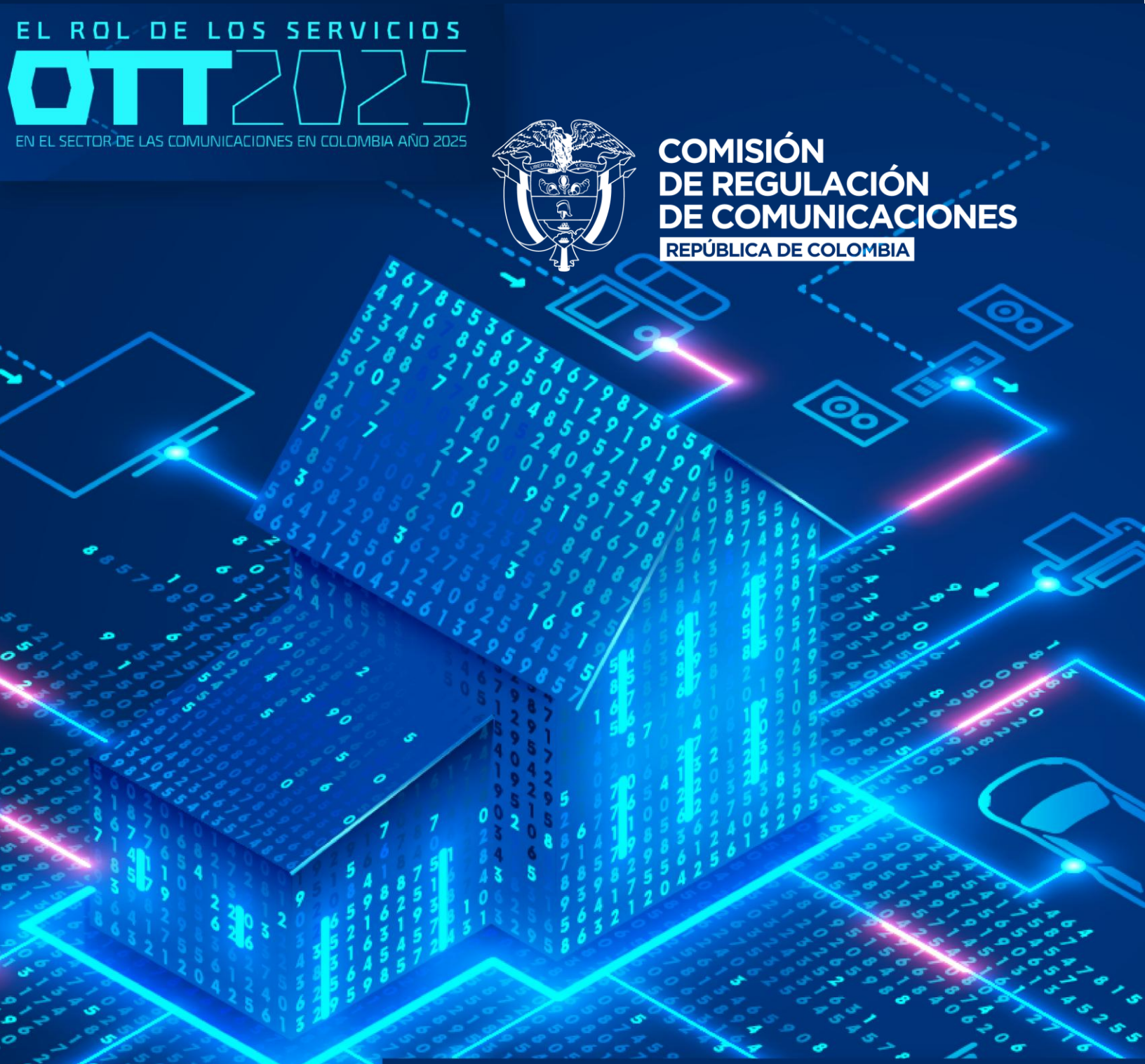
EL ROL DE LOS SERVICIOS

OTT 2025

EN EL SECTOR DE LAS COMUNICACIONES EN COLOMBIA AÑO 2025

COMISIÓN DE REGULACIÓN DE COMUNICACIONES

REPÚBLICA DE COLOMBIA



AÑO 2025

@CRCCol

/CRCCol

/CRCCol

CRCCOL

Informe metodológico
Análisis Hipótesis
Segmento RESIDENCIAL

CNC

Contenido

1. Introducción

La Comisión de Regulación de Comunicaciones (CRC) es una Unidad Administrativa Especial, de orden nacional, con independencia administrativa, técnica, patrimonial, presupuestal y con personería jurídica, la cual tiene por objeto promover la competencia en los mercados, promover el pluralismo informativo, evitar el abuso de posición dominante, regular los mercados de las redes y los servicios de comunicaciones y garantizar la protección de los derechos de los usuarios; con el fin que la prestación de los servicios sea económicamente eficiente, y refleje altos niveles de calidad, de las redes y los servicios de comunicaciones, incluidos los servicios de televisión abierta radiodifundida y de radiodifusión sonora.

Para cumplir con esta misión, es esencial que la CRC mantenga un conocimiento profundo y actualizado sobre la adopción y evolución de los servicios de comunicaciones, tanto tradicionales como en línea (Over The Top – OTT), para lo cual se hace necesario realizar estudios y análisis del sector de las comunicaciones para orientar sus decisiones regulatorias.

De acuerdo con esto, la CRC contrató al Centro Nacional de Consultoría (CNC) como consultor experto en el diseño, levantamiento, depuración, procesamiento y análisis de datos estadísticos; así como en la aplicación de encuestas que permitan recopilar información que sirva de soporte a cualquier proyecto transversal que adelante la CRC, específicamente para el rol de los servicios OTT en el sector de comunicaciones en Colombia en 2025.

2. Ficha técnica

Tabla 1. Ficha técnica

Entidad contratante	Comisión de Regulación de Comunicaciones (CRC)
Entidad financiadora de la investigación	Comisión de Regulación de Comunicaciones (CRC)
Proveedor de investigación	Centro Nacional de Consultoría S.A.
Naturaleza y temática del estudio	Estudio cuantitativo que busca caracterizar la adopción de servicios en línea de comunicaciones (over the top – OTT) e identificar su relación con los servicios de comunicaciones tradicionales en el segmento residencial en Colombia en 2025.
Población objetivo y perfil del entrevistado	Personas de 15 o más años, residentes en hogares no institucionalizados del territorio colombiano.
Cobertura	Nacional
Métodos de muestreo	Diseño de muestreo probabilístico, estratificado multietápico de elementos. Los estratos estadísticos se conforman con el cruce de las regiones geográficas y los clústeres Las Unidades Primarias del Muestreo (UPM) son los municipios, para la selección se incluye de manera forzosa a las cinco ciudades principales del país (Bogotá, Medellín, Cali, Barranquilla y Cartagena) y las demás se seleccionan usando un diseño de muestreo aleatorio simple. Las Unidades Secundarias del Muestreo (USM) son segmentos cartográficos dentro de las UPM seleccionadas, estos se conforman por manzanas cartográficas en el área urbana y por veredas en el área rural. Las Unidades Terciarias del Muestreo (UTM) son los hogares, que se seleccionan de manera sistemática, y la unidad final del muestreo son las personas, seleccionando aleatoriamente a una persona por hogar.
Descripción del marco muestral	El marco muestral se construye a partir del Marco Geoestadístico Nacional (MGN) construido por el DANE en el CNPV 2018 y actualizado al año 2023, este consta de las manzanas

	cartográficas en las cabeceras municipales y veredas en las zonas rurales en todo el territorio nacional.
Mecanismos de ponderación	En el proceso de expansión se ajustó un modelo de propensión inversa (IPW), integrando los datos de la encuesta de la CRC con la Encuesta Nacional de Calidad de Vida 2024 (ENCV2024) realizada por el DANE. El modelo, definió a priori, incorpora variables demográficas, regionales y variables comunes entre la ENCV2024 y la encuesta CRC. La estimación de los parámetros del modelo se realizó mediante el método de ecuaciones de estimación generalizadas (GEE). Posteriormente, se aplica un estimador de calibración con el método de raking, utilizando las proyecciones de población post-COVID del DANE para 2025, tanto a nivel de personas como de hogares Se construyen dos factores de expansión: uno para hogares y otro para personas. En el modelo de calibración de hogares se considera la interacción entre región, clúster, el área (urbana o rural) y el nivel socioeconómico según el Sistema Único de Información (SUI) de la Superintendencia de Servicios Públicos. En el modelo de calibración de personas se incluyen además el sexo y el rango de edad.
Tamaño de la muestra	3.043 encuestas en 58 municipios.
Nivel de confianza y error muestral	Las estimaciones para el agregado nacional tienen un margen de error máximo del 2% cuando se usa un nivel de confianza del 95%.
Cambios en la muestra	3.043 encuestas realizadas de 3.000 previstas.
Técnica de recolección de datos	Encuesta presencial en hogares.
Fecha de trabajo de campo	09 de agosto a 20 de septiembre de 2025.

Tipos de incentivos	No se emplearon incentivos.
Número de encuestadores	128 encuestadores y 16 supervisores.
Métodos de supervisión de entrevistadores	Monitorización: 10% (Escucha en vivo durante la encuesta: 2% y escucha de audios: 8%) Revisión Digital: 10% (Incluye re - contactos y escucha de audios para verificar información, encuestas de corta duración, entre otros)
Procedimientos de imputación	Ningún dato de la base de datos fue imputado: todos corresponden a los de las encuestas.
Errores no muestrales	Durante la revisión de las etapas del estudio no se identificó este tipo de error.
Subcontratación	El CNC se apoya en subcontratistas en etapas como la recolección y análisis de los datos, siempre verificando la calidad de la información entregada.

Nota 1: *El Centro Nacional de Consultoría (CNC) recolecta datos personales únicamente con fines estadísticos o de investigación atendiendo su Política de Tratamiento y Protección de Datos Personales y lo establecido en la Ley 1581 de 2012. Para garantizar lo anterior, la información que se entrega es anonimizada, con excepción de los casos en los cuales el entrevistado haya dado autorización.*

Nota 2: *Se recomienda tener precaución al analizar cifras con bases pequeñas.*

Este informe atiende los lineamientos de la norma ISO 20252:2019

3. Diseño muestral

3.1 Universo

Personas de 15 o más años, residentes en hogares no institucionalizados del territorio colombiano. Para estructurar el universo en estudio se definieron seis regiones geográficas y cuatro tipos de municipio. Las regiones geográficas se establecieron mediante la agrupación de departamentos de la manera que lo presenta la tabla 2.

Tabla 2. Departamentos por región

Región	Departamentos
Antioquia	Antioquia
Atlántica	Archipiélago de San Andrés, Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena, Sucre
Centro	Caldas, Huila, Quindío, Risaralda, Tolima
Cundinamarca	Bogotá, D.C., Cundinamarca
Oriente	Arauca, Boyacá, Casanare, Guainía, Guaviare, Meta, Norte de Santander, Santander, Vaupés, Vichada
Suroccidente	Amazonas, Caquetá, Cauca, Chocó, Nariño, Putumayo, Valle del Cauca

La tabla 3 presenta la población mayor de 15 años por región y clúster según la clasificación en el Anexo 4.9 de la Resolución CRC 5050 de 2016, adicionado mediante el artículo 6 de la Resolución CRC 7242 de 2023:

Tabla 3. Población mayor de 15 años en Colombia por región y clúster

Región	Alto y Moderado	Incipiente	Bajo	Limitado	Total
Antioquia	3.935.203	798.356	640.857	227.807	5.602.223
Atlántica	4.467.286	1.878.705	1.885.402	878.183	9.109.576
Centro	2.553.602	803.314	729.291	48.672	4.134.879
Cundinamarca	8.542.664	432.686	453.681	14.287	9.443.318

Región	Alto y Moderado	Incipiente	Bajo	Limitado	Total
Oriente	3.308.481	1.065.171	1.167.260	461.526	6.002.438
Suroccidente	3.944.751	1.291.536	1.162.475	954.975	7.353.737
Total	**26.751.987**	**6.269.768**	**6.038.966**	**2.585.450**	**41.646.171**

Por otra parte, la tabla 4 presenta la cantidad de municipios por región y clúster:

Tabla 4. Cantidad de municipios por región y clúster

Región	Alto y Moderado	Incipiente	Bajo	Limitado	Total
Antioquia	16	38	45	26	**125**
Atlántica	14	39	90	54	**197**
Centro	21	38	73	5	**137**
Cundinamarca	20	26	67	4	**117**
Oriente	19	45	167	97	**328**
Suroccidente	19	33	84	82	**218**
Total	**109**	**219**	**526**	**268**	**1.122**

3.2 Marco muestral

El marco de muestreo usado en el estudio corresponde al Marco Geoestadístico Nacional levantado con el Censo Nacional de Población y Vivienda 2018 y actualizado por el DANE en el año 2023, disponible en el Centro Nacional de Consultoría. Este dispositivo permite identificar y ubicar a cada una de las unidades estadísticas en el diseño de muestreo propuesto. Dicho marco cartográfico considera los 1.122 municipios como unidades geográficas; los hogares se agrupan dentro de manzanas cartográficas en las cabeceras municipales o veredas y centros poblados en las zonas rurales.

3.3 Método de muestreo

Se empleó un muestreo probabilístico estratificado y multietápico. En cada etapa se seleccionaron unidades estadísticas mediante muestreo aleatorio simple.

La estratificación estadística en la primera etapa consideró el cruce de regiones geográficas y el clúster. Siguiendo el diseño de mediciones previas, se estableció la inclusión obligatoria de Bogotá, Medellín, Cali, Barranquilla y Cartagena, ya que cada una de estas ciudades cuenta con más de un millón de habitantes. Del resto del país se seleccionaron un conjunto de 53 municipios, cada uno con probabilidad de inclusión menor a uno.

En la segunda etapa, dentro de cada municipio seleccionado, se eligió un conjunto de manzanas cartográficas y otro de veredas. En la tercera etapa, se tomó una muestra aleatoria simple de hogares dentro de cada manzana o vereda, según correspondía. En la última etapa, se seleccionó a una persona dentro del hogar, de acuerdo con la fecha de cumpleaños más reciente al momento de la visita del encuestador. En promedio, en cada manzana cartográfica se aplicó el cuestionario a 4 personas (de diferentes hogares) y en cada vereda a 6 personas.

Para Bogotá, Medellín, Cali, Barranquilla y Cartagena, la selección de manzanas cartográficas tomó en cuenta la estructura por nivel socioeconómico de las viviendas, con base en la información proporcionada por la Superintendencia de Servicios Públicos Domiciliarios (Superservicios). Aunque las variables sociodemográficas no formaron parte de la estratificación del diseño muestral, durante el trabajo de campo se monitoreó la distribución de los casos por sexo y edad, con el fin de asegurar que esta fuera coherente con la composición poblacional.

3.4 Algoritmo de selección de unidades estadísticas

Para las etapas en las que se aplicó un muestreo aleatorio simple (MAS) se usó el algoritmo coordinado negativo. Este algoritmo se eligió porque era apropiado para este diseño y permitió poder incluir unidades de reemplazo, mitigando el efecto de la no respuesta ya que se sigue el orden de inclusión

aleatorio determinado desde la selección inicial. El algoritmo se define a continuación:

Sea N_i el número de unidades susceptibles de selección en el conglomerado i (estratos o etapas del muestreo) y se requiere seleccionar una muestra de tamaño n_i dentro del i-ésimo conglomerado. El algoritmo se aplica de forma independiente en cada etapa o estrato y consiste en los siguientes pasos:

1. Para cada elemento j del conglomerado i se generó un número aleatorio de la distribución uniforme entre 0 y 1, esto es, $\zeta_{ij} \sim U(0,1)$.

2. Dentro del conglomerado i, se ordenó la lista de elementos con respecto a los valores aleatorios ζ_{ij}.

3. Se seleccionan los primeros n_i elementos del conglomerado i.

Para la selección de las unidades de muestreo se usaron como semillas de selección las fechas 20250521, 20250523, 20250528 en las diferentes etapas del muestreo, y se emplearon funciones de los paquetes tidyverse y janitor del software R (R Core Team 2024)

3.5 Tamaño de la muestra

El cálculo del tamaño de la muestra se realizó usando la siguiente expresión (Kish 2005):

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \qquad (1)$$

Con

$$n_0 = \frac{z_{1-\alpha/2}^2 P(1-P) DEFF}{\varepsilon^2}$$

En donde N y n son el tamaño de la población y el tamaño de la muestra respectivamente; $z_{1-\alpha/2}$ es el percentil de una distribución normal, para este caso es 1,96 para obtener un 95% de confianza; P es el parámetro trazador que se fijó en 0,5 para generar un escenario de máxima variación; $DEFF$ es el efecto de diseño que se fijó en 1,2 con base en estudios previos del CNC y ε es el margen de error máximo tolerado que se fijó en 3% según los términos de referencia.

El tamaño de muestra propuesto fue de 3.000 hogares (aplicación del cuestionario a una persona del hogar) en 58 municipios de Colombia (5 de inclusión forzosa y 53 de inclusión probabilística). Este tamaño de muestra permitía reportar indicadores de resultados con máximo 2% de error de muestreo y confiabilidad de 95%.

3.6 Encuestas efectivas

En el segmento de hogares, el tamaño de muestra realizado fue de 3.043 hogares (aplicación del cuestionario a una persona del hogar) en 58 municipios de Colombia. Este tamaño de muestra permitió reportar indicadores de resultados con máximo 2,1% de error de muestreo y confiabilidad de 95% a nivel de cada región. La tabla 5 presenta la desagregación de las encuestas efectivas realizadas por región y por clúster.

Tabla 5. Encuestas efectivas por región y clúster

Región	Alto y Moderado	Bajo	Incipiente	Limitado	Total	Margen de error
Antioquia	249	72	95	54	**470**	**5,3%**
Atlántica	232	154	146	107	**639**	**4,6%**
Centro	162	97	101	27	**387**	**5,9%**
Cundinamarca	321	77	71	13	**482**	**5,3%**
Oriente	195	116	116	75	**502**	**5,2%**
Suroccidente	204	123	123	113	**563**	**4,9%**
Total	**1.363**	**639**	**652**	**389**	**3.043**	**2,1%**

La tabla 6 muestra la desagregación de las encuestas efectivas realizadas por municipio.

Tabla 6. Encuestas efectivas por municipio

MUNICIPIO	Número de Encuestas
Acacías	79
Albania	85
Angostura	33
Aquitania	30
Armero	38
Arroyohondo	30
Barrancabermeja	116
Barranquilla	93
Betulia	40
Bogotá D.C.	286
Cajamarca	32
Carolina	25
Cartagena de Indias	81
Cartagena del Chairá	67
Cértegui	34
Concepción	21
El Carmen de Atrato	63
El Colegio	33
El Dorado	15
Elías	27
Gachalá	20
Gramalote	15
Guacamayas	25
Guacarí	90
Guaduas	38
Herveo	27
Honda	81
Jerusalén	13
La Ceja	32
Lérida	51
Medellín	217
Medina	26
Mitú	28
Moniquirá	61
Morroa	61

MUNICIPIO	Número de Encuestas
Nariño	33
Nueva granada	49
Palestina	50
Pueblo Bello	48
Puerto Nariño	16
Quimbaya	81
Recetor	22
Riofrío	27
Roldanillo	23
Samacá	55
San Cristóbal	28
San juan de Betulia	37
San Onofre	69
San Vicente Ferrer	32
Santa marta	58
Santiago	29
Santiago de Cali	181
Segovia	70
Sopó	19
Suaita	25
Tabio	16
Ventaquemada	31
Viotá	31
TOTAL	**3.043**

4 Factores de expansión

4.1 Metodología para el cálculo de factores de expansión

Los métodos de integración de datos son relativamente recientes en la literatura y han emergido como una solución al gran volumen de datos que se tiene actualmente. De esta forma, si una muestra presenta sesgos importantes entonces se pueden usar métodos de integración de datos con el propósito de lograr una inferencia a través de un modelo (Valliant 2023).

El enfoque de la inferencia basado en un modelo es un método alternativo al enfoque de la inferencia basado en el diseño. En la inferencia a partir de un modelo se tiene una muestra recolectada denotada como s_A y tenemos otra muestra que asegura la representatividad que es de tipo probabilístico y ha sido dirigida a la misma población objetivo, denotada como s_B. Para la implementación de la integración de datos se usa un soporte común, para lo cual se eligen unas variables $\mathbf{X}$ que se encuentran presentes en ambas muestras y la matriz $\mathbf{Y}$ que se refieren a las variables específicas de investigación. La siguiente tabla resume el tipo de información que se construye para la implementación del modelo de inferencia.

Muestra	Tipo	X	Y
s_B	Muestra ECV	✓	
s_A	Muestra CRC-OTT	✓	✓

En este caso se utilizó la Encuesta de Calidad de vida del DANE como fuente probabilística, luego de ajustar la cobertura para emparejar la población objetivo con el estudio de la CRC-OTT. El proceso consistió integrar esta fuente con la presente encuesta, esto permite lograr la representatividad en la segunda a partir de un modelo de integración de datos (Yang, Kim, y Song 2020). Los factores de expansión que permiten realizar una inferencia a toda la población se determinan a partir del modelo que estima probabilidad de inclusión en la muestra de CRC-OTT, que se obtiene de $\hat{\pi}_k^A = P(I_k^A = 1|\mathbf{x}_k) = \pi(\mathbf{x}_k, \hat{\boldsymbol{\theta}}_0)$, en donde $\hat{\boldsymbol{\theta}}_0$ corresponde a los valores que se obtienen de maximizar la siguiente función:

$$l(\boldsymbol{\theta}) = \sum_{k \in S_A} \log\left(\frac{\pi(\mathbf{x}_k, \boldsymbol{\theta})}{1 - \pi(\mathbf{x}_k, \boldsymbol{\theta})}\right) + \sum_{k \in S_B} d_k^B \log\left(1 - \pi(\mathbf{x}_k, \boldsymbol{\theta})\right)$$

En donde el log-verosimilitud está basado en la integración de las muestras, y d_k^B son los factores de expansión, que provienen de la ECV.

Para la estimación de hogares se aplicó un método propensión inversa (IPW por sus siglas en inglés) que consiste en los siguientes pasos bajo un modelo logístico para $\pi(\mathbf{x}_k, \widehat{\boldsymbol{\theta}}_0)$:

1. Sea I_k^A un indicador de selección dentro de la muestra CRC-OTT, s_A, de modo que $I_k^A = 1$ si $k \in s_A$ y 0 en otro caso.
2. Se ajusta el modelo $I_k^A \sim Bernoulli(\pi_k^A), logit(\pi_k^A) = \mathbf{x}_k^T\boldsymbol{\theta}$
3. Calcular el factor de propensión inversa como $w_k^0 = 1/\hat{\pi}_k^A$

La matriz de covariables incluyó la región, clúster, tenencia de computador fijo, computador portátil, Tablet, acceso a tv por suscripción, plataformas de streaming, conexión a internet fijo y tipo de señal de tv. La estimación de los parámetros del modelo se realiza con un método de estimación por ecuaciones generales (GEE por sus siglas en inglés), esto permite recuperar la representatividad de la muestra completa. Estas ecuaciones de pseudo puntaje $U(\boldsymbol{\theta}) = 0$ son derivadas de la función de log-verosimilitud $l(\boldsymbol{\theta})$ y pueden ser reemplazadas por un sistema de ecuaciones generales que permiten la estimación, para ello sea $h(x, \boldsymbol{\theta})$ un vector de funciones con la misma dimensión que $\boldsymbol{\theta}$, entonces:

$$U(\boldsymbol{\theta}) = \sum_{i \in S_A} h\,(x_i, \boldsymbol{\theta}) - \sum_{i \in S_B} \pi\,(x_i, \boldsymbol{\theta}) h(x_i, \boldsymbol{\theta}),$$

con $\mathbb{E}[U(\boldsymbol{\theta})] = 0$ para cualquier elección de $h(x, \boldsymbol{\theta})$. En principio, un estimador $\widehat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ puede obtenerse resolviendo $U(\boldsymbol{\theta}) = 0$ con la forma paramétrica elegida $\pi(x_i, \boldsymbol{\theta})$ y las funciones elegidas $h(x, \boldsymbol{\theta})$, se obtiene un estimador $\widehat{\boldsymbol{\theta}}$ que es consistente. De esta manera bajo el modelo logístico se tiene:

$$\mathbf{U}(\boldsymbol{\theta}) = \sum_{i \in S_A} \mathbf{x}_i - \sum_{i \in S_B} d_i^B\, \pi(\mathbf{x}_i, \boldsymbol{\theta}) \mathbf{x}_i.$$

Que es una función que depende de las variables auxiliares, los factores de expansión de la fracción probabilística y de la estimación de las propensiones.

4.2 Consistencia de los factores de expansión finales

En este paso se calibran los factores básicos de expansión de hogares obtenidos con el modelo, ajustándolos a las pirámides poblacionales de hogares reportadas por el DANE para el año 2025. Para el caso de las personas, el factor de expansión se construye multiplicando primero el factor de hogares calibrado por el número de personas mayores de 15 años que habitan en cada hogar, y posteriormente este resultado también se calibra para ajustarse a la distribución poblacional POST-COVID reportada por el DANE para el año 2025.

Para calibrar los factores de expansión es común aplicar metodologías de raking, postestratificación o estimadores de regresión generalizada (GREG) para corregir sesgos muestrales y no muestrales, y para mejorar la eficiencia de las estimaciones (Lu y Gelman 2003). En este caso se usó un modelo de raking debido a la multicolinealidad entre las restricciones de calibración. En el método de raking, solo se necesitan los conteos de control marginales de las variables auxiliares. El estimador ajustado por raking para hogares se asocia con un modelo que se puede escribir como:

$$\text{Minimizar:} \quad \sum_{k=1}^{N} w_k \log\left(\frac{w_k}{w_k^0}\right)$$

$$\text{Sujeto a:} \quad \sum_{k=1}^{N} w_k = N$$

$$\sum_{k=1}^{N} w_k \cdot R_{k,j,k} = T_{j,k}, \quad \forall (j,k) \text{ en región y area}$$

$$\sum_{k=1}^{N} w_k \cdot R_{k,j,l} = T_{j,l}, \quad \forall (j,l) \text{ en región y clúster}$$

$$\sum_{k=1}^{N} w_k \cdot R_{k,j,m} = T_{j,m}, \quad \forall (j,m) \text{ en región y NSE}$$

En donde:

- w_k: Peso ajustado del hogar k.
- w_k^0: Peso inicial del hogar k obtenido por el IPW.

- $R_{k,j,k}$: Indicador para la interacción entre la región (j) y el área (Urbano-Rural) (k).
- $R_{k,j,l}$: Indicador para la interacción entre la región (j) y el clúster del municipio (l).
- $R_{k,j,m}$: Indicador para la interacción entre la regíon (j) y NSE (Nivel socioeconómico) (m).
- $T_{j,k}$, $T_{j,l}$, $T_{j,m}$, T_n: Totales poblacionales conocidos para cada interacción o categoría.

Para la calibración de personas, el estimador ajustado se puede escribir como:

$$\text{Minimizar:} \quad \sum_{k=1}^{N} w_k \log\left(\frac{w_k}{w_k^0}\right)$$

$$\text{Sujeto a:} \quad \sum_{k=1}^{N} w_k = N$$

$$\sum_{k=1}^{N} w_k \cdot R_{k,j,k} = T_{j,k}, \quad \forall (j,k) \text{ en región y área}$$

$$\sum_{k=1}^{N} w_k \cdot R_{k,j,l} = T_{j,l}, \quad \forall (j,l) \text{ en región y cluster}$$

$$\sum_{k=1}^{N} w_k \cdot R_{k,j,m} = T_{j,m}, \quad \forall (j,m) \text{ en región y NSE}$$

$$\sum_{k=1}^{N} w_k \cdot R_{k,j,o} = T_{j,o}, \quad \forall (j,o) \text{ en región y sexo}$$

$$\sum_{k=1}^{N} w_k \cdot R_{k,j,p} = T_{j,p}, \quad \forall (j,o) \text{ en región y edad}$$

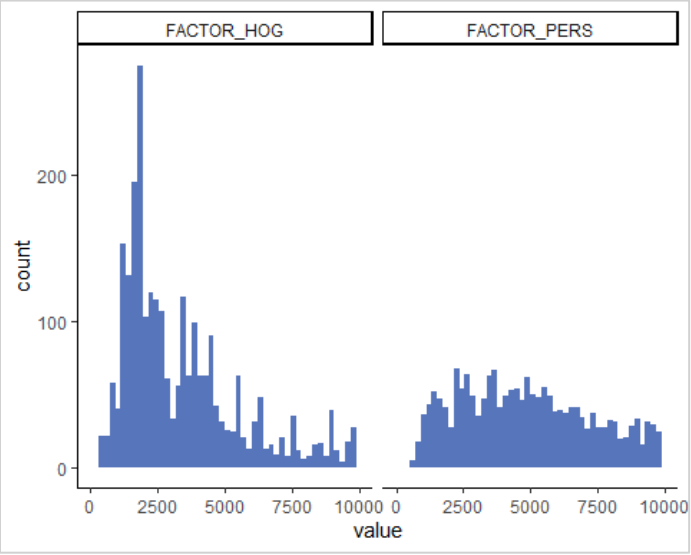
En donde:

- w_k: Peso ajustado del individuo k.
- w_k^0: Peso inicial del individuo k obtenido por el factor calibrado de hogares y la cantidad de personas en el hogar k
- $R_{k,j,k}$: Indicador para la interacción entre la región (j) y el área (Urbano-Rural) (k).

- $R_{k,j,l}$: Indicador para la interacción entre la región (j) y el clúster del municipio (l).
- $R_{k,j,m}$: Indicador para la interacción entre la región (j) y NSE (Nivel socioeconómico) (m).
- $R_{k,j,o}$: Indicador para la interacción entre la región (j) y el sexo (o).
- $R_{k,j,p}$: Indicador para la interacción entre la región (j) y el rango de edad (p).
- $T_{j,k}$, $T_{j,l}$, $T_{j,m}$, T_n, $T_{j,o}$, $T_{j,p}$: Totales poblacionales conocidos para cada interacción o categoría.

La función objetivo minimiza la divergencia de Kullback-Leibler entre los pesos ajustados (w_k) y los iniciales (w_k^0). Para la implementación se usa la librería *survey* (Lumley 2020) del software R que cuenta con el procedimiento implementado, la Figura 1. Histograma de factores de expansión calibrados ilustra la distribución de los factores de expansión a través de un histograma.

Figura 1. Histograma de factores de expansión calibrados



4.3 Soporte teórico de los modelos de integración de datos

Esta sección detalla los aspectos metodológicos para integrar fuentes de datos no probabilísticas con probabilísticas. Se enfoca en los supuestos, selección de variables, identificación de sesgos y estrategias de interoperabilidad.

4.3.1 Supuestos

La integración de fuentes con la ENCV requiere supuestos fundamentales para garantizar inferencias válidas.

a) Independencia Condicional: La inclusión en la muestra no probabilística S_A debe ser independiente de la variable de estudio y_i dado las covariables $\mathbf{x}_i$, esto es:

$$\pi_i^A = P(R_i = 1|\mathbf{x}_i, y_i) = P(R_i = 1|\mathbf{x}_i)$$

En donde π_i^A es la probabilidad de que el hogar i sea incluido en la muestra, R_i es una variable indicadora que vale 1 si la unidad i es seleccionada en la muestra y 0 en caso contrario. Este supuesto es análogo a "missing at random" (MAR) en datos faltantes y su violación introduce sesgos no corregibles sin información adicional.

b) Positividad: Implica que la probabilidad de inclusión en la muestra es mayor que cero para cualquier individuo de la población.

$$\pi_i^A > 0 \quad \forall i \in U$$

Esto se controló asegurando la cobertura de las subpoblaciones clave.

c) Variables de soporte consistentes

Las covariables $\mathbf{x}_i$ deben estar definidas y medidas de forma consistente en ambas fuentes de información, ya que discrepancias en las definiciones o en el nivel de agregación pueden comprometer la validez de la integración (Lohr, 2022).

$$\mathbf{x}_i^{\text{ECV}} \equiv \mathbf{x}_i^{\text{CRC-OTT}} \quad \text{(definiciones y mediciones)}$$

4.3.2 Selección de variables comunes

Para la selección de las variables comunes de la ECV y de la encuesta CRC-OTT se realizó una revisión detallada de los instrumentos de ambas encuestas, para lo cual se tuvieron en cuenta cuatro criterios:

1. **Relevancia para participación**: Variables que explican la probabilidad de inclusión.

2. **Poder predictivo**: Se eligen variables que se esperan que estén correlacionadas con las variables de resultado.

3. **Consistencia metodológica**: Mantener la misma definición y escalas de medición.

4. **Unidad de referencia**: que la unidad de referencia fuera siempre el hogar

4.3.3 Posibles sesgos derivados de la integración de fuentes

1. Sesgo por Violación del supuesto 1, de independencia condicional: Si $\mathbf{x}_i$ no captura factores que afectan tanto y_i como la inclusión en S_A entonces el sesgo va a persistir.

Solución: Usar diagnósticos para comparar modelos $z|\mathbf{u}$ para las variables disponibles en ambas fuentes (Chen, Li, y Wu 2020), en este caso se usaron dos fuentes probabilísticas y se eligió la que obtuvo el mejor ajuste.

b) Sesgo por ponderadores extremos: Ponderadores d_i^A con alta varianza aumentan el error cuadrático medio.

Solución: Truncar ponderadores. En este caso no se presentó esta situación, por lo que no fue necesario incorporar estrategias para recortar o trucar los ponderadores.

3. Sesgo por medición inconsistente: Si $\mathbf{x}_i$ difiere entre fuentes, en su contexto o en las escalas, los modelos fallan.

Solución: Armonizar definiciones pre-integración o usar métodos de calibración robusta (Zhang, 2019).

4.3.4 Mejoras para la interoperabilidad

Las siguientes recomendaciones se basan en Lohr (2022) y Meng (2021):

1. Usar muestras de enlace: Estas son muestras que contienen variables comunes con la muestra de estudio.

2. Metadatos estandarizados: Documentar definiciones y metodologías.

3. Validación cruzada: Comparar estimaciones con fuentes externas.

Frente al proceso de **recolección de información**, el segmento residencial fue realizado mediante encuesta presencial en hogares.

5 Estimadores y error relativo de muestreo

La metodología para la estimación de los errores estándar se basa en un muestreo complejo considerando los factores de expansión finales.

5.1. Estimadores

El estimador del total para una variable y_k medida en la persona k, que pertenece al estrato de muestreo h, se obtiene a partir del estimador de Horvitz-Thompson, así:

$$\hat{t}_{y\pi} = \sum_{k \in S} \frac{y_k}{\pi_k} = \sum_{k \in S} d_k y_k$$

En donde S es el conjunto de la muestra de unidades de observación, π_k es la pseudo-probabilidad de inclusión de la persona k, mientras que d_k es el factor de expansión asociado a la misma persona obtenido por el modelo de integración de datos con la ECV-DANE.

5.2. Estimación de la varianza

La estimación de la varianza sigue el procedimiento de un muestreo complejo de varias etapas. Para lograr la estimación de los parámetros se realiza el siguiente procedimiento:

- Se selecciona una muestra aleatoria simple con reemplazo de $\tilde{m}_h$ UPM de la muestra inicial de m_h. Sea m^*_{hi} el número de veces que la UPM i es seleccionada en la muestra.
- Se crean los nuevos ponderadores a partir de $d^*_k = d_k * m^*_h$, con d_k los factores de expansión de la investigación.
- Se calcula el estimador $\hat{\theta}$ usando d^*_k en vez de d_k.
- Se repite el proceso $B > 1$ veces. Denotando los correspondientes estimadores de Bootstrap como $\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots \hat{\theta}_{(B)}$

El estimador de la varianza basado en el método bootstrap de Rao-Wu es:

$$\hat{V}_{BT}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\theta}_{(b)} - \hat{\theta}\right)^2$$

En donde $\hat{\theta}$ es la estimación del parámetro de interés usando los factores de expansión sin el procedimiento de réplicas (con la muestra original). Este procedimiento se encuentra incorporado en el paquete survey de R en sus diferentes procedimientos: svyglm, svytotal, svymean, entre otras.

5.3. Error relativo de muestreo

El error relativo de muestreo se mide a partir del coeficiente de variación estimado, así:

$$cve(\hat{\theta}_y) = \frac{\sqrt{\hat{V}(\hat{\theta}_y)}}{\hat{\theta}_y} \times 100\%$$

El intervalo de confianza al $(1-\alpha)\%$ se obtiene como:

$$\hat{\theta}_y \pm t_{\alpha/2,df} \times \sqrt{\hat{V}(\hat{\theta}_y)}$$

donde $df =$ número de estratos. El estimador $\hat{\theta}$ corresponde al estimador del total o de la media según corresponda.

6 Metodología aplicada para la validación de hipótesis

La validación de las hipótesis planteadas se realizó mediante la estimación de modelos de regresión estadística. No obstante, los enfoques tradicionales de regresión resultan insuficientes, ya que ignorar los pesos muestrales, la estratificación o la conglomeración puede derivar en sesgos en los coeficientes estimados y, sobre todo, en una subestimación de sus errores estándar, lo que compromete la validez de las inferencias. Por esta razón, los modelos de regresión aplicados a encuestas deben ser modificados y ajustados para garantizar resultados representativos y robustos.

Por lo anterior, el ajuste de los modelos se basó en la especificación de un objeto que contiene el diseño muestral, en donde se definen los estratos, conglomerados, las unidades del muestreo y factores de expansión. Para ello se utilizaron los paquetes survey y srvyr del paquete R, mediante la función svyglm,

seleccionando la familia de distribución y el método de estimación de acuerdo con la naturaleza de la variable dependiente en cada caso:

1. Variables binarias: **Regresión logística** con enlace logit.
2. Variables de conteo: Modelo lineal generalizado (GLM):
   → **Quasi-Poisson** para datos de conteo con sobredispersión moderada[1].
   → **Binomial Negativa** para datos de conteo con sobredispersión severa.

El modelo general estimado se expresa como:

$$g(E(y_i)) = \alpha + \beta z_i + \theta X_i$$

donde:

- y_i es la variable dependiente,
- α es el intercepto,
- z_i es la variable independiente de interés,
- X_i es un vector de variables de control,
- β y θ son los coeficientes asociados, y
- $g(\cdot)$ es la función de enlace correspondiente a cada familia de distribución.

El método de estimación y la interpretación de los coeficientes dependen de la naturaleza de y_i:

a. Modelo de **Regresión Logística**

Este modelo se empleó cuando la variable dependiente fue binaria, es decir, cuando la respuesta toma valores 0 o 1 (por ejemplo, presencia/ausencia de un evento). El modelo se especifica de la siguiente manera:

[1] ***Nota:*** *Se considera sobre dispersión moderada cuando la diferencia entre la varianza y la media no es significativa (en el presente estudio la mayoría de las pruebas de hipótesis en las que se empleó un modelo Quasi-Poisson, presentan una diferencia entre estos dos valores de menos de 10). En los casos en que la varianza es mucho mayor que la media, se considera sobre dispersión severa.*

$$ln\left(\frac{\pi}{1-\pi}\right) = X^T\boldsymbol{\beta} + \varepsilon_i$$

Y la probabilidad estimada que utiliza el modelo logístico es la siguiente:

$$P(Y = 1 \mid X) = \frac{e^{X^T\beta}}{1 + e^{X^T\beta}}$$

En donde β es el vector de parámetros que cuantifica los efectos sobre la probabilidad del suceso y que se estiman usando un método de máximo verosimilitud ponderada teniendo en cuenta el diseño muestral.

b. Modelo lineal generalizado con respuesta **Quasi-Poisson**

Cuando la variable dependiente fue numérica discreta (conteos) y no presentó evidencia de sobredispersión significativa, se utilizó un GLM con familia Quasi-Poisson. El modelo adopta la siguiente forma:

$$g(E(y_i)) = \log\left(E(y_i)\right) = \alpha + \beta z_i + \theta X_i$$

La estimación se realiza mediante mínimos cuadrados ponderados generalizados.

c. Modelo Lineal Generalizado con respuesta **Binomial Negativa**

Cuando la variable dependiente también fue de conteo, pero con una sobredispersión más pronunciada o estructural, se aplicó un modelo de regresión Binomial Negativa, estimado mediante el método de máxima verosimilitud. El modelo se expresa como:

$$\log\left(E(y_i)\right) = \alpha + \beta z_i + \theta X_i$$

y asume que la varianza de y_i sigue la forma $\text{Var}(y_i) = E(y_i) + k\,E(y_i)^2$, donde k es un parámetro de dispersión.

En todas las pruebas de hipótesis incluyó un conjunto consistente de variables de control: región, rango de edad, estrato socioeconómico y sexo; en aquellos modelos donde alguna de estas variables era el objeto de la hipótesis a evaluar

(por ejemplo, edad o estrato), dicha variable no se trató como control sino como variable de análisis.

El objeto del diseño usado en cada hipótesis dependió del nivel de análisis que se estableció, nivel de persona u hogar, según correspondiera, con el objetivo de asegurar la representatividad de las estimaciones a nivel poblacional y aislar el efecto de la variable de interés frente a diferencias estructurales entre las unidades analizadas.

Debido a la especificación logarítmica de los modelos analizados, que pertenecen a la familia de modelos lineales generalizados (GLM por sus siglas en inglés), la interpretación de los coeficientes se realiza a través de su exponencial, $\exp(\beta)$. De esta manera, cada hipótesis fue contrastada considerando el signo, la magnitud y la significancia estadística del coeficiente asociado a la variable de interés mediante el test de Wald, manteniendo constantes las variables de control. En los casos en que la variable de interés corresponde a una variable categórica, se agrega adicionalmente el p-valor Chi-cuadrado.

A continuación, se presenta la estimación de los coeficientes y la respectiva validación de hipótesis partiendo de la metodología descrita previamente. Se consideró como umbral de decisión un nivel de significancia inferior al 10%.

7 Resultados validación de hipótesis

Teniendo en cuenta la metodología para la validación detallada previamente, se verificó cada prueba de hipótesis de acuerdo con el nivel de agregación descrito a continuación: todas las hipótesis de servicios mensajería móviles, mensajería y envío de paquetes, radiodifusión, de voz fija, y voz móvil fueron estimadas usando el factor de expansión a nivel de personas. Sobre las hipótesis que corresponden a los servicios audiovisuales, cuatro de ellas son estimadas a partir de expansores estadísticos a nivel de hogar dado que corresponden al uso de TV paga básica en el hogar, estrato socioeconómico del mismo y gasto en plan de internet fijo. El factor de expansión utilizado se puede

evidenciar observando la población expandida en la tabla de resultados para cada hipótesis.

6.1 Servicios audiovisuales

H2: "La probabilidad de que un hogar esté suscrito al servicio de TV paga-básica ofrecido por un operador tradicional se relaciona negativamente con la utilización del servicio audiovisual ofrecido por un OSP pago."

- Variable dependiente: Suscripción a TV paga básica
- Variable independiente: Uso servicios OSP pago
- Metodología: Modelo de regresión logística

Tabla 7. Resultados H2 servicios audiovisuales

HIPÓTESIS 2		
β	**p-value**	$Exp(\beta)$
2,222	0,000	9,230
Población expandida		
18.777.420		
Controles		
Estrato Género Rangos de edad Región		

- *Conclusión*: **Se rechaza la hipótesis.**
- Interpretación: A medida que aumenta la proporción de uso del servicio audiovisual ofrecido por un OSP pago, también aumenta la probabilidad de que un hogar esté suscrito al servicio de TV paga-básica tradicional. Esto contradice la hipótesis inicial, ya que la relación encontrada es positiva y significativa, no negativa.

H5: "Un aumento marginal en la frecuencia de uso de los servicios audiovisuales ofrecidos por un OSP se relaciona negativamente con la frecuencia de uso de los servicios de TV paga-básica."

- Variable dependiente: Frecuencia uso TV paga básica
- Variable independiente: Frecuencia uso OSP
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 8. Resultados H5 servicios audiovisuales

HIPÓTESIS 5		
β	**p-value**	$Exp(\beta)$
0,134	0,000	1,144
Población expandida		
18.777.420		
Controles		
Estrato Género Rangos de edad Región		

- *Conclusión*: **Se rechaza la hipótesis.**
- Interpretación: A medida que aumenta la frecuencia de uso de servicios audiovisuales ofrecidos por un OSP, también aumenta la frecuencia de uso de los servicios de TV paga-básica. Esto contradice la hipótesis inicial, ya que la relación encontrada es positiva y significativa, no negativa.

H7: "La edad de los usuarios se relaciona negativamente con el uso de los servicios audiovisuales ofrecidos por un OSP."

- Variable dependiente: Uso servicios OSP
- Variable independiente: Edad
- Metodología: Modelo de regresión logística

Tabla 9. Resultados H7 servicios audiovisuales

HIPÓTESIS 7			
Edad	β	**p-value**	$Exp(\beta)$
18 a 24	-0,440	0,240	0,643
25 a 34	-0,969	0,018	0,379
35 a 44	-1,421	0,002	0,241
45 a 54	-1,690	0,002	0,184
55 o más	-2,456	0,000	0,086
Chi-cuadrado		0,000	
Población expandida			
41.646.171			
Controles			
Estrato Género Región			

- *Conclusión*: **Se acepta la hipótesis**.
- Interpretación: A medida que aumenta la edad, disminuye la probabilidad de uso de servicios audiovisuales ofrecidos por un OSP. Esto confirma la hipótesis inicial, ya que se encuentra una relación negativa y significativa entre la edad y el uso de estos servicios.

H8: "Los estratos socioeconómicos altos hacen un mayor uso de los servicios audiovisuales ofrecidos por un OSP."

- Variable dependiente: Uso servicios OSP
- Variable independiente: Estrato socioeconómico
- Metodología: Modelo de regresión logística

Tabla 10. Resultados H8 servicios audiovisuales

HIPÓTESIS 8			
Edad	β	**p-value**	$Exp(\beta)$
Estrato 2	0,372	0,003	1,451
Estrato 3	0,693	0,004	2,000
Estrato 4	1,883	0,000	6,574
Estrato 5	1,502	0,000	4,490
Estrato 6	1,133	0,358	3,107
Chi-cuadrado		0,000	
Población expandida			
18.777.420			
Controles			
Género Región Rangos de edad			

- *Conclusión*: **Se acepta la hipótesis**.
- Interpretación: A medida que aumenta el nivel socioeconómico, también aumenta la probabilidad de uso de servicios audiovisuales ofrecidos por un OSP. Esto confirma la hipótesis inicial, ya que se encuentra una relación positiva y significativa entre el estrato y el uso de estos servicios.

H10: "Los usuarios que utilizan el servicio audiovisual ofrecido por un OSP con mayor intensidad tienen una mayor propensión a gastar más en planes de Internet fijo."

- Variable dependiente: Gasto en plan internet fijo
- Variable independiente: Uso intensivo servicios OSP
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 11. Resultados H10 servicios audiovisuales

HIPÓTESIS 10		
β	**p-value**	$Exp(\beta)$
0,538	0,000	1,713
Población expandida		
18.777.420		
Controles		
Estrato Género Rangos de edad Región		

- *Conclusión*: **Se acepta la hipótesis.**
- Interpretación: A mayor intensidad en el uso diario de servicios audiovisuales ofrecidos por un OSP, mayor es la cantidad de megas contratados en el plan de Internet fijo del hogar. Esto confirma la hipótesis inicial, ya que la relación encontrada es positiva y significativa.

H Nueva: "Un aumento marginal en la frecuencia de consumo de los contenidos audiovisuales ofrecidos por un OSP gratuito se relaciona negativamente con la frecuencia de consumo de los contenidos de la TV abierta."

- Variable dependiente: Frecuencia consumo TV abierta
- Variable independiente: Frecuencia uso OSP gratuito
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 12. Resultados H Nueva servicios audiovisuales

HIPÓTESIS Nueva		
β	**p-value**	$Exp(\beta)$
0,157	0,000	1,170
Población expandida		
41.646.171		
Controles		
Estrato Género Rangos de edad Región		

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: A medida que aumenta la frecuencia de consumo de contenidos audiovisuales ofrecidos por un OSP gratuito, también aumenta la frecuencia de consumo de contenidos de TV abierta. Esto contradice la hipótesis inicial, ya que la relación encontrada es positiva y significativa, no negativa.

Tabla 13. Resultados de servicios audiovisuales – tendencia

	H2	H5	H7	H8	H10	H11 (Nueva)
Hipótesis	La probabilidad de que un hogar esté suscrito al servicio de TV paga-básica ofrecido por un operador tradicional se relaciona negativamente con la utilización del servicio audiovisual ofrecido por un OSP pago.	Un aumento marginal en la frecuencia de uso de los servicios audiovisuales ofrecidos por un OSP se relaciona negativamente con la frecuencia de uso de los servicios de TV paga-básica.	La edad de los usuarios se relaciona negativamente con el uso de los servicios audiovisuales ofrecidos por un OSP.	Los estratos socio económicos altos hacen un mayor uso de los servicios audiovisuales ofrecidos por un OSP.	Los usuarios que utilizan el servicio audiovisual ofrecido por un OSP con mayor intensidad tienen una mayor propensión a gastar más en planes de Internet fijo.	Un aumento marginal en la frecuencia de consumo de los contenidos audiovisuales ofrecidos por un OSP gratuito se relaciona negativamente con la frecuencia de consumo de los contenidos de la TV abierta.
2018	Se rechaza	-	Se acepta	Se rechaza	Se rechaza	-
2019	Se rechaza	-	-	-	Se acepta	-
2021	Se rechaza	Se rechaza	Se acepta	Se rechaza	Se rechaza	-
2022	Se rechaza	Se rechaza	Se acepta	Se acepta	Se acepta	-
2023	Se acepta	Se acepta	Se rechaza	Se acepta	Se acepta	-
2024	Se acepta	Se rechaza	Se acepta	Se rechaza	Se acepta	Se acepta
2025	Se rechaza	Se rechaza	Se acepta	Se acepta	Se acepta	Se rechaza

6.2 Servicios de Mensajería Móviles

H2: "Un aumento marginal en la frecuencia de uso de los servicios de mensajería móvil ofrecidos por un OSP reduce la frecuencia de uso de los servicios de mensajería móvil tradicional."

- Variable dependiente: Frecuencia uso SMS por operador
- Variable independiente: Frecuencia uso mensajería por apps
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 14. Resultados H2 servicios de Mensajería Móviles

HIPÓTESIS 2		
β	**p-value**	$Exp(\beta)$
-0,027	0,006	0,972
Población expandida		
41.646.171		
Controles		
Estrato Género Rangos de edad Región		

- *Conclusión*: **Se acepta la hipótesis**.
- Interpretación: A medida que aumenta la frecuencia de uso de servicios de mensajería móvil ofrecidos por un OSP, disminuye la cantidad de mensajes enviados por medios tradicionales (SMS). Esto confirma la hipótesis inicial, ya que se encuentra una relación negativa y significativa entre ambas formas de mensajería.

H4: "Los estratos socioeconómicos altos hacen mayor uso de los servicios de mensajería móvil ofrecidos por un OSP."

- Variable dependiente: Uso mensajería OSP
- Variable independiente: Estrato socioeconómico
- Metodología: Modelo lineal generalizado con respuesta Quasi-Poisson

Tabla 15. Resultados H4 servicios de Mensajería Móviles

HIPÓTESIS 4			
Edad	β	**p-value**	$Exp(\beta)$
Estrato 2	-0,011	0,625	0,989
Estrato 3	-0,001	0,940	1,002
Estrato 4	-0,039	0,065	0,962
Estrato 5	-0,007	0,834	0,993
Estrato 6	0,156	0,010	1,170
Chi-cuadrado		0,000	
Población expandida			
31.444.168			
Controles			
Género Rangos de edad Región			

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: No se encuentra una asociación general significativa entre los estratos altos y el uso de estos servicios.

H6: "Los usuarios que utilizan el servicio de mensajería móvil ofrecido por un OSP con mayor intensidad tienen una mayor propensión a gastar más en planes móviles de alto valor."

- Variable dependiente: Gasto en planes móviles alto valor
- Variable independiente: Uso intensivo mensajería OSP
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 16. Resultados H6 servicios de Mensajería Móviles

HIPÓTESIS 6		
β	**p-value**	$Exp(\beta)$
-0,017	0,059	0,982
Población expandida		
5.409.428		
Controles		
Estrato Género Región Rangos de edad		

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: A medida que aumenta la intensidad de uso del servicio de mensajería móvil ofrecido por un OSP, se reduce la propensión a estar en un plan pospago de mayor valor. Esto contradice la hipótesis inicial, ya que se encuentra una relación negativa y significativa entre el uso intensivo de mensajería OSP y el gasto en planes móviles.

Tabla 17. Resultados de servicios de mensajería móviles – tendencia

	H2	H4	H6
Hipótesis	Un aumento marginal en la frecuencia de uso de los servicios de mensajería móvil ofrecidos por un OSP reduce la frecuencia de uso de los servicios de mensajería móvil tradicional.	Los estratos socioeconómicos altos hacen mayor uso de los servicios de mensajería móvil ofrecidos por un OSP.	Los usuarios que utilizan el servicio de mensajería móvil ofrecido por un OSP con mayor intensidad tienen una mayor propen sión a gastar más en planes móviles de alto valor.
2018	-	-	Se rechaza
2019	-	Se rechaza	Se acepta
2021	Se acepta	Se rechaza	Se rechaza
2022	Se rechaza	Se acepta	Se rechaza
2023	Se rechaza	Se acepta	Se acepta
2024	Se acepta	Se acepta	Se acepta
2025	Se acepta	Se rechaza	Se rechaza

6.3 Servicios de Mensajería y envío de paquetes

H1: "La probabilidad de que un usuario utilice medios tradicionales para envío de paquetes se relaciona negativamente con la utilización de aplicaciones para su envío."

- Variable dependiente: Uso medios tradicionales para envío de paquetes
- Variable independiente: Uso apps para envío de paquetes
- Metodología: Modelo de regresión logística

Tabla 18. Resultados H1 servicios de Mensajería y envío de paquetes

HIPÓTESIS 1		
β	**p-value**	$Exp(\beta)$
-38,311	0,000	0,000
Población expandida		
3.860.777		
Controles		
Región Rangos de edad		

- *Conclusión*: **Se acepta la hipótesis**.
- Interpretación: Los usuarios que usan aplicaciones para envío de paquetes tienen menor probabilidad de usar medios tradicionales para el mismo fin. Esto confirma la hipótesis, ya que la relación encontrada es negativa y significativa.

H2: "La probabilidad de que un usuario utilice medios tradicionales para envío de documentos se relaciona negativamente con la utilización aplicaciones para su envío."

- Variable dependiente: Uso medios tradicionales para envío de documentos
- Variable independiente: Uso apps para envío de documentos
- Metodología: Modelo de regresión logística

Tabla 19. Resultados H2 servicios de Mensajería y envío de paquetes

	HIPÓTESIS 2		
	β	**p-value**	$Exp(\beta)$
	-43,828	0,000	0,000
	Población expandida		
	2.971.343		
	Controles		
	Estrato Género Región		

- *Conclusión*: **Se acepta la hipótesis**.
- Interpretación: Los usuarios que usan aplicaciones para el envío de documentos tienen menor probabilidad de usar medios tradicionales para el mismo fin. Esto confirma la hipótesis, ya que la relación encontrada es negativa y significativa.

H3 (Nueva): "La probabilidad de que un usuario utilice medios tradicionales para envío o recepción de dinero se relaciona negativamente con la utilización aplicaciones para su envío."

- Variable dependiente: Uso medios tradicionales para envío o recepción de dinero
- Variable independiente: Uso apps o páginas web para envío o recepción de dinero
- Metodología: Modelo de regresión logística

Tabla 20. Resultados H3 servicios de Mensajería y envío de paquetes

HIPÓTESIS 3		
β	**p-value**	$Exp(\beta)$
-19,514	0,000	0,000
Población expandida		
12.558.209		
Controles		
Estrato Región		

- *Conclusión*: **Se acepta la hipótesis**.
- Interpretación: Los usuarios que usan aplicaciones para enviar o recibir dinero tienen menor probabilidad de usar medios tradicionales para el mismo fin. Esto confirma la hipótesis, ya que la relación encontrada es negativa y significativa.

Tabla 21. Resultados de servicios de mensajería y envío de paquetes – tendencia

	H1	H2	H3 (Nueva)
Hipótesis	La probabilidad de que un usuario utilice medios tradicionales para envío de paquetes se relaciona negativamente con la utilización aplicaciones para su envío.	La probabilidad de que un usuario utilice medios tradicionales para envío de documentos se relaciona negativamente con la utilización aplicaciones para su envío.	La probabilidad de que un usuario utilice medios tradicionales para envío o recepción de dinero se relaciona negativamente con la utilización aplicaciones para su envío.
2018	-	-	-
2019	-	-	-
2021	**	**	-
2022	**	**	-
2023	Se rechaza	Se rechaza	-
2024	Se rechaza	Se rechaza	-
2025	Se acepta	Se acepta	Se acepta

6.4 Servicios de Radiodifusión

H1: Un aumento en la intensidad de uso (tiempo) de servicios de música/ noticias/ entretenimiento a través de internet disminuye la frecuencia de uso de emisoras radiodifundidas.

- Variable dependiente: Frecuencia uso emisoras tradicionales
- Variable independiente: Uso OSP de audio
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 22. Resultados H1 servicios de radiodifusión

HIPÓTESIS 1		
β	**p-value**	$Exp(\beta)$
-0,005	0,171	0,995
Población expandida		
41.646.171		
Controles		
Estrato Género Región Rangos de edad		

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: No hay evidencia suficiente para afirmar que existe una relación negativa entre ambas formas de consumo. Por lo tanto, se rechaza la hipótesis, ya que no se encuentra una asociación significativa entre el tiempo dedicado a servicios en línea y la frecuencia de escucha de radio tradicional.

H3: "Las personas jóvenes hacen un mayor uso de servicios de música a través de un OSP."

- Variable dependiente: Uso música por OSP
- Variable independiente: Edad
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 23. Resultados H3 servicios de radiodifusión

HIPÓTESIS 3			
Edad	β	**p-value**	$Exp(\beta)$
18 a 24	-0,545	0,015	0,579
25 a 34	-0,990	0,002	0,371
35 a 44	-1,229	0,000	0,293
45 a 54	-1,485	0,000	0,227
55 o más	-3,084	0,000	0,043
Chi-cuadrado		0,000	
Población expandida			
41.646.171			
Controles			
Estrato Género Región			

- *Conclusión*: **Se acepta la hipótesis.**
- Interpretación: A medida que aumenta la edad, disminuye el tiempo de escucha de música por servicios OSP. Esto confirma la hipótesis, ya que se encuentra una relación negativa y significativa entre la edad y el uso de servicios musicales por internet.

H5: "El uso de servicios de música a través de OSP pagos y no pagos disminuye la frecuencia de uso de emisoras radiodifundidas."

- Variable dependiente: Frecuencia uso radio tradicional
- Variable independiente: Uso música por OSP
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 24. Resultados H5 servicios de radiodifusión

HIPÓTESIS 5		
β	**p-value**	$Exp(\beta)$
-0,007	0,093	0,992
Población expandida		
41.646.171		
Controles		
Estrato Género Rangos de edad Región		

- *Conclusión*: **Se acepta la hipótesis.**
- Interpretación: A medida que aumenta el uso de servicios de música por OSP, disminuye el tiempo de escucha de radio tradicional. Esto confirma la hipótesis, ya que se encuentra una relación negativa, aunque débil, entre ambas formas de consumo de audio.

Tabla 25. Resultados de servicios de radiodifusión – tendencia

	H1	H3	H5
Hipótesis	Un aumento en la intensidad de uso (uso (tiempo) de servicios de música/ noticias/ entretenimiento a través de internet disminuye la frecuencia de uso de emisoras radiodifundidas.	Las personas jóvenes hacen un mayor uso de servicios de música a través de un OSP.	El uso de servicios de música a través de OSP pagos y no pagos disminuye la frecuencia de uso de emisoras radiodifundidas.
2018	-	-	-
2019	-	-	-
2021	Se rechaza	Se acepta	Se acepta
2022	Se rechaza	Se rechaza	Se acepta
2023	Se rechaza	Se rechaza	Se rechaza
2024	Se rechaza	Se rechaza	Se rechaza
2025	Se rechaza	Se acepta	Se acepta

6.5 Servicios de Voz fija

H3: "La probabilidad de que un usuario utilice teléfono fijo para llamadas nacionales se relaciona negativamente con el uso de aplicaciones en línea para hacer o recibir llamadas."

- Variable dependiente: Uso teléfono fijo para llamadas nacionales
- Variable independiente: Uso de aplicaciones para llamadas
- Metodología: Modelo de regresión logística

Tabla 26. Resultados H3 servicios de voz fija

	HIPÓTESIS 3		
	β	**p-value**	$Exp(\beta)$
	0,261	0,388	1,299
	Población expandida		
	41.646.171		
	Controles		
	Estrato Género Región Rangos de edad		

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: No hay evidencia estadística suficiente para afirmar que existe una relación entre ambas formas de comunicación. Por lo tanto, se rechaza la hipótesis, ya que no se encuentra una asociación significativa entre el uso de aplicaciones para llamadas y el uso del teléfono fijo.

Tabla 27. Resultados de servicios de voz fija – tendencia

	H3
Hipótesis	La probabilidad de que un usuario utilice teléfono fijo para llamadas nacionales se relaciona negativamente con el uso de aplicaciones en línea para hacer o recibir llamadas.
2018	-
2019	-
2021	Se rechaza
2022	Se rechaza
2023	Se rechaza
2024	Se rechaza
2025	Se rechaza

6.6 Servicios de Voz móvil

H1: "La probabilidad de que un usuario utilice el servicio de voz móvil ofrecido por un operador tradicional se relaciona negativamente con la utilización del servicio de voz ofrecido por un OSP**."**

- Variable dependiente: Uso voz móvil tradicional
- Variable independiente: Uso del servicio de voz ofrecido por un OSP
- Metodología: Modelo de regresión logística

Tabla 28. Resultados H1 servicios de voz móvil

HIPÓTESIS 1		
β	**p-value**	$Exp(\beta)$
0,794	0,002	2,212
Población expandida		
38.464.061		
Controles		
Estrato Género Región Rangos de edad		

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: Los usuarios que usan aplicaciones para llamadas también tienen mayor probabilidad de usar el servicio de voz móvil tradicional. Esto contradice la hipótesis, ya que la relación encontrada es positiva y significativa, no negativa.

H2: "La probabilidad de que un usuario utilice el servicio de voz móvil ofrecido por un operador tradicional se relaciona negativamente con la frecuencia de uso del servicio de voz ofrecido por un OSP."

- Variable dependiente: Uso voz móvil tradicional
- Variable independiente: Frecuencia voz por OSP
- Metodología: Modelo de regresión logística

Tabla 29. Resultados H2 servicios de voz móvil

HIPÓTESIS 2		
β	**p-value**	$Exp(\beta)$
0,012	0,675	1,012
Población expandida		
38 464.061		
Controles		
Estrato Género Región Rangos de edad		

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: No se encuentra una asociación general significativa.

H3: "Un aumento marginal en la frecuencia de uso de los servicios de voz ofrecidos por un OSP se relaciona negativamente con la frecuencia de uso de los servicios de voz móvil tradicionales."

- Variable dependiente: Frecuencia voz móvil tradicional
- Variable independiente: Frecuencia voz por OSP
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 30. Resultados H3 servicios de voz móvil

HIPÓTESIS 3		
β	**p-value**	$Exp(\beta)$
-0,136	0,000	0,872
Población expandida		
38.464.061		
Controles		
Estrato Género Rangos de edad Región		

- *Conclusión*: **Se acepta la hipótesis**.
- Interpretación: A medida que aumenta la frecuencia de uso del servicio de voz por OSP, disminuye la frecuencia de uso del servicio de voz móvil tradicional. Esto confirma la hipótesis, ya que se encuentra una relación negativa y significativa entre ambas formas de comunicación.

H4: "La edad de los usuarios se relaciona negativamente con la frecuencia de uso de los servicios de voz ofrecidos por un OSP."

- Variable dependiente: Frecuencia voz OSP
- Variable independiente: Edad
- Metodología: Modelo lineal generalizado con respuesta Quasi-Poisson

Tabla 31. Resultados H4 servicios de voz móvil

HIPÓTESIS 4			
Edad	β	**p-value**	$Exp(\beta)$
18 a 24	-0,243	0,030	0,783
25 a 34	-0,328	0,001	0,720
35 a 44	-0,294	0,003	0,744
45 a 54	-0,517	0,000	0,596
55 o más	-0,674	0,000	0,509
Chi-cuadrado		0,000	
Población expandida			
38.464.061			
Controles			
Estrato Género Región			

- *Conclusión*: **Se acepta la hipótesis**.
- Interpretación: A medida que aumenta la edad, disminuye la frecuencia de uso de servicios de voz por OSP. Esto confirma la hipótesis, ya que se encuentra una relación negativa y significativa entre la edad y el uso de servicios de voz por internet.

H5: "Los estratos socioeconómicos altos tienen mayor frecuencia de uso de los servicios de voz ofrecidos por un OSP."

- Variable dependiente: Frecuencia voz OSP
- Variable independiente: Estrato socioeconómico
- Metodología: Modelo lineal generalizado con respuesta Quasi-Poisson

Tabla 32. Resultados H5 servicios de voz móvil

HIPÓTESIS 5			
Edad	β	**p-value**	$Exp(\beta)$
Estrato 2	0,087	0,219	0,915
Estrato 3	0,069	0,321	0,933
Estrato 4	-0,065	0,384	0,982
Estrato 5	-0,035	0,817	0,965
Estrato 6	-3,430	0,015	0,032
Chi-cuadrado		0,000	
Población expandida			
38.464.061			
Controles			
Género Rangos de edad Región			

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: Los coeficientes para los estratos socioeconómicos ESTRATO2 a ESTRATO5 no son estadísticamente significativos, y el coeficiente para ESTRATO6 resulta negativo y significativo. Esto contradice la hipótesis inicial, ya que no se observa una relación positiva y significativa entre el nivel socioeconómico y el uso de servicios de voz.

H7: "Los usuarios que utilizan el servicio de voz ofrecido por un OSP con mayor intensidad tienen una mayor propensión a gastar más en planes móviles de alto valor."

- Variable dependiente: Gasto en plan móvil
- Variable independiente: Intensidad uso voz OSP
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 33. Resultados H7 servicios de voz móvil

HIPÓTESIS 7		
β	**p-value**	$Exp(\beta)$
0,002	0,927	1,002
Población expandida		
5.409.428		
Controles		
Estrato Género Región Rangos de edad		

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: No se encuentra una asociación general significativa.

H8: "Un aumento marginal en la frecuencia de uso de los servicios de mensajería móvil ofrecidos por un OSP disminuye la frecuencia de uso de los servicios de voz móvil tradicionales."

- Variable dependiente: Frecuencia voz tradicional
- Variable independiente: Frecuencia uso mensajería OSP
- Metodología: Modelo lineal generalizado con respuesta Quasi-Poisson

Tabla 34. Resultados H8 servicios de voz móvil

HIPÓTESIS 8		
β	**p-value**	$Exp(\beta)$
0,024	0,000	1,024
Población expandida		
38.464.061		
Controles		
Estrato Género Región Rangos de edad		

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: A medida que aumenta el uso de servicios de mensajería por OSP, también aumenta la frecuencia de uso de servicios de voz móvil tradicionales. Esto contradice la hipótesis, ya que se encuentra una relación positiva y significativa, en lugar de negativa, entre ambas formas de comunicación.

H9: El uso de los servicios de mensajería móvil ofrecidos por un OSP reduce la frecuencia de uso de los servicios de voz móvil tradicionales.

- Variable dependiente: Frecuencia uso voz móvil tradicional
- Variable independiente: Uso mensajería OSP
- Metodología: Modelo lineal generalizado con respuesta binomial negativa

Tabla 35. Resultados H9 servicios de voz móvil

HIPÓTESIS 9		
β	**p-value**	$Exp(\beta)$
0,961	0,000	2,616
Población expandida		
38.464.061		
Controles		
Estrato Género Rangos de edad Región		

- *Conclusión*: **Se rechaza la hipótesis**.
- Interpretación: A medida que aumenta el uso de aplicaciones de mensajería por OSP, también aumenta la frecuencia de uso de servicios de voz móvil tradicionales. Esto contradice la hipótesis, ya que se encuentra una relación positiva y significativa, en lugar de negativa, entre ambas formas de comunicación.

Tabla 36. Resultados de servicios de voz móvil – tendencia

	H1	H2	H3	H4	H5	H7	H8	H9
Hipótesis	La probabilidad de que un usuario utilice el servicio de voz móvil ofrecido por un operador tradicional se relaciona negativamente con la utilización del servicio de voz ofrecido por un OSP.	La probabilidad de que un usuario utilice el servicio de voz móvil ofrecido por un operador tradicional se relaciona negativamente con la frecuencia de uso del servicio de voz ofrecido por un OSP.	Un aumento marginal en la frecuencia de uso de los servicios de voz ofrecidos por un OSP se relaciona negativamente con la frecuencia de uso de los servicios de voz móvil tradicionales.	La edad de los usuarios se relaciona negativamente con la frecuencia de uso de los servicios de voz ofrecidos por un OSP.	Los estratos socioeconómicos altos tienen mayor frecuencia de uso de los servicios de voz ofrecidos por un OSP.	Los usuarios que utilizan el servicio de voz ofrecido por un OSP con mayor intensidad tienen una mayor propensión a gastar más en planes móviles de alto valor.	Un aumento marginal en la frecuencia de uso de los servicios de mensajería móvil ofrecidos por un OSP disminuye la frecuencia de uso de los servicios de voz móvil tradicionales.	El uso de los servicios de mensajería móvil ofrecidos por un OSP reduce la frecuencia de uso de los servicios de voz móvil tradicionales.
2018	Se rechaza	-	-	Se acepta	Se rechaza	Se rechaza	Se rechaza	-
2019	Se acepta	Se acepta	-	Se acepta	Se rechaza	Se rechaza	Se rechaza	-
2021	Se rechaza	Se acepta	Se acepta	Se acepta	Se rechaza	Se rechaza	Se acepta	Se rechaza
2022	Se rechaza	Se rechaza	Se acepta	Se rechaza	Se acepta	Se acepta	Se rechaza	Se rechaza
2023	Se rechaza	Se rechaza	Se rechaza	Se rechaza	Se acepta	Se acepta	Se rechaza	Se rechaza
2024	Se rechaza	Se rechaza	Se rechaza	Se rechaza	Se acepta	Se rechaza	Se rechaza	Se rechaza
2025	Se rechaza	Se rechaza	Se acepta	Se acepta	Se rechaza	Se rechaza	Se rechaza	Se rechaza

8 Referencias

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Kish, Leslie. 2005. *Statistical design for research*. John Wiley & Sons.