido. En primer lugar, se destaca que los resultados obtenidos mediante el abordaje telefónico y online difieren considerablemente, tendiendo a respuestas positivas en la modalidad telefónica. Por otra parte, como diseño muestral se han utilizado muestreos no probabilísticos, particularmente muestreos por cuotas. Debido a esto, no se garantiza la representatividad de la muestra y no se puede evaluar precisión, imposibilitando conocer cuán confiables son los resultados obtenidos.

En el año 2023, desde Avalian surge el interés de llevar a cabo el proyecto de la Encuesta de Satisfacción con la colaboración de la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario a través de un convenio. Se busca presentar una estrategia disruptiva que cumpla con las aspiraciones de la encuesta actual, minimizando los sesgos y maximizando la confianza de los resultados.

Resulta fundamental elegir un muestreo de tipo probabilístico que permita obtener una muestra representativa, a partir de la cual se puedan proyectar los resultados de ésta en la población, cuantificar la incertidumbre y la validez de los resultados y eliminar sesgos de selección e intervenciones humanas.

Por otra parte, se plantea trabajar sobre los posibles sesgos generados por los diferentes métodos de recolección de datos.

Otro aspecto importante es crear un instrumento de recolección de datos simple y ágil, haciendo preguntas comprensibles y con escalas sencillas que no generen ambigüedad, y, principalmente, trabajar sobre la extensión del formulario para evitar fatiga por parte del encuestado y, por lo tanto, su abandono.

En una primera instancia, se realiza un análisis de antecedentes para poder identificar puntos de mejora y se plantean los objetivos de la encuesta actual. Luego, para desarrollar el diseño muestral se evalúa cómo estará conformada la población objetivo y se definen los dominios de estimación, las variables de interés, los niveles de precisión con los que se quiere trabajar, la forma de contacto, etcétera.

A partir de las definiciones realizadas, se considera usar un diseño muestral estratificado de GF, siendo el método de estimación el de simple expansión. Teniendo en cuenta que existe más de una variable de interés y más de un dominio de estimación, la adjudicación de la muestra sigue un enfoque multivariado y multidominio.

En una primera instancia, se planifica el relevamiento de datos en octubre del 2023, fecha habitual de realización de la encuesta. Debido al contexto general, influenciado, entre otras



cosas, por el cronograma electoral y la situación económica, se decidió postergar la recolección de datos para el año 2024.

## **2. Objetivos**

### **2.1 Objetivo general**

Estudiar y aplicar diseños muestrales para encuestas de opinión en la empresa Avalian que tengan en cuenta distintos niveles de estratificación y distintos niveles de dominios de estimación.

### **2.2 Objetivos específicos**

- Estudio de antecedentes de los diseños muestrales utilizados en el pasado en encuestas similares dentro de la empresa.
- Búsqueda bibliográfica de diseños muestrales con más de una variable de estratificación y con distintos niveles de dominios de estimación.
- Determinación de tamaños de muestra que se ajusten a niveles de precisión para cada uno de los niveles de dominios de estimación y a los costos operativos.
- Determinación del diseño muestral que cumpla con los requisitos pretendidos y planteo de posibles soluciones para el problema de la no respuesta.

### 3. Antecedentes

El análisis de antecedentes se presenta para las encuestas de satisfacción realizadas en los años 2017, 2018, 2019 y 2022. Si bien existieron ediciones anteriores, desde el año 2017 se cuenta con documentación disponible para el análisis.

Debido al impacto de la pandemia de COVID-19 sobre la población y los sistemas de salud, la encuesta de satisfacción se discontinuó en los años 2020 y 2021. Las encuestas de los años 2017 y 2019 se trabajaron con colaboradores y recursos internos de la empresa mientras que en los otros años se trabajó con distintas consultoras.

Al no contar con informes metodológicos que detallen en profundidad los diseños muestrales utilizados, el análisis de antecedentes puede llegar a ser limitado en algunos aspectos de interés. Sólo se dispone de documentos de trabajo de las versiones en cuestión (presentaciones de resultados, formularios utilizados, bases de datos, etc.).

A continuación, se presenta el análisis de antecedentes de la Encuesta de Satisfacción de Asociados de Avalian según distintos aspectos de interés:

#### 3.1 Objetivos de las encuestas anteriores

La encuesta de satisfacción, en términos generales, mantiene un mismo objetivo principal a través de los años. En cambio, los objetivos específicos se modifican en las distintas ediciones.

A través de todas las encuestas estudiadas, el objetivo principal es conocer el nivel de satisfacción general que tienen los asociados con el desempeño de Aca Salud/Avalian. Respecto a los objetivos específicos, los que se mantienen en el tiempo son:

- Medir el nivel de satisfacción entre los asociados con respecto a los diferentes puntos que componen el servicio y la atención brindada por Avalian.
- Relevar las causas de los inconvenientes ocurridos entre los asociados y Avalian.

En los primeros tres años se agrega medir la imagen de ACA Salud percibida por los asociados. A partir del 2018, se incorpora también el objetivo específico de identificar oportunidades de mejora; y desde el 2019, el de determinar el conocimiento y el uso del servicio de telemedicina E-doc, una aplicación que permite realizar consultas médicas las 24 hs a través de videollamada. En la última encuesta se busca también medir las preferencias de canales de atención: presencial, telefónico y online.

En la Tabla 1 se resumen los objetivos específicos presentes en cada encuesta.

**Tabla 1:** Objetivos específicos de las encuestas en los distintos años

Objetivo específico	2017	2018	2019	2022
Medir el nivel de satisfacción entre los asociados con respecto a los diferentes puntos que componen el servicio y la atención brindada por Avalian	X	X	X	X
Relevar las causas de los inconvenientes ocurridos entre los asociados y Avalian	X	X	X	X
Medir la imagen de ACA Salud percibida por los asociados	X	X	X	
Identificar oportunidades de mejora		X	X	X
Determinar el conocimiento y el uso del servicio de telemedicina E-doc			X	X
Medir las preferencias de canales de atención: presencial, telefónico y online				X

### 3.2 Población objetivo

En las primeras tres encuestas analizadas, la población objetivo se restringe a los titulares de los GF activos en determinada fecha, mientras que en la encuesta realizada en el año 2022, se incorporan los familiares mayores de 18 años (para las definiciones que se consideran en esta tesina ver ANEXO 1).

Debido a cuestiones legales, limitaciones de contacto o características de los asociados que pueden afectar su satisfacción y provocar sesgo, se aplican criterios de exclusión. Los mismos se modifican año a año y se presentan en la Tabla 2.

**Tabla 2:** Criterios de exclusión de la población en los distintos años

Año	Criterios de exclusión
2017	Empleados de Aca Salud, afiliados sin número telefónico y afiliados con antigüedad menor a tres meses
2018	Empleados de Aca Salud, afiliados sin número telefónico o email y afiliados que tienen credenciales de planes no comercializables
2019	Empleados de Aca Salud, afiliados sin número telefónico o email y afiliados que tienen credenciales de planes no comercializables
2022	Empleados de Avalian, afiliados sin número telefónico o email, afiliados que tienen credenciales de planes no comercializables y menores de 18 años

### 3.3 Método de selección

En todas las instancias se trabaja con diseños muestrales no probabilísticos, particularmente con muestreos por cuotas. Las variables elegidas para definir las cuotas varían en todas las ediciones.

Durante el 2017 se trabaja con región, plan, edad y una variable construida que identifica asociados migrados a nuevos planes y asociados genuinos/nativos. Este último control responde a un objetivo específico de ese año.

En el año 2018, las cuotas se plantean según sexo y edad. Además de estas cuotas, en el año 2019, se agrega el tipo de credencial y región. Respecto al año 2022, se incorpora también el tipo de segmento. Las categorías en las que se agrupa la edad varían año a año (ver ANEXO 1).

### 3.4 Forma de contacto

A excepción del año 2017 donde sólo se realizan encuestas telefónicas, en los demás años se utiliza una modalidad híbrida de encuestas telefónicas y online. En las encuestas online, el envío es masivo y se recolectan las respuestas de voluntarios hasta completar las cuotas, mientras que en las encuestas telefónicas se hace un barrido de los teléfonos hasta completar el número de encuestas deseadas. Bajo la modalidad híbrida, aproximadamente el 30% de la

muestra se obtiene de manera telefónica, mientras que el 70% restante mediante formularios online. En la documentación no se justifica el motivo de la elección de esta distribución. Se destaca que los resultados obtenidos mediante el abordaje telefónico y online difieren considerablemente, tendiendo a respuestas positivas en la modalidad telefónica.

### 3.5 Instrumento de recolección de datos

Se utilizan cuestionarios con preguntas principalmente cerradas y algunas preguntas abiertas. La extensión del formulario varía cada año, pero se destacan formularios extensos. En el año 2022 algunos asociados tenían la posibilidad de responder hasta 83 preguntas.

### 3.6 Tamaños de muestra

La Tabla 3 muestra los tamaños poblacionales y tamaños de muestras en los distintos años. En el año 2017 la muestra es de 371 encuestas, lo cual representa el 0,7% del tamaño de la población. En los últimos tres años (2018, 2019 y 2022) el tamaño de las muestras aumenta a valores entre 1.600 a 2.100 aproximadamente, llegando a representar un 3% aproximadamente de la población.

**Tabla 3:** Tamaño de la muestra y tamaño de la población en los distintos años

Año	Tamaño de la población aproximados	Número de encuestas planificadas	Número de encuestas realizadas
2017	60.000	Telefónicas: 400	Telefónicas: 371
2018	63.000	No hay información	<b>Total:</b> 1.942 Telefónicas: 600 Online: 1.342
2019	71.000	<b>Total:</b> 1.661 Telefónicas: 450 Online: 1.211	<b>Total:</b> 1.654 Telefónicas: 443 Online: 1.211
2022	94.600	<b>Total:</b> 2.000 Telefónicas: 500 Online: 1.500	<b>Total:</b> 2.109 Telefónicas: 501 Online: 1.608

### 3.7 Indicadores de satisfacción

La satisfacción de los asociados de Avalian se mide a través de cuatro temáticas: satisfacción general, recomendación, inconvenientes y recompra. Según necesidades particulares, en algunos años se incorporan nuevos aspectos.

En el cuestionario, la forma de indagar sobre estos temas varía en el tiempo. Se incluyen preguntas relacionadas a la satisfacción general y la satisfacción en determinados aspectos que componen el servicio y la atención brindada por Avalian, al nivel de recomendación de la empresa a familiares/amigos, a la presencia de inconvenientes, sus causas y resolución y si volvería a elegir Avalian como empresa de medicina prepaga.

Los indicadores elegidos para presentar los resultados también son diferentes según los años. Lo mismo sucede con las variables de clasificación consideradas para desagregar los resultados obtenidos a través de estos indicadores.

Se presentan estimaciones tanto para el total de los asociados como para la desagregación dada por algunas de las variables usadas como cuotas y el método de recolección (telefónico/online).

#### 3.7.1 Satisfacción General

El nivel de satisfacción general se mide a través de la pregunta “¿Qué tan satisfecho te encontrás con Avalian?”, con distintas opciones de respuesta según la edición. En el año 2017, se trabaja con una escala ordinal de cinco categorías con opciones que varían entre “Muy satisfecho” y “Muy insatisfecho”. Los resultados se presentan a través de una distribución porcentual, haciendo foco en el porcentaje de asociados que expresan estar satisfechos o muy satisfechos con la empresa. En cambio, las últimas tres encuestas analizadas utilizan una escala ordinal de diez categorías con valores de 1 a 10 en la que 10 es muy satisfecho y 1 para nada satisfecho. En estos casos se utiliza el promedio como indicador principal.

#### 3.7.2 Recomendación

La medición del nivel de recomendación se realiza a través de las preguntas: “¿En qué medida recomendarías a Avalian como prestador de Medicina Prepaga?” o “¿Recomendarías Aca Salud/Avalian a familiares o amigos?”. En las primeras tres encuestas analizadas, se plantean escalas ordinales con categorías que varían desde “Sí, seguro” a “Definitivamente no”. La presentación de los resultados se hace con la distribución porcentual.

Particularmente, el indicador principal es el porcentaje que respondió “Sí, seguro” o “Sí, probablemente”. En el año 2019 se incorpora adicionalmente el índice *Net Promoter Score* (NPS), el cual es el indicador principal en el año 2022. Este índice utiliza una escala ordinal que toma diez categorías donde 1 es “Seguramente NO lo recomendaría” y 10 es “Seguramente lo recomendaría”. Los encuestados son clasificados según sus respuestas en promotores (9 y 10), neutros (7 y 8) y detractores (1 a 6), y se calcula el NPS como la diferencia entre el porcentaje de promotores y detractores.

### 3.7.3 Inconvenientes

Para medir inconvenientes se utiliza la pregunta: “¿Tuviste inconvenientes con algún servicio ofrecido por Avalian?”, siendo “Sí” o “No” las respuestas posibles. Se presenta el porcentaje de asociados que tuvieron inconvenientes con Avalian. En algunas instancias se indaga sobre la causa del inconveniente (servicio y/o atención) y su resolución.

### 3.7.4 Recompra

Para evaluar recompra se pregunta: “¿Si tuviese que elegir nuevamente su cobertura médica optaría por Avalian?” y se mide el porcentaje de asociados que volverían a elegir Avalian.

### 3.7.5 Otros indicadores

En determinadas ediciones, según necesidades particulares del momento, se incorporan adicionalmente indicadores específicos que miden aspectos de interés. En el 2017 es de interés analizar la imagen de Aca Salud, para lo cual se pregunta: “Qué opinión tiene sobre la imagen de Aca Salud en general?” y se mide el porcentaje de asociados que manifiestan percibir una imagen “Buena” o “Muy buena”. En el año 2019, se incluye el índice *Customer Loyalty Index* (CLI)<sup>2</sup> para medir la lealtad de los clientes a largo plazo, incorporando los valores de satisfacción, recomendación y recompra. En la última encuesta realizada (año

---

<sup>2</sup> Es una herramienta estandarizada que mide tres preguntas de recomendación, recompra y satisfacción en una escala de 6 puntos (1=100,...,6=0) para poder clasificar al asociado en tres categorías: *Leal* (otorgan 80 a 100 puntos en las tres preguntas), *Satisfecho* (puntuaron con 60 puntos al menos una de las tres preguntas) y *En riesgo* (puntuaron con 40 o menos alguna de las 3 preguntas). El índice CLI se calcula como el promedio entre el CLI de recomendación, de recompra y de satisfacción, y varía entre 0 y 100. Un valor de CLI entre 0 y 70 indica que se está por debajo del nivel esperado de lealtad, entre 70 y 80 significa que se alcanza el valor promedio esperado y mayor a 80 que supera la lealtad esperada.



2022), se usa el índice TRI\*M<sup>3</sup> que permite medir la fortaleza de la relación con los asociados, dada por la preferencia de los mismos y el desempeño de Avalian.

### 3.8 Recomendaciones basadas en antecedentes

Con el fin de cumplir con los objetivos definidos para la nueva encuesta, en base a las experiencias pasadas se recomienda:

- Evaluar si la incorporación de los familiares mayores de 18 años a la población es una decisión acertada. Asociados pertenecientes a un mismo grupo familiar podrían tener opiniones que se vean afectadas por la experiencia del resto de los integrantes del grupo familiar. Es decir que la satisfacción del asociado no se basa sólo en su experiencia individual. Se plantean escenarios donde se indaga sobre la experiencia de los titulares, de todos los asociados o del grupo familiar.
- Trabajar sobre los posibles sesgos que estén generando diferencias entre las respuestas online y telefónicas. Es decir, buscar reducir los sesgos causados por el método de recolección de datos.
- Generar un instrumento de recolección de datos simple y ágil, haciendo preguntas comprensibles y con escalas sencillas que no generen ambigüedad. Principalmente trabajar sobre la extensión del formulario para evitar fatiga y, por lo tanto, su abandono.
- Optar por un muestreo probabilístico que garantice la representatividad de la muestra, permitiendo extrapolar los resultados obtenidos a toda la población, evaluar la precisión de los resultados y eliminar sesgos de selección.

---

<sup>3</sup> Es un indicador que resume en un solo número dos dimensiones: Performance y Preferencia. La Performance se define como el desempeño general de la compañía, considerando sólo el servicio brindado al cliente, mientras que la Preferencia agrega la perspectiva del mercado ya que contextualiza la percepción de la empresa en comparación con otras compañías. En la documentación no se dispone de información específica sobre su cálculo.

## 4. Encuesta de Satisfacción 2023

### 4.1 Materiales

Previo a la obtención de la base de datos con la cual se trabaja, se define cómo estará conformada la población objetivo. De las posibles alternativas de trabajar con titulares, con todos los asociados o con GF, se opta como unidad elemental esta última opción. Los motivos de la elección se desarrollan en el apartado “ 4.3 Población objetivo”.

Por lo tanto, se dispone de una base de datos que contiene información referida a los GF asociados a Avalian que se encuentran activos al 25/10/2023. Sobre esta población inferencial, se excluyen aquellos GF donde el titular:

- es colaborador de Avalian o cuenta con beneficio de ex-empleado,
- es menor de 18 años,
- posee credenciales de planes no comercializables,
- pertenece a una empresa particular<sup>4</sup>,
- se encuentra en el Registro Único de Postulantes (RUP),
- no supera 6 meses de antigüedad.

Se excluyen también aquellos GF que hayan incurrido en casos de falseamientos o que en el último año controlado acumulan una cuota menor a un valor mínimo de referencia<sup>5</sup>.

De esta manera, se obtiene la población objetivo sobre la cual se van a realizar inferencias.

Para cada uno de estos GF, la base de datos cuenta con la siguiente información:

- Región<sup>6</sup> a la cual pertenece: A, B, C, L, N, P
- Segmento<sup>6</sup> al cual pertenece: A, B, C, D, E
- Credencial<sup>6</sup> que posee: A, B, C, D
- Nivel de uso del grupo: Bajo – Medio - Alto<sup>7</sup>
- Cantidad de integrantes en el GF
- Indicador de si en el GF hay cónyuge
- Indicador de cantidad de datos de contacto válidos del titular: de email, de celular, de teléfono fijo

---

<sup>4</sup> Por cuestiones de privacidad se reserva el nombre de la empresa.

<sup>5</sup> Por cuestiones de confidencialidad no se especifica el valor de referencia.

<sup>6</sup> Debido a cuestiones de confidencialidad, la variable es renombrada por contener información sensible.

<sup>7</sup> Variable creada por el equipo que relaciona la cuota y el gasto acumulado en el último año calendario. Debido a cuestiones de confidencialidad no se especifica el criterio de clasificación.

- Indicador de cantidad de datos de contacto válidos del cónyuge: de email, de celular, de teléfono fijo
- Email del titular
- Email del cónyuge
- Celular del titular
- Celular del cónyuge
- Teléfono fijo del titular
- Teléfono fijo del cónyuge

En una primera instancia se trabaja con este marco muestral. Luego, a lo largo de la tesina, se presentan nuevas propuestas en cuanto a la categorización de segmento y a la exclusión de la credencial D. Se discute acerca de la posibilidad de agrupar determinadas categorías de segmento, para que se presente con sólo dos categorías: F<sup>8</sup> y G<sup>9</sup>. De esta manera, en el desarrollo se trabaja con cuatro bases de datos distintas según la inclusión o no de la credencial D y la categorización de segmento en dos o cinco categorías.

## 4.2 Objetivos de la Encuesta de Satisfacción 2023

El objetivo general es conocer el nivel de satisfacción general que tienen los asociados con el desempeño de Avalian.

Los objetivos específicos son:

- Medir el nivel de satisfacción entre los asociados con respecto a los diferentes puntos que componen el servicio y la atención brindada por Avalian.
- Relevar las causas de los inconvenientes ocurridos entre los asociados y Avalian.
- Identificar oportunidades de mejora.

## 4.3 Población objetivo

Entendiendo que el grado de satisfacción de cada asociado respecto al desempeño de la empresa es concordante entre los integrantes de un mismo GF, se propone trabajar con la experiencia de los grupos, dejando atrás las experiencias individuales de cada asociado.

---

<sup>8</sup> Surge de agrupar los segmentos B, D y E presentados con las cinco categorías.

<sup>9</sup> Surge de agrupar los segmentos A y C presentados con las cinco categorías.

Se define a la población objetivo como el conjunto de GF asociados a Avalian que se encuentren activos al 25/10/2023, aplicándose los criterios de exclusión mencionados anteriormente en la Sección 4.1.

En esta edición, a diferencia de las anteriores, se incluye a los asociados sin datos de contacto, entendiendo que sus opiniones tienen similar validez que la del resto de los asociados con datos de contacto. La recuperación de estos datos se trabaja dentro del proyecto de Calidad de Datos, por colaboradores del sector de Planeamiento Económico Financiero y Operacional en Avalian. Dado que un marco muestral completo y actualizado es fundamental para que sea válido extrapolar los resultados de la muestra en la población, se asume desde ese proyecto el compromiso de obtener la información faltante.

La incorporación de los criterios de exclusión para asociados que pertenecen a la empresa particular o se encuentran en el RUP se debe a que, dadas las condiciones particulares del grupo, la opinión de los mismos puede verse afectada tanto de manera positiva como negativa. En el caso de los GF que no superan los 6 meses de antigüedad, se considera que no cuentan con el recorrido suficiente para medir satisfacción. En caso de incorporarlos se corre el riesgo de que su opinión sólo esté vinculada al proceso de contratación y no al servicio y la atención recibida.

El integrante que responderá el formulario contando la experiencia del GF es el titular del grupo, o en caso de corresponder, el cónyuge. Esto se debe a que se considera que son los informantes mejor calificados dentro del grupo, lo cual se debe a su responsabilidad en el pago de cuotas y en la decisión de mantener el servicio.

En Avalian, al 25/10/2023 hay una totalidad de **124.333 GF** activos. Aplicando los criterios de exclusión, se tiene una población objetivo de **88.839 GF** (71.5% respecto a la totalidad de GF).

La información correspondiente a los datos de contacto disponibles para los titulares pertenecientes a la población objetivo se presenta en la Tabla 4.

**Tabla 4:** Cantidad y porcentaje de titulares según datos de contacto disponibles

Total titulares	Con mail	Con teléfono (celular y/o fijo)	Con mail y/o teléfono	Sin mail y sin teléfono
88.839	84.520 (95.1%)	85.727 (96.5%)	87.736 (98.8%)	1.103 (1.2%)

En la Tabla 4 puede observarse que de los 88.839 titulares, se tienen 87.736 GF (98.8%) con titulares que tienen al menos un dato de contacto válido.

Además, de los 88.839 GF pertenecientes a la población objetivo, se tienen **23.817 GF** (26.8%) en los cuales dentro de sus integrantes existe un cónyuge. La información de los datos de contacto disponibles para los cónyuges se presenta en la Tabla 5.

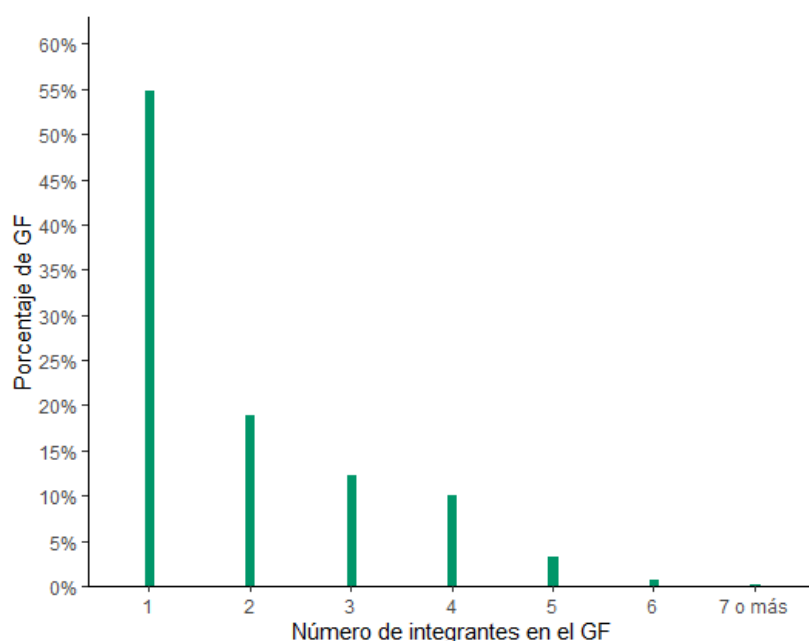
**Tabla 5:** Cantidad y porcentaje de cónyuges en los GF según datos de contacto disponibles

Total GF con cónyuge	Con mail	Con teléfono (celular o fijo)	Con mail y/o teléfono	Sin mail y sin teléfono
23.817	9.029 (37.9%)	5.604 (23.5%)	9.761 (41.0%)	14.056 (59.0%)

En la Tabla 5 puede observarse que hay un 59.0% de GF con cónyuge que no tienen información de contacto de éste.

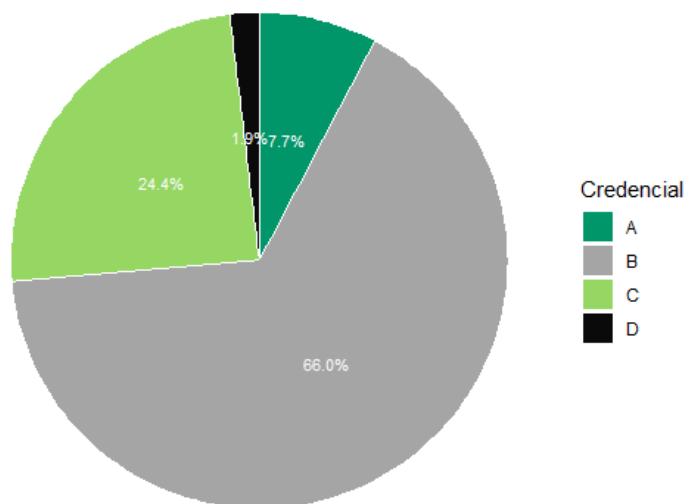
A continuación, se presenta el análisis descriptivo de la población objetivo en función de variables como cantidad de integrantes en el GF, región, segmento, nivel de uso y credencial. En primer lugar, interesa ver cómo se distribuye la cantidad de GF según el número de integrantes que tiene el GF. En la Figura 1 puede observarse que un 54.5% de los GF son de un único asociado, es decir, están compuestos únicamente por titulares. Existe un 18.9% de GF que tienen dos integrantes, mientras que hay un 12.4% y 10.1% de GF con 3 y 4 asociados respectivamente. Además, hay un porcentaje muy pequeño de GF con 5 integrantes o más.

**Figura 1.** Distribución de GF según el número de integrantes en el GF al 25/10/2023



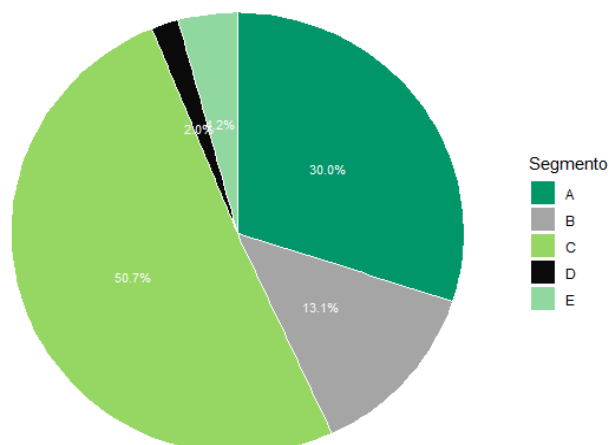
Con respecto a la distribución según credencial, una gran parte de la población tiene la credencial B (66.0%), mientras que un 24.4% tiene la credencial C, un 7.7% la credencial A y un 1.9% la credencial D (Figura 2).

**Figura 2.** Distribución de GF según credencial al 25/10/2023



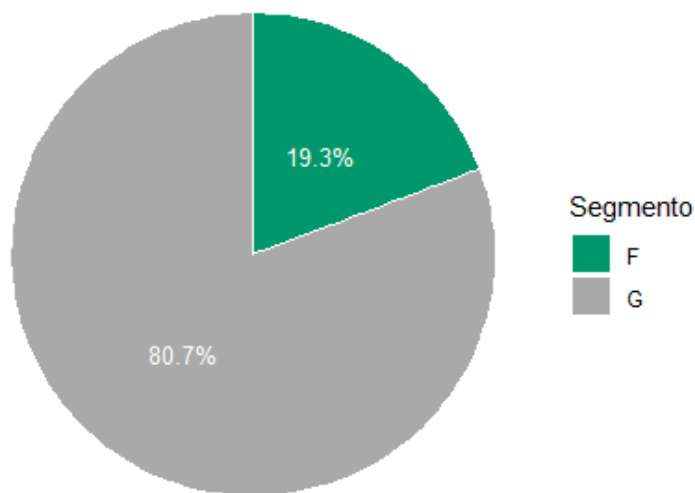
En la figura 3 puede observarse la distribución de los GF según las cinco categorías de segmento. La mayor parte de la población se concentra en el segmento C (50.7%) y en el A (30.0%), mientras que el segmento B, E y D tienen un 13.1%, 4.2% y 2.0% de GF respectivamente.

**Figura 3.** Distribución de GF según segmento (5 categorías) al 25/10/2023



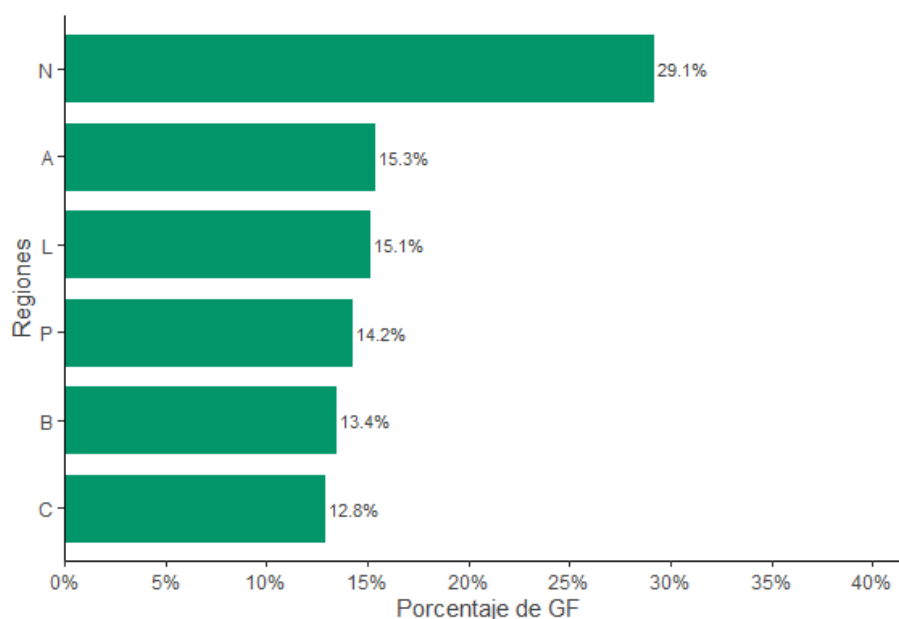
En la figura 4 se presenta la distribución de GF según las dos categorías de segmento F y G, que surgen de agrupar los segmentos B, D y E en el caso de F, y A y C en el caso de G. El 80.7% de los GF pertenecen al segmento G, mientras que el 19.3% restante al F.

**Figura 4.** Distribución de GF según segmento (2 categorías) al 25/10/2023



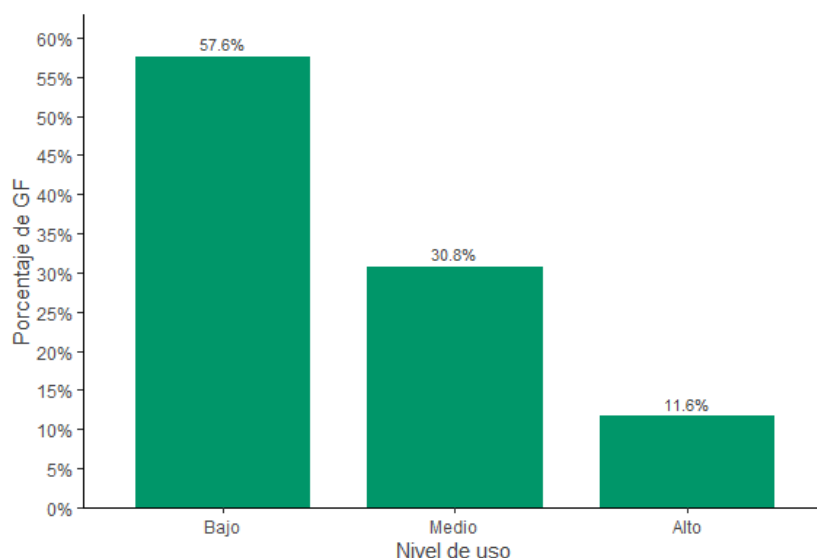
Respecto a la distribución según región, en la Figura 5 se observa que un 29.1% de los GF son de la región N, mientras que el resto se distribuye aproximadamente de manera similar entre el resto de las regiones (valores entre el 13% y 15% aproximadamente).

**Figura 5.** Distribución de GF según región al 25/10/2023



En la Figura 6 se presenta la distribución de GF según el nivel de uso. El 57.6% de los GF presentan un nivel de uso bajo, mientras que un 30.8% y un 11.6%, medio y alto respectivamente.

**Figura 6.** Distribución de GF según el nivel de uso al 25/10/2023



#### 4.4 Parámetros de interés

El nivel de satisfacción general y de recomendación son los indicadores principales. Para medir satisfacción general el parámetro elegido es la proporción de GF que mencionan estar satisfechos o muy satisfechos con Avalian. Se releva la información por medio de la pregunta “Considerando la experiencia del grupo familiar, ¿Qué tan satisfecho te encontrarás con Avalian?”, utilizando una escala ordinal de cinco categorías que van desde “Muy satisfecho” a “Muy insatisfecho”.

En cuanto a recomendación, el indicador elegido para medir este aspecto es el NPS. Aunque se trata de la diferencia de dos proporciones, por cuestiones de definición del diseño muestral se consideran dos parámetros de interés principal:

- la proporción de GF promotores de la empresa,
- la proporción de GF detractores de la empresa.

La información se releva a través de la pregunta “¿En qué medida recomendarías Avalian a algún familiar o amigo?” con una escala ordinal de 1 a 10 donde 1 es “Seguramente NO lo recomendaría” y 10 es “Seguramente lo recomendaría”. En función de sus respuestas, los GF son clasificados en promotores (9 y 10), neutros (7 y 8) y detractores (1 al 6), y se calcula el NPS como la diferencia entre el porcentaje de promotores y detractores.



Recompra e inconvenientes se miden en la encuesta, pero no son abordados como parámetros principales. Es decir, sólo satisfacción general y recomendación se contemplan como parámetros al momento de definir el diseño muestral.

## 4.5 Dominios de interés

Se definen a continuación las subpoblaciones o niveles de la población sobre los cuales se quieren realizar estimaciones con un cierto nivel de precisión.

Se opta por utilizar cuatro tipos de dominios de estimación: región comercial, credencial, segmento y nivel de uso.

En una primera instancia, se proponen los siguientes dominios de estimación:

- Región: A, B, C, L, N, P
- Credencial: A, B, C, D
- Segmento: A, B, C, D, E
- Nivel de uso: Bajo, Medio, Alto

A lo largo de la tesina se proponen distintos escenarios en cuanto a la reducción de categorías dentro del tipo de dominio credencial y el segmento, pero siempre trabajando con los cuatro tipos de dominios definidos con anterioridad.

## 4.6 Cuestionario y forma de contacto

Dado que la extensión y claridad del cuestionario impacta sobre la tasa de respuesta y la calidad de la información brindada por el encuestado, se realizan modificaciones sobre el que fue utilizado en el año 2022. Se redujo la extensión de éste estableciendo un mínimo de doce preguntas y un máximo de catorce. Además, se simplifica la formulación de las preguntas y las escalas utilizadas. El cuestionario se presenta en el ANEXO 2.

A diferencia de años anteriores donde el encuestado respondía por teléfono o mail según le correspondiese, este año se busca obtener la respuesta del GF por cualquier medio de contacto. El diseño muestral se define de manera tal que la forma de contacto no genere sesgo significativo en la respuesta.

Siguiendo un orden preestablecido y definido en conjunto con el área de Comunicación al Asociado dentro de Avalian, los medios de contacto elegidos son: email, SMS, WhatsApp,

llamados telefónicos e intervención de las agencias<sup>10</sup>. Además, se trabaja en una estrategia de comunicación que contribuya a aumentar las tasas de respuesta (ver ANEXO 3).

---

<sup>10</sup> Entidad tercerizada que establece una relación contractual con Avalian, mediante la cual comercializa sus productos y/o brinda servicios de atención a sus asociados, a cambio de una retribución económica.

## 5. Diseño muestral

En base a las decisiones tomadas y frente al requerimiento de obtener estimaciones de tres parámetros de interés en las categorías de cuatro tipos de dominios de estimación, se opta por utilizar un muestreo estratificado de GF, con adjudicación de la muestra siguiendo un enfoque multivariado y multidominio. Este diseño se enmarca en los muestreos probabilísticos. El método de estimación utilizado es por simple expansión. El desarrollo de la metodología se presenta en el apartado 5.1.

### 5.1 Determinación del tamaño de la muestra y su adjudicación

En la mayoría de las encuestas, se tiene el objetivo de brindar estimaciones precisas de diferentes parámetros en distintos dominios de interés. Además, existen restricciones presupuestarias y logísticas para el diseño y desarrollo de la encuesta, debiendo ésta planearse adecuadamente para poder brindar estimaciones de buena calidad de los parámetros de interés. Decisiones como el tamaño muestral, el número de estratos y la constitución de los mismos, y la adjudicación de la muestra en los estratos, no son triviales y pueden afectar la calidad de los resultados.

La metodología que se utiliza para determinar el tamaño de la muestra y su adjudicación a través de los estratos busca maximizar la calidad de los resultados dentro de las restricciones presupuestarias, contemplando múltiples variables y múltiples dominios. Es decir, la adjudicación de la muestra a través de los estratos es óptima.

Este concepto de adjudicación óptima fue trabajado inicialmente por Tschuprow (1923) y luego formalizado por Neyman (1934), quienes buscaban minimizar la variancia de las estimaciones en diseños estratificados con una sola variable. Posteriormente, el algoritmo de Bethel (1989) extendió la adjudicación de Neyman al caso multivariado.

En el presente apartado se desarrolla la metodología. En primer lugar, se definen los estratos y se presenta el concepto de adjudicación óptima. Luego, se desarrolla la adjudicación multivariada de la muestra en muestreos estratificados que requieren estimaciones de múltiples variables únicamente para la población completa. Finalmente, se desarrolla el caso multivariado y multidominio utilizado en esta tesina. Además, se presenta un paquete estadístico del software R utilizado para determinar el tamaño de la muestra y su adjudicación a través de los estratos bajo el diseño muestral en cuestión: *R2BEAT* (“*R ‘to’ Bethel Extended Allocation for Two-stage sampling*”).

### 5.1.1 Definición de estratos

El primer paso al trabajar con un muestreo estratificado es dividir a la población en estratos formados de acuerdo a variables conocidas a priori para todas las unidades de la población. En el caso de variables cualitativas, una de las maneras de construir los estratos es como combinación de las categorías de los tipos de dominios de estimación. En esta tesina, se denomina tipo de dominio a la credencial, región, segmento y nivel de uso, y dominios a las categorías correspondientes. Por ejemplo, un estrato surge de la combinación de la credencial C, nivel de uso Bajo, región B y segmento A (C-Bajo-B-A). Dado que se tienen cuatro credenciales, tres niveles de uso, seis regiones y cinco segmentos, se tienen 360 estratos. Sin embargo, dado que hay estratos que no contienen ningún GF, resultan 294 estratos.

Debe tenerse en cuenta que a medida que se tengan más dominios de estimación o más tipos de dominios, habrá un mayor número de estratos y por lo tanto puede generar un mayor tamaño de muestra para poder asegurar al menos dos unidades por estrato.

### 5.1.2 Adjudicación óptima

Para el cálculo del tamaño de la muestra se establecen restricciones respecto al error máximo deseado de las estimaciones de interés, para cada uno de los dominios. Además, se determina la asignación en cada estrato bajo el concepto de adjudicación óptima.

El problema de adjudicación óptima fue estudiado inicialmente por Tschuprow (1923) y formalizado por Neyman (1934), quienes desarrollaron la teoría para el caso univariado. Desde entonces, diversos investigadores han tratado con el caso multivariado, proponiendo distintas soluciones, las cuales pueden clasificarse en dos grupos: el enfoque promedio de las variancias y la programación convexa.

El primer enfoque se basa en hallar la adjudicación óptima para un promedio ponderado de las variancias en cada estrato, definiendo previamente un peso para cada variable considerada. Dalenius (1953), Yates (1960), Folks y Antle (1965), Hartley (1965) y Kish (1976) discutieron métodos referidos a este enfoque.

Por otra parte, el segundo enfoque consiste en especificar restricciones de desigualdad para cada variancia y, minimizando los costos de adjudicación, hallar la solución que satisfaga todas las restricciones. Para ello se utiliza la programación convexa, la cual es una rama de la optimización matemática que se centra en resolver problemas donde la función objetivo y las restricciones cumplen con la propiedad de convexidad. Dalenius (1957), Yates (1960), Kokan (1963), Hartley (1965), Kokan y Khan (1967), Chatterjee (1968, 1972), Huddleston,

Claypool y Hocking (1970), Bethel (1985) y Chromy (1987) estudiaron métodos clasificados en este enfoque. La presente metodología es una generalización del algoritmo de Bethel contenida en este enfoque.

Cada enfoque tiene sus ventajas y desventajas. El método del promedio ponderado es computacionalmente simple, intuitivo y puede resolverse asumiendo un costo fijo. Sin embargo, la elección de los pesos es arbitraria y las propiedades de optimalidad no son tan claras. Por su parte, el método que utiliza programación convexa brinda la solución óptima, aunque el costo resultante puede que no sea aceptable requiriendo continuar con la búsqueda de una solución óptima que se encuentre dentro de los requerimientos económicos (Bethel 1989).

Bethel (1989) presenta una expresión para la adjudicación óptima de la muestra estableciendo restricciones de desigualdad en términos de multiplicadores de Lagrange. Obtiene dos resultados que reflejan las ventajas del segundo enfoque. El primero refiere a la posibilidad de escalar los resultados obtenidos bajo la adjudicación óptima multivariada de manera de obtener una adjudicación que es óptima bajo restricciones proporcionales a las originales. Por lo tanto, si la solución óptima es muy costosa, puede reducirse proporcionalmente al costo considerado aceptable. Las consecuencias en la precisión de las estimaciones pueden determinarse directamente.

El segundo resultado trata de la utilización de los “precios sombra”, cantidades que muestran la sensibilidad de los costos respecto a las restricciones de las variancias, para evaluar la rentabilidad del diseño muestral. Se obtienen mediante las derivadas parciales de la función de costos de adjudicación con respecto a las restricciones de las variancias.

### 5.1.3 Adjudicación multivariada: estimaciones para la población completa

Para presentar la teoría referida a la adjudicación multivariada y multidominio, se opta por desarrollar primero el caso multivariado en el cual las estimaciones se calculan únicamente para la población completa. Dicho caso fue estudiado por Bethel, por lo que se presenta a continuación su algoritmo, el cual es convergente, sencillo de programar y fácil de usar debido a que no requiere la especificación de valores iniciales. Luego, se extiende al caso multidominio.

#### 5.1.3.1 Modelo de adjudicación

Bajo un muestreo estratificado de selección de unidades con igual probabilidad y sin reemplazo, la variancia del estimador del total de una variable de interés genérica se expresa:

$$var(\hat{Y}_j) = V_j' = V_j + V_{0j} = \sum_{h=1}^H \frac{N_h^2}{n_h} s_{j,h}^2 - \sum_{h=1}^H N_h s_{j,h}^2$$

donde:

$\hat{Y}_j$ : total estimado para la j-ésima variable de interés, con  $j = 1, \dots, J$

$N_h$ : tamaño del estrato  $h$ , con  $h = 1, \dots, H$

$n_h$ : tamaño muestral del estrato  $h$

$s_{j,h}^2$ : variancia muestral de la variable  $j$  en el estrato  $h$

$V_j$ : parte de la variancia influenciada por la adjudicación

$V_{0j}$ : parte de la variancia que no es influenciada por la adjudicación.

Se define, de forma lineal, una función de costos:

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

donde  $C_0$  es el costo fijo de la encuesta que no depende del tamaño de la muestra ni de la adjudicación,  $C$  es el costo variable y  $C_h$  con  $h = 1, \dots, H$  es el costo por unidad muestral en el estrato  $h$ .

El problema de adjudicación tratado bajo el segundo enfoque, en forma general, consiste en minimizar la función  $C$  bajo las siguientes restricciones:

$$V_j' \leq V_j^* \quad j = 1, \dots, J$$

siendo  $V_j^*$  la máxima variancia especificada del estimador del total de la j-ésima variable.

Bajo esta metodología particular y por la simplicidad de su resolución, la función de costos  $C$  es reformulada por:

$$C = f(\mathbf{X}) = \sum_{h=1}^H \frac{C_h}{X_h}, \quad C_h > 0, \quad h = 1, \dots, H \quad (1)$$

donde:

$$X_h = \begin{cases} \frac{1}{n_h} & \text{si } n_h \geq 1 \\ \infty & \text{en otro caso} \end{cases}$$

y  $\mathbf{X} = (X_1, \dots, X_H)'$ .

El término constante para costos fijos no afecta la minimización, por lo que no se lo considera para simplificar la notación.

Las restricciones  $V_j' \leq V_j^*$  son de la forma:

$$\sum_{h=1}^H a_{h,j} X_h \leq 1, \quad X_h > 0, \quad j = 1, \dots, J \quad (2)$$

donde  $a_{h,j} = \frac{N_h^2 S_{hj}^2}{V_j^* - V_{0j}} \geq 0$ .

El problema de adjudicación óptima consiste entonces en:

minimizar  $f(\mathbf{X})$

sujeto a  $\mathbf{a}_j' \mathbf{X} \leq 1$ ,

donde  $\mathbf{a}_j$  es el  $j$ -ésimo vector columna de la matriz  $\mathbf{A} = \{\mathbf{a}_{h,j}\}$ .

La resolución de este sistema permite obtener los tamaños muestrales en los estratos ( $n_h$ ) que satisfacen las precisiones establecidas.

#### 5.1.3.2 Problema de adjudicación óptima

En el caso univariado, donde  $J = 1$ , la adjudicación óptima es conocida. El mínimo de (1) sujeto a  $\mathbf{a}_1' \mathbf{X} \leq 1$ , siendo todas las componentes del vector  $\mathbf{X}$  mayores a cero, denotado por  $\mathbf{X}^*$ , para un estrato  $h$ , es:

$$X_h^* = \begin{cases} \sqrt{C_h} / \left( \sqrt{a_{h,1}} \sum_{k=1}^H \sqrt{C_k a_{k,1}} \right) & \text{si } a_{h,1} > 0, \quad 1 \leq h \leq H \\ \infty \text{ en otro caso} \end{cases} \quad (3)$$

A continuación, se extiende (3) para  $J > 1$ . Dado que la función  $f(\mathbf{X})$  en (1) es estrictamente convexa para componentes de  $\mathbf{X}$  mayores a cero ( $\mathbf{X} > \mathbf{0}$ ) y las restricciones en (2) son lineales, se aplican los resultados básicos de la programación convexa. Uno de ellos es que siempre existe una solución óptima  $\mathbf{X}^*$ , lo cual fue demostrado por Kokan y Khan (1967). Entonces, por el teorema de Kuhn-Tucker (1951), se demuestra que existen multiplicadores de lagrange  $\lambda_j \geq 0$ , con  $j = 1, \dots, J$ , tal que:

$$\nabla f(\mathbf{X}^*) + \sum_{j=1}^J \lambda_j \mathbf{a}_j = 0 \quad (4)$$

donde  $\nabla$  denota el gradiente, y

$$\lambda_j (\mathbf{a}_j' \mathbf{X}^* - 1) = 0, \quad \text{con } j = 1, \dots, J. \quad (5)$$

Si las componentes del vector  $\mathbf{X}$  son mayores a cero, satisface  $\sum_{j=1}^J \lambda_j \mathbf{a}_j' \mathbf{X} \leq \sum_{j=1}^J \lambda_j$ , entonces, combinando (4) y (5):

$$-\mathbf{X}' \nabla f(\mathbf{X}^*) = \sum_{j=1}^J \lambda_j \mathbf{a}_j' \mathbf{X} \leq \sum_{j=1}^J \lambda_j = \sum_{j=1}^J \lambda_j \mathbf{a}_j' \mathbf{X}^* = -\mathbf{X}^{*'} \nabla f(\mathbf{X}^*). \quad (6)$$

Por convexidad,

$$f(\mathbf{X}) - f(\mathbf{X}^*) \geq (\mathbf{X} - \mathbf{X}^*)' \nabla f(\mathbf{X}^*) \quad \text{para todo } \mathbf{X} > \mathbf{0} \text{ y } \mathbf{X}^* > \mathbf{0}.$$

Entonces, por (6):

$$f(\mathbf{X}) - f(\mathbf{X}^*) \geq (\mathbf{X} - \mathbf{X}^*)' \nabla f(\mathbf{X}^*) \geq 0.$$

Y por lo tanto,  $\mathbf{X}^*$  es el mínimo de  $f(\mathbf{X})$  sujeto a las condiciones:

$$\sum_{j=1}^J \lambda_j \mathbf{a}_j' \mathbf{X} \leq \sum_{j=1}^J \lambda_j \quad \text{para todo } \mathbf{X} > \mathbf{0}.$$

Dado que la incorporación de constantes multiplicativas positivas no afecta la minimización de  $f$ ,  $\mathbf{X}^*$  también minimiza  $f(\mathbf{X})$  sujeto a las restricciones:  $\sum_{j=1}^J \alpha_j^* \mathbf{a}_j' \mathbf{X} \leq 1$  y  $\mathbf{X} > \mathbf{0}$ , donde

$$\alpha_j^* = \lambda_j / \sum_{j=1}^J \lambda_j.$$

Partiendo de (3), Bethel demostró que la expresión para la adjudicación óptima multivariada

consiste en aplicar la suma ponderada  $\sum_{j=1}^J \alpha_j^* \mathbf{a}_j$ :

$$X_h^* = \begin{cases} \sqrt{C_h} / \left( \sqrt{\sum_{j=1}^J \alpha_j^* a_{h,j}} \sum_{k=1}^H \sqrt{C_k \sum_{j=1}^J \alpha_j^* a_{k,j}} \right) & \text{si } \sum_{j=1}^J \alpha_j^* a_{h,j} > 0, \quad 1 \leq h \leq H \\ \infty & \text{en otro caso} \end{cases} \quad (7)$$



Para determinar simultáneamente los valores óptimos  $X_h^*$  y  $\alpha_j^*$  se usan algoritmos numéricos.

Dado que  $\mathbf{X}^*$  minimiza  $f(\mathbf{X})$  sujeto a  $\mathbf{a}_j' \mathbf{X} \leq 1$  con  $\mathbf{X} > \mathbf{0}$  para  $1 \leq j \leq J$ , entonces  $m\mathbf{X}^*$  minimiza  $f(m\mathbf{X})$  sujeto a  $\mathbf{a}_j'(m\mathbf{X}) \leq m$  con  $\mathbf{X} > \mathbf{0}$  para  $1 \leq j \leq J$ . Por lo tanto, es posible escalar las restricciones en las variancias (CV's) a un factor  $m$  (o  $\sqrt{m}$ ) si los costos de la encuesta son muy elevados.

### 5.1.3.3 Sensibilidad de los costos de la encuesta respecto a las restricciones de variancias

Resulta útil conocer cómo la solución óptima se comporta frente a cambios en las restricciones de las variancias. En la práctica, esto sirve para determinar el impacto que tienen en los costos operacionales de la encuesta, cambios en la precisión de las estimaciones. Es posible detectar cambios pequeños en la precisión que pueden reducir sustancialmente los costos. Bethel demostró que, para ello, pueden utilizarse los “precios sombra” dados por las derivadas parciales de la función de costos respecto a cada restricción de variancia. La fórmula para la  $k$ -ésima restricción de variancia es  $\partial f(\mathbf{X}^*)/\partial V_k^*$ , donde  $f(\mathbf{X}^*)$  es el costo de la adjudicación óptima y  $V_k^*$  el límite superior de la  $k$ -ésima restricción de variancia.

### 5.1.3.4 Algoritmos computacionales

En esta sección se profundiza acerca de los algoritmos numéricos utilizados para hallar tanto  $\mathbf{X}^*$  como  $\alpha_k^*$ . Estos valores óptimos son obtenidos a través del promedio ponderado  $\sum_{j=1}^J \alpha_j^* \mathbf{a}_j$ .

Se define  $\delta_{h,j}$  como:

$$\delta_{h,j} = \begin{cases} 1 & \text{si } h = j \\ 0 & \text{si } h \neq j \end{cases}$$

Se define  $\bar{\mathbf{X}}(\boldsymbol{\alpha})$ , siendo  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_J)'$ , como:

$$\bar{\mathbf{X}}_h(\boldsymbol{\alpha}) = \begin{cases} \sqrt{C_h} / \left( \sqrt{\sum_{j=1}^J \alpha_j a_{h,j}} \sum_{k=1}^H \sqrt{C_k \sum_{j=1}^J \alpha_j a_{h,j}} \right) & \text{si } \sum_{j=1}^J \alpha_j a_{h,j} > 0, \quad 1 \leq h \leq H \\ \infty & \text{en otro caso} \end{cases}$$

Notar que  $\bar{\mathbf{X}}(\boldsymbol{\alpha}^*) = \mathbf{X}^*$ .

El algoritmo iterativo para hallar  $\mathbf{X}^*$  se define como:

1) Tomar  $\alpha_j^{(1)} = \delta_{1j} \quad 1 \leq j \leq J$

2) Para  $n \geq 2$ , hallar un índice  $k$  tal que:

$$(\mathbf{a}_k - \mathbf{a}_j)' \bar{\mathbf{X}}(\boldsymbol{\alpha}^{(n)}) \geq 0, \quad 1 \leq j \leq J.$$

La solución de asignación óptima actual viola la restricción con el mayor margen. Si  $\mathbf{a}_k' \bar{\mathbf{X}}(\boldsymbol{\alpha}^{(n)}) \leq 1$ , terminar el algoritmo. De otra manera, hallar  $t^{(n)} \in (0, 1)$  tal que

$$f(\bar{\mathbf{X}}(t^{(n)}\delta_k + (1 - t^{(n)})\boldsymbol{\alpha}^{(n)})) \geq f(\bar{\mathbf{X}}(t\delta_k + (1 - t)\boldsymbol{\alpha}^{(n)})), \quad \forall t \in [0, 1]$$

3) Tomar  $\alpha_j^{(n+1)} = t^{(n)}\delta_{kj} + (1 - t^{(n)})\alpha_j^{(n)}$

4) Terminar cuando  $|\alpha_j^{(n+1)} - \alpha_j^{(n)}| < \varepsilon$  con  $1 \leq j \leq J$ , siendo  $\varepsilon$  un criterio de convergencia predeterminado.

El algoritmo toma como valor inicial  $\alpha_1$ , lo cual es completamente arbitrario. Podría utilizarse cualquiera de los  $\alpha_j$  con  $1 \leq j \leq J$ . En la práctica, una buena opción es la restricción para la cual la adjudicación óptima (fórmula (3)) tiene el mayor costo.

Bethel (1985) muestra que el algoritmo converge rápidamente en la mayoría de los problemas. Chromy (1987) presenta un algoritmo incluso más simple y de convergencia más rápida, que tampoco requiere valores iniciales. Aunque no exista prueba formal de convergencia, en la práctica se han observado buenas propiedades de la misma. Se adapta a la notación de la siguiente manera:

$$\alpha_j^{(1)} = 1/J$$

y en el caso en el que  $n \geq 2$ :

$$\alpha_j^{(n)} = \alpha_j^{(n-1)} \left( \mathbf{a}_j' \bar{\mathbf{X}}(\boldsymbol{\alpha}^{(n-1)}) \right)^2 / \sum_{j=1}^J \alpha_j^{(n-1)} \left( \mathbf{a}_j' \bar{\mathbf{X}}(\boldsymbol{\alpha}^{(n-1)}) \right)^2, \quad 1 \leq j \leq J.$$

#### 5.1.4 Adjudicación multivariada: estimaciones para múltiples dominios

En la mayoría de las encuestas se requieren estimaciones no sólo para la población completa sino también para ciertas subpoblaciones o dominios de estudio. Estos dominios se corresponden con una partición de la población bajo estudio dada por un tipo de dominio. Por ejemplo, un tipo de dominio es la credencial y los dominios A, B, C y D. A su vez, las

estimaciones pueden ser requeridas para más de un tipo de dominio, siendo particiones alternativas de la misma población.

En este caso, Falorsi et al. (1998) realiza una generalización de la solución hallada previamente. Esto se debe a que el tamaño de la muestra debe ser tal que satisfaga la precisión requerida para cada uno de los dominios y variables de interés.

Se define:

$d$ : tipo de dominio genérico, con  $d = 1, \dots, D$

$k_d$ : dominio genérico de tipo  $d$ , con  $k_d = 1, \dots, K_d$

$H_{k_d}$ : número de estratos pertenecientes al dominio  $k_d$

La función objetivo (1) es la misma, mientras que el sistema de restricciones es:

$$\sum_{h=1}^{H_{k_d}} \frac{N_h^2}{n_h} S_{j,h}^2 - \sum_{h=1}^{H_{k_d}} N_h S_{j,h}^2 \leq V_{j,k_d}^*, \quad j = 1, \dots, J; \quad d = 1, \dots, D; \quad k_d = 1, \dots, K_d \quad (8)$$

donde  $V_{j,k_d}^*$  es el límite superior de la variancia muestral de la estimación del total de la variable  $j$  en el dominio  $k_d$ .

De manera similar a lo realizado anteriormente, (8) puede reescribirse como:

$$\sum_{h=1}^H a_{j,k_d,h} X_h \leq 1, \quad j = 1, \dots, J; \quad d = 1, \dots, D; \quad k_d = 1, \dots, K_d$$

donde  $a_{j,k_d,h} = \frac{N_h^2 S_{j,h}^2 \delta_{k_d,h}}{\sum_{h=1}^H N_h S_{j,h}^2 \delta_{k_d,h} + V_{j,k_d}^*}$  (9)

siendo  $\delta_{k_d,h} = 1$  si  $h \in k_d$  o  $\delta_{k_d,h} = 0$  en otro caso.

Al definir un índice  $r$  cuyos valores estén en correspondencia con los valores del vector identificado por los tres índices  $(d, k_d, j)$ , el sistema de restricciones es:

$$\sum_{h=1}^H a_{r,h} X_h \leq 1, \quad r = 1, \dots, R \text{ donde } R = J \sum_{d=1}^D K_d \quad (10)$$

siendo totalmente equivalente a (2).

Retomando (7), la solución óptima que minimiza (1) bajo las restricciones (10) es:

$$X_h^* = \sqrt{C_h} / \left( \sqrt{\sum_{r=1}^R \alpha_r^* a_{h,r}} \sum_{k=1}^H \sqrt{C_k \sum_{r=1}^R \alpha_r^* a_{k,r}} \right) \quad (11)$$

siendo  $\alpha_r^* = \lambda_r / \sum_{r=1}^R \lambda_r$ .

De esta manera, se obtienen los tamaños muestrales de cada estrato que satisfacen los requerimientos de precisión.

#### 5.1.4.1 Algoritmos resolutivos

Esta metodología propone una generalización del algoritmo de Bethel para el caso en el que hay múltiples dominios.

El algoritmo halla la solución óptima iterativamente. Se comienza con una solución inicial ( $v = 1$ ) que coincide con la solución óptima en el caso univariado para la primera variable en el primer dominio ( $r = 1$ ). Con esta solución, la función objetivo asume un valor pequeño y las restricciones restantes  $r = 2, \dots, R$  no son satisfechas. A medida que se procede con los pasos siguientes ( $v = 2, \dots, V$ ), el tamaño de muestra aumenta por lo que la función objetivo asume valores mayores  $f(\mathbf{X}^{(v)}) \geq f(\mathbf{X}^{(v-1)})$  y se logran satisfacer todas las restricciones.

Bethel demostró que el algoritmo converge y se puede determinar  $\alpha^*$  y  $\mathbf{X}^*$  simultáneamente de manera que  $0 \leq f(\mathbf{X}^{(v)}) \leq f(\mathbf{X}^*)$ .

Sin embargo, debido a la complejidad computacional de este algoritmo, especialmente en el caso de múltiples dominios, se usa el algoritmo propuesto por Chromy (1987) cuya implementación es más reciente y converge a la solución óptima más rápidamente. Este algoritmo se ilustra a continuación.

Sea  $\mathbf{A} = \{a_{r,h}\}$  la matriz de orden  $R$  y  $H$ , cuyos elementos están definidos por (9) y siendo  $a_r$  la  $r$ -ésima fila de  $\mathbf{A}$ . El algoritmo Chromy es un algoritmo iterativo cuyo primer paso consiste en calcular el valor de  $\mathbf{X}$  de acuerdo a (11), estableciendo cada elemento de  $\alpha$  igual a  $\frac{1}{R}$ . Si esta solución satisface todas las restricciones, el algoritmo para. Caso contrario, el algoritmo calcula  $\mathbf{X}^{(v)}$  en correspondencia al vector  $\alpha^{(v)}$  cuyo elemento genérico está dado por la siguiente expresión:

$$\alpha_r^{(v)} = \frac{\alpha_r^{(v-1)} (a_r \mathbf{X}^{(v-1)})^2}{\sum_{r=1}^R \alpha_r^{(v-1)} (a_r \mathbf{X}^{(v-1)})^2} \quad 1 \leq r \leq R$$

donde  $\mathbf{X}(\alpha^{(v-1)})$  denota el valor  $\mathbf{X}$  obtenido en (6) estableciendo  $\alpha = \alpha^{v-1}$ .

Estos algoritmos no aseguran que la solución óptima satisfaga  $n_h \leq N_h$  para todo  $h$ , lo cual fue tratado y solucionado por el software *MAUSS-R* y por lo tanto, contenido en el paquete

*R2BEAT*. Estos programas brindan la posibilidad de especificar que se haga un censo en aquellos estratos en los cuales  $n_h > N_h$ , calculando el tamaño de la muestra bajo estas condiciones.

### 5.1.5 Paquete R2BEAT del software R

El problema de adjudicación multivariado y multidominio que se presenta cuando se requieren estimaciones para más de una variable de interés y más de un dominio de estimación fue abordado computacionalmente por Fasulo et al. (2021) a través del paquete estadístico del software R llamado *R2BEAT*. Este paquete, que brinda una generalización del algoritmo de Bethel, sirve para determinar el tamaño de la muestra y su adjudicación a través de ciertos estratos en diseños muestrales estratificados multivariados y multidominios en una y dos etapas. Es una extensión de *MAUSS-R* ("Multivariate Allocation of Units in Sampling Surveys") (Barcaroli et al., 2020), software creado anteriormente para diseños muestrales estratificados multivariados y multidominios en una etapa. *R2BEAT* incorpora la solución brindada por *MAUSS-R* a través de la función *beat.1st*, pero permite además determinar la adjudicación de la muestra en diseños muestrales complejos en dos etapas mediante la función *beat.2st*. En el desarrollo de esta tesina se trabaja sólo con el caso a una etapa.

#### 5.1.5.1 Bases de datos requeridas

Para obtener una muestra compleja usando *R2BEAT* deben primero prepararse los archivos de entrada requeridos por la función. El paquete requiere que se especifiquen dos *dataframes* referidos a ciertas características de la población, a las variables de interés y a los niveles de precisión. El primero de ellos, fue construido previamente en Microsoft Excel y se lo denominó *STRATUM*. El segundo, es especificado en el código de R y en general se lo llama *cv*. Cabe aclarar que el paquete no brinda funciones para preparar estos *dataframes*, sino que deben crearse manualmente siguiendo ejemplos.

##### 5.1.5.1.1 Base STRATUM

Esta base de datos contiene información respecto al tamaño de cada estrato, a los dominios a los cuales pertenecen los estratos y estimaciones de la media y variabilidad de las variables de interés. En forma general, se requiere para cada estrato un registro de los siguientes campos:

- etiqueta del estrato  $h$ ,  $h = 1, \dots, H$

- $N_h$ : tamaño del estrato  $h$
- para cada uno de los tipos de dominios, el dominio al cual pertenece el estrato  $h$ . Las columnas referidas a los tipos de dominios, se nombran con  $DOM1$ , ...,  $DOMD$ , lo cual se llama código del tipo de dominio.
- media poblacional de cada una de las  $J$  variables en el estrato  $h$ .

En el caso de variables cuantitativas, se calcula como:

$$m_{j,h} = \frac{\sum_{i=1}^{N_h} y_{j,hi}}{N_h}$$

donde  $y_{j,hi}$  es el valor de la  $j$ -ésima variable en la  $i$ -ésima unidad del estrato  $h$ .

En el caso de variables cualitativas, se define una variable dicotómica, y la media corresponde a la proporción  $p_{j,h}$  del valor 1 de la variable dicotómica  $y_j$ .

$$m_{j,h} = \frac{Y_{j,h}}{N_h} = p_{j,h}$$

donde  $Y_{j,h}$  es el total de la variable  $j$  en el estrato  $h$ .

- desvío estándar poblacional de cada una de las  $J$  variables de interés en el estrato  $h$ .

Para variables cuantitativas, se calcula como:

$$s_{j,h} = \sqrt{\frac{\sum_{i=1}^{N_h} (y_{j,hi} - m_{j,h})^2}{N_h - 1}}$$

En el caso de variables categóricas:

$$s_{j,h} = \sqrt{p_{j,h}(1 - p_{j,h})}$$

- estrato de inclusión forzosa: indicador de si se realizará un censo o se tomará una muestra en el estrato  $h$
- costo del trabajo de campo en el estrato  $h$ .

A continuación se presenta la estructura de la base de datos con la que se trabaja junto con los nombres de las columnas. El dataset *STRATUM* tiene una estructura particular y es muy importante que se respeten los nombres de las columnas (a excepción de las dos primeras), para que el código se ejecute exitosamente.

- STRATUM: indica el número del estrato  $h$ , con  $h = 1, \dots, H$ .
- ESTRATOS: etiquetas de los estratos, indica los estratos particulares dados como combinación de las categorías de los tipos de dominios.
- DOM1: indica la categoría de la credencial a la cual pertenece el estrato.

- DOM2: indica la categoría del nivel de uso a la cual pertenece el estrato.
- DOM3: indica la categoría de la región a la cual pertenece el estrato.
- DOM4: indica la categoría del segmento a la cual pertenece el estrato.
- DOM5: se establece como 1 para todos los estratos. Indica que se obtendrán estimaciones para el total de GF (nivel global).
- N: número de unidades de la población que pertenecen al estrato.
- M1: media poblacional de satisfacción en el estrato.

Para el cálculo de la media de satisfacción en el estrato  $h$ , con  $h = 1, \dots, H$ , se define una variable dicotómica, y la media corresponde a la proporción  $p_{1,h}$  del valor 1 de la variable dicotómica  $y_1$ :

$$m_{1,h} = \frac{Y_{1,h}}{N_h} = p_{1,h}$$

donde  $N_h$  es el tamaño del estrato  $h$  y  $Y_{1,h}$  es el total de satisfechos en el estrato  $h$ .

- S1: desviación estándar de satisfacción en el estrato. El cálculo del desvío de satisfacción en el estrato  $h$  es:

$$s_{1,h} = \sqrt{p_{1,h}(1 - p_{1,h})}$$

- M2: media de promotores en el estrato. De manera análoga a satisfacción, el cálculo de la media de promotores en el estrato  $h$  es:

$$m_{2,h} = \frac{Y_{2,h}}{N_h} = p_{2,h}$$

- S2: desviación estándar de promotores en el estrato. El cálculo del desvío de promotores en el estrato  $h$  es:

$$s_{2,h} = \sqrt{p_{2,h}(1 - p_{2,h})}$$

- M3: media de detractores en el estrato. Análogamente:

$$m_{3,h} = \frac{Y_{3,h}}{N_h} = p_{3,h}$$

- S3: desviación estándar de detractores en el estrato. Análogamente:

$$s_{3,h} = \sqrt{p_{3,h}(1 - p_{3,h})}$$

- COST: se establece como 1 para todos los estratos ya que no hay diferencias en los costos de realización de las encuestas entre los estratos, es decir, no hay diferencias en los costos del trabajo de campo.

- CENS: se establece como 0 para aquellos estratos en los que se quiera realizar un muestreo; y como 1 si se quiere realizar un censo dentro del estrato.

Respecto a la especificación de las primeras ocho columnas, no surgen dificultades. Los nombres de los estratos vienen dados por las combinaciones de los dominios de estimación, y a partir del marco muestral del listado de GF se puede determinar el tamaño de la población en cada uno de estos estratos. La columna COST, como se mencionó anteriormente, contiene el número uno en todos los estratos. Sin embargo, al querer terminar de construir esta base de datos se presentan algunas dificultades que requieren tomar ciertas decisiones.

En principio, se decide establecer *CENS* como 1 en aquellos estratos cuyo tamaño poblacional sea menor o igual a 10, entendiendo que no se pueden obtener estimaciones confiables de una muestra de una población muy chica.

Por otro lado, es necesario definir los valores de medias y desvíos estándares de los parámetros de interés. Cabe aclarar que se utiliza la información disponible de la encuesta realizada en el año 2022. Sin embargo, el tipo de dominio nivel de uso no se utilizó anteriormente y no se cuenta con información específica. Es por eso que se decide utilizar para los tres niveles de uso, las estimaciones de las variables obtenidas en estratos formados como combinación de los tipos de dominios de credencial, segmento y región. Además, para aquellos casos en los que no hay información (credencial D no incluida el año 2022 y estratos con tamaño de muestra igual a cero en la encuesta 2022) se supone una proporción de satisfechos de 0.6, de promotores de 0.5 y de detractores de 0.3. El mismo criterio es considerado en el caso en que el tamaño de la muestra en el estrato en el año 2022 es menor o igual a 10, entendiendo que no es confiable la estimación obtenida con una muestra pequeña. Estas cantidades son definidas en función de lo observado en líneas generales para el resto de los estratos, pero están basadas en suposiciones y se espera tener información más precisa en años posteriores. En particular, se observa que en general para tamaños de muestra del 2022 mayores a 10, la proporción de satisfechos es mayor a 0.60; mientras que la de detractores siempre es menor a 0.38. En el caso de promotores, se opta por la mayor heterogeneidad (dada por una proporción de 0.5) ya que hay proporciones tanto altas como bajas.

#### 5.1.5.1.2 Base *CV*

Este *dataframe* contiene información de la precisión que se desea garantizar en las estimaciones de cada uno de los parámetros de interés para cada tipo de dominio. La precisión es expresada en términos de coeficientes de variación, lo cual se define como el cociente entre el error estándar del estimador y el valor poblacional. Específicamente, se



establece para cada tipo de dominio y cada variable de interés, el valor máximo permitido del coeficiente de variación (CV) del estimador del parámetro de interés. Para su creación, debe especificarse el código del tipo de dominio y el CV máximo esperado de las  $P$  variables de interés.

Es común establecer la misma precisión para todos los dominios. Sin embargo, si no es necesario especificar un límite de error para cierta variable en un determinado dominio, debe indicarse un valor alto del CV, como por ejemplo 1.

En este análisis, son especificadas distintas propuestas de *dataframe cv*. Esto se debe a que se decidió seleccionar el nivel de precisión con el que se va a trabajar dependiendo de los tamaños muestrales y el presupuesto destinado a la encuesta. De manera general, se presentan CV del 3%, 5%, 7%, 10% y 12%. Estos valores son totalmente aceptables y brindarán resultados válidos y confiables. La idea es garantizar mayor precisión a nivel total, resignando un poco de la misma para los tipos de dominios. Las propuestas parten de un CV del 3% a nivel total y un 5% para los tipos de dominios hasta un 5% a nivel total y un 12% para los tipos de dominios. Debe tenerse en cuenta que si se requiere mayor precisión, el tamaño de muestra será mayor.

#### 5.1.5.2 Archivos de salida

Una vez especificada la función *beat.lst* junto con los argumentos requeridos, el software brinda como salida una lista denominada *output\_beth* de cuatro elementos, los cuales se listan a continuación:

- *n*: tamaño de muestra por estrato.
- *file\_strata*: es un dataframe con la misma estructura que *STRATUM* pero con una columna adicional llamada *n* que corresponde al tamaño de muestra por estrato.
- *alloc*: es un dataframe que contiene para cada uno de los estratos, el tamaño de la muestra requerido si se utiliza adjudicación óptima, proporcional o igual.
- *sensitivity*: es un dataframe que sirve para realizar el análisis de sensibilidad. En primer lugar, contiene para cada estimador de interés y cada tipo de dominio de estimación, los errores muestrales esperados junto con los obtenidos realmente al realizar la adjudicación. Además, en una columna denominada *sensitivity 10%*, proporciona cuánto es necesario aumentar o disminuir el tamaño de la muestra para lograr una disminución o incremento respectivamente del CV de cierta estimación en un cierto tipo de dominio en un 10%.

## 5.2 Estimación de parámetros

Antes de definir el diseño muestral, se plantearon como indicadores principales el nivel de satisfacción general y el nivel de recomendación. La medición de estos aspectos se realiza mediante tres parámetros de interés:

- la proporción de GF que mencionan estar satisfechos o muy satisfechos con Avalian,
- la proporción de GF promotores de la empresa,
- la proporción de GF detractores de la empresa.

A su vez, se planteó el interés de obtener estimaciones de estos parámetros no sólo para el total de GF de Avalian, sino también para distintos dominios de estimación dados por las categorías de la credencial, el segmento, la región y el nivel de uso del GF. Las estimaciones de estos parámetros se realizan siguiendo la teoría del muestreo estratificado.

### 5.2.1 Estimación de parámetros bajo muestreo estratificado

Bajo el muestreo estratificado, la población  $U$  de  $N$  unidades muestrales (GF) se divide en  $H$  subpoblaciones, denominadas estratos. Cada unidad muestral pertenece únicamente a un estrato. Se tienen  $N_h$  GF en el  $h$ -ésimo estrato, siendo estos valores conocidos, y se cumple

$$\sum_{h=1}^H N_h = N.$$

Se toman muestras aleatorias simples de GF en cada estrato, las cuales son independientes entre sí. Es decir, de las  $N_h$  unidades poblacionales en el estrato  $h$  se seleccionan aleatoriamente  $n_h$ , siendo este conjunto definido como  $G_h$ .

Como fue mencionado, los parámetros de interés son proporciones. Una proporción se define como la media de una variable que toma valores 0 y 1. Por lo tanto, para hacer inferencias sobre proporciones, se utiliza el estimador de la media y de la variancia poblacional.

En una primera instancia, se presenta la teoría referida a este tipo de parámetro, para luego aplicar reemplazos necesarios que permitan obtener el estimador para una proporción.

#### 5.2.1.1 Estimación de los parámetros de interés

Se definen las cantidades poblacionales:

$y_{hi}$ : valor de la variable para la  $i$ -ésima unidad en el estrato  $h$

$$Y_h = \sum_{i=1}^{N_h} y_{hi} : \text{total de la variable en el estrato } h$$

$$Y = \sum_{h=1}^H Y_h : \text{total poblacional}$$

$$\bar{y}_{hU} = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h} : \text{media poblacional en el estrato } h$$

$$\bar{y}_U = \frac{Y}{N} = \frac{\sum_{h=1}^H \sum_{i=1}^{N_h} y_{hi}}{N} : \text{media poblacional}$$

$$S_h^2 = \sum_{i=1}^{N_h} \frac{(y_{hi} - \bar{y}_h)^2}{N_h - 1} : \text{variancia poblacional de la variable en el estrato } h.$$

Dentro de cada estrato, las estimaciones bajo muestreo simple al azar para la media, el total y la variancia son:

$$\begin{aligned} \bar{y}_h &= \frac{\sum_{i \in G_h} y_{hi}}{n_h} \\ \hat{Y}_h &= \frac{N_h}{n_h} \sum_{i \in G_h} y_{hi} = N_h \bar{y}_h \\ s_h^2 &= \sum_{i \in G_h} \frac{(y_{hi} - \bar{y}_h)^2}{n_h - 1}. \end{aligned} \quad (12)$$

Entonces, se estima  $\bar{y}_{hU}$  con  $\bar{y}_h$  y  $Y_h$  con  $\hat{Y}_h$ . Resulta que el estimador de  $Y$  es:

$$\hat{Y}_{est} = \sum_{h=1}^H \hat{Y}_h = \sum_{h=1}^H N_h \bar{y}_h. \quad (13)$$

Mientras que el estimador de  $\bar{y}_U$  es:

$$\bar{y}_{est} = \frac{\hat{Y}_{est}}{N} = \sum_{h=1}^H \frac{N_h}{N} \bar{y}_h \quad (14)$$

que es un promedio ponderado de las medias muestrales por estrato. Los pesos corresponden a los tamaños relativos del estrato.

Estos estimadores comparten las propiedades de los estimadores bajo muestreo simple al azar:

- $\bar{y}_{est}$  y  $\hat{Y}_{est}$  son estimadores insesgados de  $\bar{y}_U$  y  $Y$  respectivamente.

- Debido a que las muestras tomadas de los estratos son independientes y se conoce

$V(\hat{Y}_h)$  por la teoría del muestreo simple al azar, se tiene:

$$V(\hat{Y}_{est}) = \sum_{h=1}^H V(\hat{Y}_h) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) N_h^2 \frac{s_h^2}{n_h}. \quad (15)$$

- Se puede obtener un estimador insesgado de  $V(\hat{Y}_{est})$  sustituyendo la cantidad poblacional  $S_h^2$  por la estimación muestral  $s_h^2$ , definido en (12):

$$\circ \quad \hat{V}(\hat{Y}_{est}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) N_h^2 \frac{s_h^2}{n_h}, \quad (16)$$

$$\circ \quad \hat{V}(\bar{y}_{est}) = \frac{1}{N^2} \hat{V}(\hat{Y}_{est}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) \left(\frac{N_h}{N}\right)^2 \frac{s_h^2}{n_h}. \quad (17)$$

El error estándar estimado de un estimador es la raíz cuadrada de la variancia estimada:

$$\widehat{SE}(\bar{y}_{est}) = \sqrt{\hat{V}(\bar{y}_{est})}.$$

- Si los tamaños muestrales dentro de cada estrato son grandes o el diseño muestral tiene un número grande de estratos, puede calcularse un intervalo de confianza aproximado  $100(1 - \alpha)\%$  (IC) para la media:

$$\bar{y}_{est} \pm z_{\alpha/2} \widehat{SE}(\bar{y}_{est}).$$

Para estimar una proporción, se define  $y_{hi} = 1$  si la unidad  $i$  presenta la característica de interés y  $y_{hi} = 0$  en otro caso, y se utilizan ecuaciones (13) a (17) con  $\bar{y}_h = \hat{p}_h$  y  $s_h^2 = [n_h/(n_h - 1)]\hat{p}_h(1 - \hat{p}_h)$ . El estimador, por lo tanto, resulta:

$$\hat{p}_{est} = \sum_{h=1}^H \frac{N_h}{N} \hat{p}_h$$

La estimación de la variancia del estimador es:

$$\hat{V}(\hat{p}_{est}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) \left(\frac{N_h}{N}\right)^2 \frac{\hat{p}_h(1 - \hat{p}_h)}{n_h - 1}$$

Sencillamente se puede también estimar el número total de unidades con características específicas dadas por el valor 1 de la variable, el cual es la suma de los totales estimados por estrato:

$$\hat{Y}_{est} = \sum_{h=1}^H N_h \hat{p}_h$$

Y la variancia estimada del estimador:

$$\widehat{V}(\widehat{Y}_{est}) = N^2 \widehat{V}(\widehat{p}_{est}).$$

Para un dominio  $k_d$ , se puede definir  $y_{hik_d} = y_{hi}$  si la unidad  $i$  del estrato  $h$  pertenece al dominio  $k_d$  y 0 en otro caso y aplicar las fórmulas presentadas anteriormente sobre esta nueva variable. Luego, para la estimación de un parámetro en un dominio particular, esta definición de una nueva variable equivale a sumar los estratos que conforman el mismo.

## 6. Resultados

En esta sección se presentan tamaños de muestra obtenidos, junto con su adjudicación y la manera de obtener el listado de GF seleccionados. Además, se presenta una estrategia para tratar la no respuesta.

### 6.1 Obtención del tamaño de la muestra

Como es mencionado en la Sección “4.1 Materiales”, se trabaja con cuatro marcos muestrales con el fin de evaluar la inclusión de la credencial D y la categorización del segmento en dos o cinco categorías. Por lo tanto, se obtienen tamaños muestrales y su adjudicación a través de los estratos para cuatro posibles escenarios. Estos escenarios varían entonces en cuanto a las categorías de los tipos de dominios credencial y segmento, manteniendo las mismas seis regiones y tres categorías de nivel de uso, las cuales se presentan a continuación:

- **Escenario 1:** Credencial: A, B, C, D  
Segmento: A, B, C, D, E
- **Escenario 2:** Credencial: A, B, C  
Segmento: A, B, C, D, E
- **Escenario 3:** Credencial: A, B, C, D  
Segmento: F, G
- **Escenario 4:** Credencial: A, B, C  
Segmento: F, G

El planteo de cuatro escenarios se debe a que, en una primera instancia, cuando fue considerado únicamente el escenario 1, el tamaño muestral era muy grande, incluso disminuyendo la precisión. Por lo tanto, se buscan alternativas que puedan reducir dicho tamaño. Una de ellas es disminuir la cantidad de dominios de estimación, lo cual resulta en una posible reducción del tamaño de la muestra.

La categorización de segmento en F y G está basada en las encuestas realizadas en los años anteriores. Estas categorías surgen de agrupar los segmentos originales A y C en F, y B, D y E en G.

En cuanto a la credencial D, se considera una posible exclusión debido a un pedido hecho por parte de las gerencias de realizar un análisis particular de esta credencial. Por lo tanto, para poder estudiar con mayor detalle las características particulares de este plan, se plantea la posibilidad de realizar una encuesta aparte para ese conjunto de asociados.

Para cada uno de los escenarios, se presentan catorce propuestas de tamaños muestrales que varían en los coeficientes de variación planteados.

**Tabla 6:** Tamaño de la muestra para las catorce propuestas de los cuatro escenarios

Propuesta	Coeficiente de variación					Tamaños de muestra			
	Total	Tipo de dominio				Cinco segmentos		Dos segmentos	
		Región	Credencial	Segmento	Nivel uso	Con D	Sin D	Con D	Sin D
1	3%	5%	5%	5%	5%	7.218	6.682	6.449	5.937
2		5%	10%	10%	10%	5.647	5.500	5.582	5.453
3		7%	7%	7%	7%	4.122	3.760	3.537	3.182
4		7%	10%	10%	10%	3.382	3.175	3.128	2.936
5		7%	12%	12%	12%	3.237	3.091	3.052	2.911
6		10%	12%	12%	12%	2.990	2.860	2.714	2.602
7	5%	5%	10%	10%	10%	5.647	5.500	5.582	5.453
8		7%	7%	7%	7%	4.123	3.760	3.526	3.177
9		7%	10%	10%	10%	3.382	3.175	3.116	2.929
10		7%	12%	12%	12%	3.237	3.091	3.038	2.904
11		10%	10%	10%	10%	2.312	2.078	1.844	1.622
12		10%	10%	10%	12%	2.292	2.054	1.820	1.597
13		10%	12%	12%	12%	2.062	1.875	1.691	1.527
14		12%	12%	12%	12%	1.772	1.582	1.347	1.173

Los tamaños de muestra obtenidos van desde 1.173 hasta 7.218. Puede notarse, en la Tabla 6, cómo disminuye el tamaño de muestra al aumentar los coeficientes de variación máximo permitidos en los tipos de dominios.

El menor tamaño de muestra se alcanza en el escenario 4 (dos segmentos sin D) con un CV del 5% a nivel global, es decir, para la población de GF de Avalian, y un máximo del 12% para los dominios región, credencial, nivel de uso y segmento. Para garantizar estos niveles de precisión es necesario que el tamaño de muestra efectivo sea de 1.173 GF. Debido a que se trata de la primera experiencia en este tipo de metodología y por cuestiones de recursos, se elige el tamaño menor de muestra para la encuesta actual.

## 6.2 Tratamiento de la no respuesta

Además del error de estimación dado por la variación aleatoria de muestreo que está presente cuando se mide solamente un subconjunto de la población (muestra), puede presentarse una fuente adicional de error por la presencia de no respuesta.

La no respuesta ocurre cuando no se logra localizar a las unidades muestrales o cuando estas se niegan a responder. Debido a que las unidades que no responden pueden tener respuestas distintas a los que sí lo hacen, la consecuencia principal es la posible existencia de sesgo en las estimaciones.

*R2BEAT* asegura los niveles de precisión planteados siempre que el tamaño muestral efectivo sea el especificado por el paquete. Desafortunadamente, el mismo es afectado por la no

respuesta, reduciéndolo y, por lo tanto, debilitando la precisión y eficiencia de las estimaciones. Por lo tanto, es importante considerar la no respuesta al planear el diseño muestral e implementar estrategias que busquen preservar la precisión de las estimaciones.

Uno de los métodos más comunes es el sobremuestreo. Esta técnica se usa para incrementar el tamaño muestral en cada estrato de acuerdo con la tasa de no respuesta presente en dicho estrato, de manera de obtener al final de la etapa de recopilación de datos, el tamaño de muestra deseado que asegura los requerimientos de precisión esperados para los supuestos considerados. Estas tasas usualmente se derivan de encuestas anteriores, llevadas a cabo con la misma técnica de recopilación de datos y en condiciones similares, o de suposiciones confiables de quienes diseñan la encuesta. Cuanto más preciso sea el conocimiento de la tasa de respuesta y de las características de quienes no responden, más preciso será el procedimiento de sobremuestreo.

De esta manera, se propone aumentar el tamaño de muestra considerando el porcentaje de tasa de respuesta, asumiendo que la no respuesta no está asociada a la característica de la encuesta. Sin embargo, este porcentaje es desconocido, siendo necesario realizar una suposición del mismo. En este caso, se propone suponer una tasa de respuesta del 50% en cada uno de los estratos. Siendo el tamaño muestral de 1.173 GF, en este caso, la muestra inflada es de 2.346 GF.

En la presente encuesta, se recomienda realizar un seguimiento de la no respuesta que permita contar con información más precisa de la tasa de respuesta en los años siguientes, así como también acerca de las características de quienes no responden.

Cabe destacar que este método tiene como objetivo preservar el tamaño de muestra inflándolo para contemplar la no respuesta. Sin embargo, puede no ser suficiente para evitar completamente los sesgos presentes en las estimaciones finales causados por la falta de respuesta, siendo necesario utilizarse otros métodos de tratamiento de no respuesta luego de la etapa de recopilación de datos, como, por ejemplo, a través del ajuste de pesos por post-estratificación o calibración y la estimación de variancias contemplando dichos ajustes.

## 6.3 Selección de grupos familiares

Para obtener el listado de GF seleccionados para responder la encuesta, se usa la función *strata* del paquete *sampling*. Es necesario brindar el marco muestral de GF que, en este caso particular, debe contener como información esencial las siguientes columnas:

- Número del grupo familiar.



- ESTRATOS: etiquetas de los estratos. Esta columna coincide con la de la base de datos STRATUM.
- DOM1: indica la categoría de la credencial a la cual pertenece el estrato.
- DOM2: indica la categoría del nivel de uso a la cual pertenece el estrato.
- DOM3: indica la categoría de la región a la cual pertenece el estrato.
- DOM4: indica la categoría del segmento a la cual pertenece el estrato.
- DOM5: se establece como 1 para todos los estratos.

Puede agregarse información adicional, como por ejemplo los datos de contacto.

## 7. Conclusiones

Avalian realiza cada año la Encuesta de Satisfacción de Asociados con el objetivo de conocer el nivel de satisfacción general de sus asociados con el desempeño de la empresa. Dentro del cuestionario, se incluyen preguntas vinculadas a la satisfacción general del asociado, al nivel de recomendación a familiares/amigos, a la recompra y a la existencia de inconvenientes, sus causas y resolución.

Resulta fundamental que los resultados obtenidos en la encuesta sean confiables, ya que en base a ellos se acciona buscando mejorar el servicio y aumentar la satisfacción del asociado. Sin embargo, en las últimas encuestas realizadas, esta confianza se ha visto afectada tanto por el diseño muestral elegido, muestreos por cuotas, así como por los sesgos causados por el método de recolección de datos. Es por ello, que en el año 2023 desde Avalian surge el interés de llevar a cabo el proyecto de la Encuesta de Satisfacción con la colaboración de la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario a través de un convenio, para presentar una estrategia que minimice los sesgos y maximice la confianza de los resultados.

El objetivo principal de esta tesina, por lo tanto, es investigar y aplicar un diseño muestral para la Encuesta de Satisfacción de Asociados de Avalian que brinde resultados confiables a la hora de tomar decisiones. A partir del análisis de antecedentes, de la determinación de los objetivos de la encuesta actual y de la definición de la población objetivo, de las variables de interés, de los dominios de estimación, entre otras cuestiones, se tiene que el diseño debe tener en cuenta distintos niveles de estratificación y distintos niveles de dominios de estimación.

En este sentido, se utilizó un diseño muestral probabilístico, particularmente un muestreo estratificado de GF con adjudicación de la muestra siguiendo un enfoque multivariado y multidominio. En comparación con las encuestas realizadas en los últimos años, se podrá conocer el nivel de satisfacción y de recomendación para todos los GF de la población objetivo y para los dominios considerados, y una medida de error que indique el nivel de precisión de las estimaciones. Esta metodología resulta innovadora para Avalian, que utilizaba diseños muestrales no probabilísticos en sus encuestas anteriores.

A su vez, se trabaja sobre los posibles sesgos que generan diferencias en los métodos de recolección de datos, buscando la respuesta del GF cualquiera sea el medio de contacto. Además, se trabaja en el instrumento de recolección de datos, creando un formulario simple y ágil, con un máximo de catorce preguntas, simplificando la formulación de las mismas y las

escalas utilizadas. Se establece también una estrategia de comunicación que contribuya a aumentar las tasas de respuesta.

Frente al requerimiento de obtener estimaciones de tres parámetros de interés en cuatro dominios de estimación, se presenta un problema de adjudicación de la muestra multivariado y multidominio. Además, se busca maximizar la calidad de los resultados dentro de las restricciones presupuestarias, por lo cual la adjudicación de la muestra debe ser óptima. Por lo tanto, la metodología que se utiliza para determinar el tamaño de la muestra y su adjudicación a través de los estratos debe tener en cuenta estos dos aspectos. Se utiliza entonces una generalización del algoritmo de Bethel, utilizado en diseños muestrales estratificados cuando se tienen múltiples variables y múltiples dominios.

Fasulo et al. (2021) abordó este problema de adjudicación multivariada y multidominio, creando un paquete en el software R llamado *R2BEAT*, que brinda una generalización del algoritmo de Bethel para determinar el tamaño de la muestra y su adjudicación a través de ciertos estratos en diseños muestrales estratificados multivariados y multidominios en una y dos etapas.

Utilizando este paquete, se obtuvieron, para cada uno de los escenarios planteados, distintos tamaños de muestra según distintos niveles de precisión. Los mismos fueron presentados a los responsables de la encuesta en la empresa, y se eligió el menor tamaño de 1.173 GF. El mismo corresponde al escenario 4 (sin la credencial D y segmento presentada en dos categorías) con niveles de precisión del 5% a nivel global y del 12% para las desagregaciones dadas por región, segmento, nivel de uso y credencial. Para garantizar esta confianza en los resultados, efectivamente deben responder la encuesta la totalidad de los 1.173 GF seleccionados.

Teniendo en cuenta que, en general, la no respuesta es inevitable, se optó por inflar la muestra suponiendo una tasa de respuesta del 50%. En este caso, la muestra inflada corresponde a 2.346 GF, y se presenta el listado de los GF que componen la muestra.

Se espera que a futuro se obtenga un mayor conocimiento de la tasa de respuesta de los asociados de Avalian, así como también de los valores supuestos, necesarios para la obtención del tamaño de la muestra, de la media y de la variabilidad de los parámetros de interés. De esta manera, se podrá lograr una mejora en el diseño muestral.

De igual manera, se esperan avances en el proyecto de Calidad de Datos de manera que la información faltante de los datos de contacto de los GF esté completa y actualizada, aspecto fundamental del marco muestral en un diseño probabilístico.

## 8. Bibliografia

- Barcaroli, G., Buglielli, T., Vitiis, C. D. (2020). *MAUSS-R: Multivariate Allocation of Units in Sampling Surveys (R package Version 2.4)*. URL <https://www.istat.it/en/methods-and-tools/methods-and-it-tools/design/design-tools/mauss-r>.
- Bethel, J. W. (1985). *An optimum allocation algorithm for multivariate surveys*. Proceedings of the Survey Research Section, American Statistical Association, pp. 209-212.
- Bethel, J. (1989). *Sample Allocation in Multivariate Surveys*. Survey Methodology 15 (1): 47–57.
- Buglielli, T., De Vitiis, C., Barcaroli, G. (2013). *MAUSS-R: Multivariate Allocation of Units in Sampling Surveys*. [https://www.istat.it/it/files/2011/02/user\\_and\\_methodological\\_manual.pdf](https://www.istat.it/it/files/2011/02/user_and_methodological_manual.pdf)
- Chatterjee, S. (1968). *Multivariate stratified surveys*. Journal of the American Statistical Association, 63, 530-534.
- Chatterjee, S. (1972). *A study of optimum allocation in multivariate stratified surveys*. Skandinavisk Actuarietidskrift, 55, 73-80.
- Chromy J. (1987). *Design Optimization with Multiple Objectives*, Proceedings of the Survey Research Methods Section, American Statistical Association, pp.194-199.
- Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.
- Dalenius, T. (1953). *The multivariate sampling problem*. Skandinavisk Actuarietidskrift, 36, 92-102.
- Dalenius, T. (1957). *Sampling in Sweden*. Stockholm: Almquist and Wicksall.
- Fasulo, A., Barcaroli, G., Cianchetta, R., Falorsi, S., Guandalini, A., Pagliuca, D., Terribili, M. D. (2021). *R2BEAT: Multistage Sampling Design and Sample Selection (Version 1.0.2)*. <https://barcaroli.github.io/R2BEAT/>
- Fasulo, A., Barcaroli, G., Guandalini, A., Pagliuca, D., Terribili, M. D. (2023). *Two-Stage Sampling Design and Sample Selection with the R Package R2BEAT*. The R Journal 15(3):191-213.
- Folks, J.L., and Antel, C.E. (1965). *Optimum allocation of sampling units when there are R responses of interest*. Journal of the American Statistical Association, 60, 225-233.

- Forsyth, G.E. (1968). *On the asymptotic directions of the s-dimensional optimum gradient method*. Numerische Mathematik, 11, 57-76.
- Hartley, H.O. (1965). *Multiple purpose optimum allocation in stratified sampling*. Proceedings of the Social Statistics Section. American Statistical Association, 258-261.
- Huddleston, H.F., Claypool, P.L., and Hocking, R.R. (1970). *Optimum sample allocation to strata using convex programming*. Applied Statistics, 19, 273-278.
- Kish, L. (1976). *Optima and proxima in linear sample designs*. Journal of the Royal Statistical Society A., 139, 80-95.
- Kokan, A.R. (1963). *Optimum allocation in multivariate surveys*. Journal of the Royal Statistical Society A., 126, 557-565.
- Kokan A.R., Khan S. (1967), *Optimum allocation in multivariate surveys: an analytical solution*, Journal of the Royal Statistical Society B., No. 29, pp. 115-125
- Kuhn H.W., Tucker A.W. (1951), *Nonlinear Programming*, Proceedings of II Berkley Symposium Mathematical Statistics and Probability.
- Lohr, S. L. (2009). *Sampling: design and analysis*. Pacific Grove, CA: Duxbury Press.
- Neyman, J. (1934). *On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection*. Journal of the Royal Statistical Society 97 (4): 558–625.
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Tschuprow, A. A. (1923). *On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations*. Metron 2: 646–83.
- Yates, F. (1960). *Sampling methods for Censuses and Surveys*. London: Charles Griffin and Company.

## 9. Anexo

### 9.1 Anexo 1: Definiciones de Avalian

**Asociados migrados a nuevos planes y asociados genuinos/nativos:** Con la finalidad de reducir la cantidad de planes comercializables, durante el año 2017 se creó un número reducido de nuevos planes, los cuales serían los únicos que se ofrecerían en el proceso de venta. Los planes existentes hasta ese momento quedarían fuera de dicho proceso. La separación entre socios migrados y socios nativos surge debido a que se les brindó a los asociados activos la posibilidad de migrar a los nuevos planes o permanecer en el plan contratado.

**Credencial:** Elemento que acredita que un individuo es asociado a la entidad y reúne las condiciones adecuadas para utilizar los servicios. Cada una de las credenciales agrupa diferentes tipos de planes presentes en Avalian. Existen credenciales de planes que comercializa la empresa: A, B, C y D, siendo A y C los de mayor cobertura, y credenciales de planes no comercializables.

**Región:** Hace referencia a una división del territorio nacional en zonas menores.

**Segmento:** Clasificación del asociado titular de acuerdo a la modalidad de contratación, teniendo en cuenta la derivación de aportes y el sujeto de facturación.

Según el sujeto de facturación:

- **G:** asociados cuyo ingreso se produce por un proceso personal de decisión y que son responsables del pago del plan contratado.
- **F:** asociados cuyo ingreso no se produce por un proceso personal de decisión y que no son responsables del pago del plan contratado.

Siendo las categorías A, B, C y D sub divisiones de las categorías anteriores, basadas en la derivación o no de aportes. La categoría E incluye convenios particulares.

**Titular individual:** Toda persona con capacidad jurídica de contratar (desde los 18 años o emancipado), con residencia en la República Argentina y que acredite fehacientemente el cumplimiento del régimen tributario que le corresponda. Según la política comercial vigente podrá ser titular toda aquella persona a partir de los 12 años como responsable de la contratación: padre, madre y/o tutor.

**GF:** Grupo familiar. Integrado por:

- El titular, que deberá ser una persona con capacidad jurídica de contratar.

- Su cónyuge o persona que conviva con el socio titular en unión de hecho.
- Hijos solteros hasta los veinticinco (25) años inclusive, que estén a exclusivo cargo del socio titular y que acrediten que se encuentren cursando estudios regulares oficialmente reconocidos por la autoridad pertinente.
- Hijos incapacitados con certificado de discapacidad vigente y a cargo del socio titular, mayores de veintiún (21) años.
- Hermanos incapacitados del socio titular, mayores de dieciocho (18) años, cuya curatela haya sido acordada por autoridad judicial

**RUP:** Registro Único de Postulantes. Es el procedimiento para la selección de aspirantes a ingresar como personal de la Municipalidad de Rosario, creado a partir del compromiso del estado municipal como respuesta a las necesidades de innovación en políticas de gestión del personal.

## 9.2 Anexo 2: Cuestionario

A continuación se presenta una propuesta de cuestionario de la Encuesta de Satisfacción 2023 desarrollado a la fecha, el cual será enviado mediante Formularios de Google:

Te invitamos a responder estas preguntas para que evalúes tu experiencia y la de tu Grupo familiar en Avalian. Recordá que la Encuesta es confidencial y no te llevará más de 5 minutos.

### **Satisfacción:**

Considerando la experiencia del grupo familiar ¿Qué tan satisfecho te encontrás con Avalian? [RU<sup>11</sup>](#)

- Muy satisfecho
- Satisfecho
- Ni satisfecho/ Ni insatisfecho
- Insatisfecho
- Muy insatisfecho

### **Satisfacción motivos:**

¿Qué aspectos consideras más importantes a la hora de evaluar a una empresa de medicina prepaga?

Elegí hasta tres opciones.

- Calidad de la atención médica recibida
- Red de prestadores / cartilla médica
- Cobertura (alcance de servicios contratados)
- Agilidad y facilidad de uso de los servicios brindados
- Relación precio - calidad
- Cordialidad y eficiencia del personal de Avalian

<sup>11</sup> Respuesta única: sólo puede seleccionar una opción.

Otro: \_\_\_\_\_

¿Qué tan satisfecho te encontrás con los siguientes aspectos? [ROTAR<sup>12</sup>](#)

Aspectos	Muy satisfecho	Satisfecho	Ni satisfecho/ni insatisfecho	Insatisfecho	Muy insatisfecho	No sabe/no usé
Calidad de la atención médica recibida						
Red de prestadores / cartilla médica						
Cobertura (alcance de servicios contratados)						
Agilidad y facilidad de uso de los servicios brindados						
Relación precio - calidad						
Cordialidad y eficiencia del personal en centros de atención de Avalian						

**Recomendación:**

¿En qué medida recomendarías Avalian a algún familiar o amigo en una escala donde 1 es "Seguramente NO lo recomendaría" y 10 es "Seguramente lo recomendaría"? [RU](#)

1	2	3	4	5	6	7	8	9	10
Seguramente NO lo recomendaría					Seguramente lo recomendaría				

**Recompra:**

¿Si tuvieses que elegir nuevamente tu cobertura médica, optarías por Avalian? [RU](#)

Si

No

**Inconvenientes:**

¿Tuviste inconvenientes con alguno de nuestros servicios? [RU](#)

Si -> pasar a pregunta xx

No -> pasar a pregunta xx

<sup>12</sup> En cada formulario se rota el orden de las categorías.



¿Qué problemas tuviste ? [RM<sup>13</sup> - ROTAR](#)

- Tuve problemas con la cartilla
- El prestador me cobró un importe extra (plus)
- Falta de turnos en la atención médica
- Problemas en la facturación (cobranzas-aportes)
- Incumplimiento con una promoción o descuento
- Atención poco cordial por parte del personal de Avalian
- Inconvenientes para resolver un reclamo
- El alcance del servicio contratado (cobertura) no fue el adecuado
- Los tiempos de respuesta por parte de Avalian no fueron adecuados
- Existieron problema con la aplicación/página web
- Otros. Especificar (FIXED)

En general ¿resolvimos tus inconvenientes? [RU](#)

- Si
- No

**Uso - [RM](#)**

¿En cuáles de estas situaciones recurriste a nuestros servicios en el último año? Podés elegir más de una opción

- No usamos los servicios
- Consulta
- Avalian E-doc
- Guardia y servicio de urgencia
- Internación, cirugía.
- Tratamiento prolongado, medicación frecuente (anticoncepción, hipertensión, entre otras)
- Tratamiento y/o uso de medicación de alta complejidad

**General:**

Si considerás agregar algo más, **dejanos tu comentario**

### 9.3 Anexo 3: Seguimiento encuesta

Los medios de contacto elegidos para responder la encuesta son: email, SMS, WhatsApp, llamados telefónicos e intervención de las agencias. Junto con el área de Comunicación al Asociado dentro de Avalian se definió el orden en que se utilizarán estos medios de contacto. Se plantean distintos caminos de seguimiento de la encuesta según si el titular tiene cónyuge, mail y teléfono.

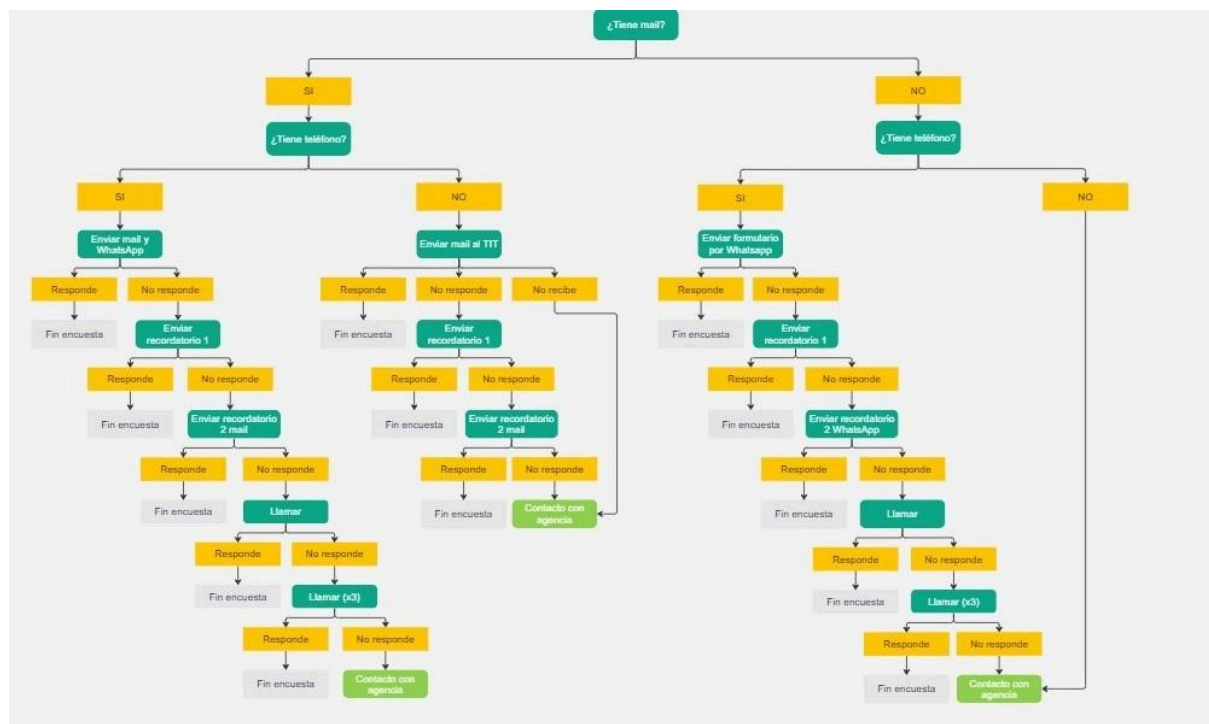
En líneas generales, se prioriza la obtención de la respuesta por medio del envío de mails y WhatsApps al titular del GF. En caso de no recibir respuesta, se envían hasta dos recordatorios mediante los mismos medios. Luego, de igual manera mediante mails y WhatsApps, se continúa con los cónyuges de los mismos. Si aún persiste la falta de respuesta,

---

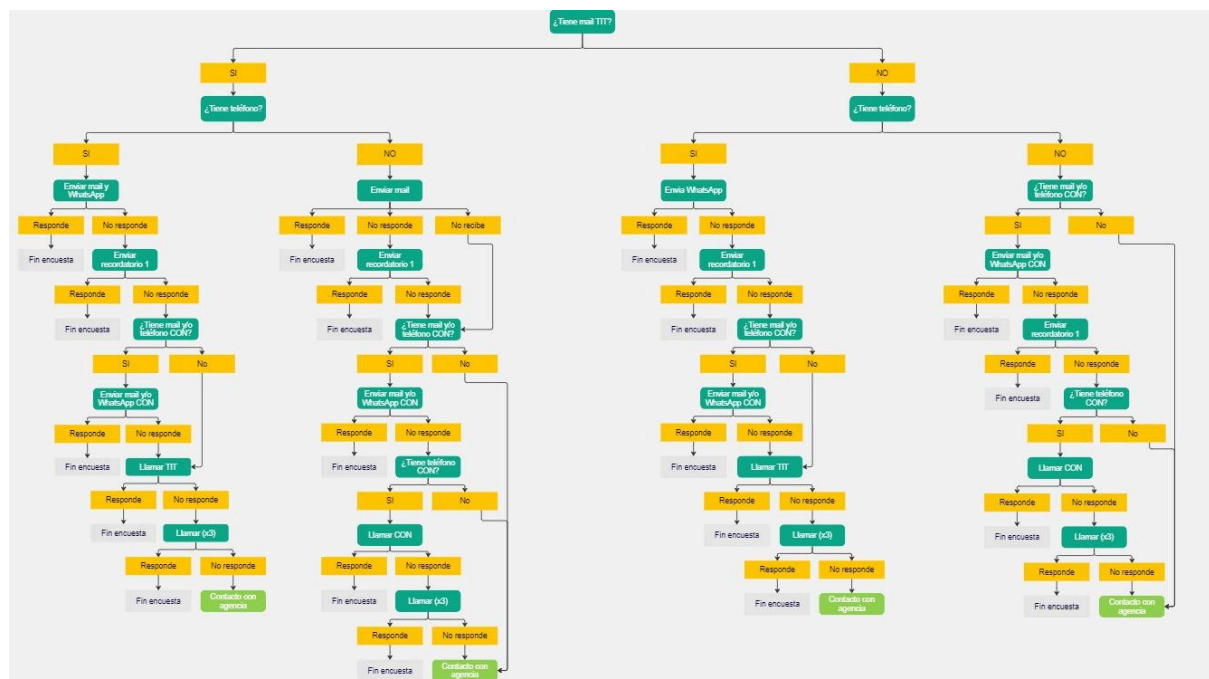
<sup>13</sup> Respuesta múltiple: puede seleccionar más de una opción.

se realizan llamados telefónicos al titular o al cónyuge en caso de que sea el único contacto telefónico disponible. En última instancia intervienen las agencias.

Para aquellos titulares que no tengan cónyuge, el seguimiento de la encuesta es el siguiente:



En cambio, para los que sí tienen es el siguiente:



## 9.4 Anexo 4: Código de R utilizado

El enlace al repositorio de GitHub con el código de R es:

<https://github.com/sofisuares/Tesina-Encuesta-de-Satisfaccion>