



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

EVALUACIÓN DE EFECTIVIDAD DE CAMPAÑAS MULTICANAL EN UNA
COMPAÑÍA DE TELECOMUNICACIONES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

FELIPE IGNACIO AYALA MIRANDA

PROFESOR GUÍA:
MARCEL GOIC FIGUEROA

MIEMBROS DE LA COMISIÓN:
DANIEL SCHWARTZ PERLROTH
SEBASTIÁN RÍOS PÉREZ

Este trabajo ha sido financiado por CEINE

SANTIAGO DE CHILE
2016

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: FELIPE IGNACIO AYALA MIRANDA
FECHA: 2016
PROF. GUÍA: SR. MARCEL GOIC FIGUEROA

EVALUACIÓN DE EFECTIVIDAD DE CAMPAÑAS MULTICANAL EN UNA COMPAÑÍA DE TELECOMUNICACIONES

Las empresas, para lograr mantenerse como entes vigentes y dominadores en su determinado mercado, necesitan mantener una disposición permanente a la innovación y al intentar no solo optimizar sus procesos actuales, sino que buscar nuevas formas de hacer las cosas y romper los paradigmas en los que están inmersos.

De esa necesidad nace el presente trabajo, el cual busca implementar un estudio experimental generado con clientes reales, datos reales y con resultados observables de forma transparente para su análisis posterior por medio de la innovación en la aplicación de la estrategia comunicativa de parte de la empresa hacia los clientes. Este estudio se basa en los canales de comunicación, en como estos pueden ser usados en secuencia y en distintos órdenes y en como la reiteración en la comunicación de una oferta puede impactar sobre la propensión a la compra de esta por parte de un cliente.

Esto es relevante no solo desde la perspectiva de la empresa, al hallar aquellas combinaciones de canales de comunicación que generen un nivel más alto de ventas que las estrategias clásicas de comunicación, sino que también desde la perspectiva del cliente, el cual cada vez más rechaza los canales más tradicionales de comunicación (Call Center por ejemplo) en favor de canales más digitales y en los cuales es el cliente quien tiene un rol más activo en la interacción con la empresa, entregándole así una mejor experiencia a los clientes, aspecto fundamental para asegurar la sostenibilidad de la empresas en el largo plazo.

La metodología de implementación del diseño experimental en cuestión considera la asignación aleatoria de los individuos a observar sobre distintos tratamientos construidos en función de los canales disponibles y de hipótesis de estudio relevantes para la empresa de multicontactabilidad y multicanalidad. La aleatoriedad es fundamental, dado que así se logra minimizar el riesgo de sesgo en los resultados solamente por la selección de los prospectos. Por medio de análisis estadísticos se logra contrastar cuál de los tratamientos testeados es aquel que genera mejores resultados concretos en la cantidad de ventas generadas y en la efectividad obtenida.

A mis padres.

Agradecimientos

Primero y principalmente agradezco a mis padres por todas las oportunidades que me han entregado en la vida y por permitirme ser el mejor ser humano que puedo llegar a ser a través de su enseñanza, su rigor y su cariño.

A mi hermano por enseñarme a nunca transar mis valores y a ser el mejor hombre que puedo ser.

A mi abuela por enseñarme a siempre ser íntegro, cortés y darme la educación fundamental.

A mi primo por enseñarme a ser un profesional de excelencia y tenaz.

A mis amigos por los buenos momentos y experiencias vividas.

A los profesores Marcel, Daniel y Sebastián por permitirme trabajar con ellos y apoyarme un 110 % en todo lo que necesité durante el desarrollo de mi memoria.

A toda la gente del Ceine que me permitió hacer de la experiencia de la memoria algo muy grato.

A todos ellos, ¡Gracias totales!

Tabla de Contenido

| | |
|---|-----------|
| 1. Introducción | 1 |
| 1.1. Descripción general | 1 |
| 1.2. Descripción de la empresa | 3 |
| 1.3. Planteamiento del problema | 5 |
| 1.4. Objetivos | 7 |
| 1.4.1. Objetivo general | 7 |
| 1.4.2. Objetivos específicos | 7 |
| 1.5. Alcances | 7 |
| 1.5.1. Campañas históricas | 7 |
| 1.5.2. Diseño experimental | 7 |
| 1.6. Resultados esperados | 8 |
| 2. Marco Teórico | 9 |
| 2.1. Multicanalidad en Marketing | 9 |
| 2.2. Rol del canal | 11 |
| 2.3. Modelos de propension | 14 |
| 2.4. Test de hipótesis | 16 |
| 2.5. Diseño experimental | 17 |
| 2.6. Análisis de experimentos | 19 |
| 2.6.1. Análisis de poder estadístico | 20 |
| 2.6.2. Análisis de test de hipótesis | 21 |
| 3. Metodología | 22 |
| 3.1. Análisis exploratorio de campañas anteriores | 22 |
| 3.1.1. Análisis y selección de datos | 23 |
| 3.1.2. Tratamiento de <i>valores atípicos y valores faltantes</i> | 24 |
| 3.1.3. Análisis descriptivo | 24 |
| 3.1.4. Extracción de dinámicas | 25 |
| 3.1.5. Construcción de inputs adicionales | 26 |
| 3.1.6. Regresión logística | 27 |
| 3.1.7. Descripción de interacción variables | 28 |
| 3.2. Diseño experimental | 29 |
| 3.2.1. Selección de prospectos | 29 |
| 3.2.2. Asignación de condiciones y tamaño muestral | 30 |
| 3.2.3. Hipótesis a testear | 35 |
| 3.2.4. Análisis de resultados | 36 |

| | |
|--|-----------|
| 4. Resultados | 41 |
| 4.1. Análisis de campañas anteriores | 41 |
| 4.1.1. Análisis de efectividad | 41 |
| 4.1.2. Regresión logística | 44 |
| 4.2. Diseño experimental | 58 |
| 4.2.1. Efectividad Call | 58 |
| 4.2.2. Efectividad Email | 60 |
| 4.2.3. Efectividad Total | 61 |
| 4.2.4. Comparación tratamientos - Analisis post hoc de poder estadístico . . | 63 |
| 4.2.5. Comparación tratamientos - Test de hipótesis | 65 |
| | |
| Conclusión | 68 |
| | |
| Anexos | 71 |
| | |
| A. Descripción del mercado | 72 |
| | |
| B. Ejemplo utilizacion <i>G-Power</i> | 74 |
| | |
| C. Boxplots | 75 |
| | |
| D. Construcción binarias (código R) | 77 |
| | |
| Bibliografía | 79 |

Índice de Tablas

| | |
|--|----|
| 2.1. Nivel de invasividad - interactividad - accionabilidad por canal | 13 |
| 3.1. Ejemplo de tabla de contingencia | 25 |
| 3.2. Tratamientos testeados Enero - Multicontacto | 31 |
| 3.3. Tratamientos testeados Enero - Multicanal 2 contactos | 32 |
| 3.4. Tratamientos testeados Marzo - Multicontacto | 33 |
| 3.5. Tratamientos testeados Marzo - Multicanal 2 contactos | 33 |
| 3.6. Tratamientos testeados Marzo - Multicanal 3 contactos | 33 |
| 3.7. Análisis a priori de poder estadístico | 34 |
| 4.1. Efectividad campañas Outbound (Fuente: Data provista por la empresa) . . | 42 |
| 4.2. Contraste efectividad campañas CQC vs parque (Fuente: Data provista por la empresa) | 43 |
| 4.3. Modelo agregado (muestras balanceadas 50-50) | 45 |
| 4.4. Modelo agregado (muestras balanceadas 60-40) | 46 |
| 4.5. Modelo agregado (muestras balanceadas 70-30) | 47 |
| 4.6. Modelo agregado (muestras balanceadas 80-20) | 48 |
| 4.7. Modelo campaña 2DA LN PERS | 49 |
| 4.8. Modelo campaña Bolsas Min/SMS | 50 |
| 4.9. Modelo campaña Bolsas Nem Personas | 51 |
| 4.10. Modelo campaña Cross Full Bam Venta 4G | 52 |
| 4.11. Modelo campaña Up Nem Personas | 53 |
| 4.12. Modelo campaña Up Voz Móvil | 54 |
| 4.13. Efectividad tratamientos Call | 58 |
| 4.14. Efectividad tratamientos E-Mail | 60 |
| 4.15. Cantidades reportadas tratamientos E-Mail | 61 |
| 4.16. Efectividad total tratamientos | 62 |
| 4.17. Análisis post-hoc de poder estadístico | 64 |
| 4.18. Test de hipótesis para contraste de tratamientos | 66 |
| 4.19. Test de hipótesis para contraste de tratamientos específicos | 67 |
| 4.20. Test de hipótesis para contraste de tratamientos (agregado por cantidad de contactos) | 67 |
| A.1. Líneas fijas en servicio por año (Fuente: Subtel) | 73 |

Índice de Ilustraciones

| | |
|---|----|
| 1.1. Abonados móviles 30 días por empresa (Total: 23 683 351) (Fuente: Subtel) . | 3 |
| 1.2. Líneas fijas por empresa (Total: 3 416 916) (Fuente: Subtel) | 4 |
| 2.1. Ejemplo de modelo de integración VS LTV (Fuente: Paper número [12, pp. 99]) | 11 |
| 2.2. Escala de información-acción de canales | 12 |
| 2.3. Roles fundamentales canales | 13 |
| 2.4. Efecto mediador | 18 |
| 3.1. Representación gráfica del test post-hoc(muestras de 3.000 individuos) | 39 |
| 4.1. Efectividad histórica Call Center | 42 |
| 4.2. Contraste efectividad histórica Call Center CQC vs Parque | 44 |
| 4.3. Modelo campaña 2DA LN PERS | 49 |
| 4.4. Modelo campaña Bolsas Min/SMS | 50 |
| 4.5. Modelo campaña Bolsas Nem Personas | 51 |
| 4.6. Modelo campaña Cross Full Bam Venta 4G | 52 |
| 4.7. Modelo campaña Up Nem Personas | 53 |
| 4.8. Modelo campaña Up Voz Móvil | 54 |
| A.1. Abonados contrato 30 días por empresa (Total: 7 374 909) (Fuente: Subtel) . | 72 |
| A.2. Abonados prepago 30 días por empresa (Total: 16 308 442) (Fuente: Subtel) | 72 |
| A.3. Líneas fijas en servicio por año (Fuente: Subtel) | 73 |
| B.1. Interfaz software G-Power | 74 |
| C.1. Definición de cuartiles data | 75 |

Capítulo 1

Introducción

1.1. Descripción general

Las empresas, cualquiera sea su rubro o mercado, deben tener como principal misión la de lograr aumentar su valor y lograr ser mejores para los stakeholders relevantes. Para ello, el foco debe estar puesto fundamentalmente en el driver principal detrás de la generación de valor que pueden tener las compañías: el cliente. Esto es aún más cierto en la industria de los servicios, donde la percepción de valor de los clientes más relevante, debido a las relaciones de largo plazo que en general se establecen entre ambas partes. Aún más, las malas prácticas de gestión pueden ser muy caras para estos negocios, debido a lo difícil que es borrar las malas experiencias que un consumidor experimenta.

En Chile, la industria de las telecomunicaciones y particularmente los servicios móviles son un caso interesante, dado que es un mercado altamente concentrado, con 3 empresas que dominan el market share - Telefónica Movistar, Entel y Claro Chile - pero con varias otras comenzando a participar del mercado y que en el largo plazo pueden presentar una amenaza a los líderes. El constante cambio de las tecnologías, específicamente de las redes (internet móvil 3G a 4G, internet hogar ADSL a Fibra Óptica, etc.) y las modificaciones regulatorias recientes (portabilidad, fin de la larga distancia nacional, entre otros) hacen necesaria la adaptación frente a dichos cambios y lograr además ajustar la propuesta de valor para diferenciarse de la competencia en el mercado. En general, los clientes se muestran más seducidos por aquellas propuestas de valor que más se ajusten a sus necesidades específicas y que se hagan cargo de sus particularidades. El riesgo de no tomar esto en consideración radica en que los servicios se comienzan a convertir en meros commodities, convirtiendo al mercado en una carrera de reducción de costos y en consecuencia, de competencia en precios, dejando de lado el valor de la especificidad que se les puede entregar a los clientes.

La componente de esa especificidad que el presente trabajo de memoria de título explora tiene que ver con la manera en que se comunican las ofertas a los clientes, es decir, el canal que se utiliza para que dicha oferta se entregue de la mejor manera posible. En términos concretos, este proyecto se concentra en evaluar la administración de campañas multicanal, es decir, el entender como los distintos canales de marketing directo se pueden combinar a través de

distintas maneras para lograr entregar un mensaje a tipos específicos de clientes por medio de estrategias de comunicación y así determinar cuáles combinaciones de canales, órdenes de utilización de estos y distintas interacciones que se detecten logran mejores resultados de las campañas que se efectúan.

La administración de este tipo de campañas agrega una componente de complejidad a la gestión estratégica de la promoción. Si bien la administración disconexa e independiente de los distintos canales de comunicación con los que cuentan las empresas entrega la ventaja de que la administración general se simplifica (al asignar áreas y tareas independientes para la operación de cada canal), ello trae consigo la desventaja de la ineficiencia que presenta este enfoque. Se abre la posibilidad por ejemplo de situaciones en que se contacta a un mismo cliente por 2 canales separados, pero con solo el primer contacto se logró motivar al cliente al consumo - por lo que se hizo un gasto de capital innecesario al reiterar la comunicación. O bien, podrían darse el caso en que los esfuerzos se focalizan sencillamente en un grupo privilegiado de clientes (los que entregan más valor a la compañía), dejando de lado a clientes que potencialmente podrían ser muy valiosos para la compañía, pero que requieren de mayores estímulos antes de tomar una decisión, y por tanto, requieren una estrategia comunicativa más a medida.

Bajo este marco, el problema de negocios que le interesa solucionar a la compañía se puede resumir en tres puntos fundamentales:

- **Ganar eficiencia y economizar al mejorar elección de canal de contacto.** Ciertos canales poseen un costo de utilización muy alto - particularmente el Call Center, que si bien se encuentra externalizado, su operación es cara no solo en el uso de recursos económicos, sino también de recursos humanos - por lo que es necesario para la empresa identificar canales más económicos que logren transmitir las distintas ofertas hacia los clientes. Es decir, economizar la operación del marketing directo.
- **Mejorar eficiencia propia de las campañas.** Lograr que las tasas de conversión de las campañas que se efectúen sea más alta de lo que es actualmente, y en consecuencia, lograr una mejor rentabilidad por campaña, dados los costes particularmente altos de ciertos canales. La implementación de una estrategia conjunta de canales, considerando múltiples contactos por medio de distintos canales, busca ello al redistribuir la carga asignada de clientes a cada canal específico.
- **Lograr una mejora en la experiencia del cliente.** Al evitar un "bombardeo" sobre los clientes (excesivos contactos con un cliente particular) y al lograr entregarle una experiencia más personalizada en la forma en la que se le comunica una oferta particular - clientes podrían preferir enterarse de una oferta por medio de un e-mail, mientras que otros quizás por medio de un llamado telefónico.

Debido a que en la empresa no existe actualmente una coordinación entre las campañas generadas por medio de los distintos canales de marketing directo, y a que en general la utilización de los canales tienen distintos roles u objetivos (por lo que las campañas de los distintos canales no son perfectamente comparables), el desarrollo del presente trabajo busca por medio de la implementación de un diseño experimental, extraer resultados que permitan generar inferencias y conclusiones referido a las mejores formas de contactabilidad que se pueden impulsar hacia los clientes y de esa forma lograr sacar un mayor provecho de las

campañas de marketing directo que se generan. La aplicación de dicho diseño experimental busca combinar los distintos canales de marketing directo y explorar como a través de las posibles sinergias que se puedan encontrar, los clientes responden a las ofertas que se les presentan, levantando así data y resultados relevantes de la implementación de campañas bajo una nueva lógica de contactos.

Además, existe una necesidad de explorar con mayor fuerza nuevos canales debido a la operación actual del Call Center, la cual se desempeña bajo un fuerte régimen de incentivos por ventas, lo que en ocasiones lleva a que se produzca un reciclaje"de bases de datos, pasando a llevar la inteligencia que se genera en la extracción de prospectos y por ende

1.2. Descripción de la empresa

El mercado de las telecomunicaciones es uno que se encuentra altamente concentrado, particularmente en lo que respecta a los servicios de telefonía. Esto es claro al observar estadísticas de la subtel que muestran que del total de abonados móviles en Chile, prácticamente un 97 % se encuentran distribuidos en Claro (22.72 %), Entel (35.61 %) y Movistar (38.30 %). Esto incluyendo tanto líneas de contrato como de prepago.

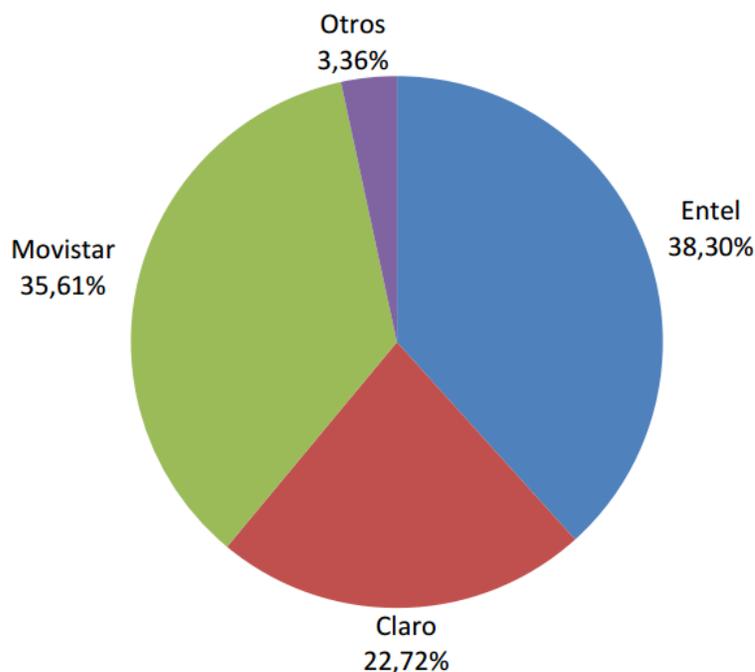


Figura 1.1: Abonados móviles 30 días por empresa (Total: 23 683 351) (Fuente: Subtel)

En el mundo fijo la situación es similar, manteniéndose un claro liderazgo de parte de Telefónica Movistar (46.3 %) respecto a la cantidad de líneas fijas pero con VTR (20.7 %) como competidor potente y Entel (13.1 %) manteniendo un tercer lugar en el mercado 1.2.

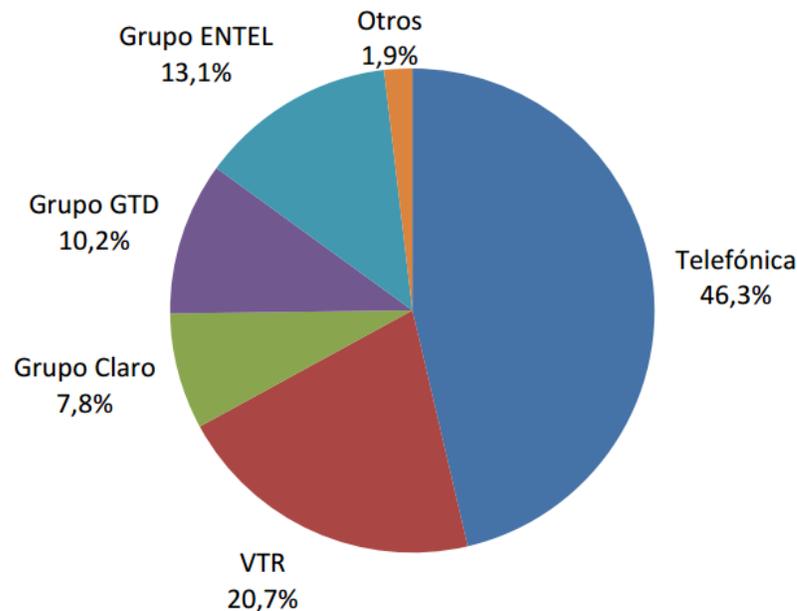


Figura 1.2: Líneas fijas por empresa (Total: 3 416 916) (Fuente: Subtel)

Los órdenes de magnitud de los datos antes presentados muestran que el mercado de los servicios móviles presenta mayores oportunidades de generación de valor a través del upselling de servicios, por ejemplo a través del traspaso de abonados prepago a abonados a contrato, estableciendo una relación más sólida y consistente con los clientes, generándose flujos de ingresos más constantes.

En relación al desempeño particular de los esfuerzos de marketing de la empresa en análisis de la presente memoria, la administración se lleva adelante por medio de una dinámica de campañas, planificadas generalmente en forma mensual (y en ocasiones particulares, en forma quincenal). A grandes rasgos, el proceso comienza por la definición de la oferta, la cual involucra decisiones definidas desde la planificación más estratégica, determinada por ejemplo en función de objetivos comerciales de largo plazo (atender a un grupo particular de clientes, enfocarse en un grupo etario específico). Posterior a ello, se determina por medio de la investigación analítica, qué clientes se podrán prospectar o seleccionar como candidatos para campañas particulares, pareando de esa manera una oferta con un cliente específico. En general, se asigna un cliente a solo una campaña, evitando así un agobio sobre el cliente por ser candidato a más de una campaña y acotando en consecuencia el número máximo de asignación a campaña a una por cliente.

Los canales utilizados en esta instancia son principalmente los de marketing remoto clásicos, como lo es el Call Center y el IVR, el correo electrónico, los mensajes de texto y multimedia principalmente. Otros canales como el Sat Push - mensajes multimedia inmediatamente desplegados tras la recepción al estilo de un banner - se están introduciendo desde hace algunos meses, pero su utilización aún se encuentra en estado de prueba. No obstante lo anterior, en general el canal que más utilización tiene en la administración de las campañas es el Call Center, mientras que los demás canales más que nada se utilizan como canales

informativos o de fidelización (a través de las ofertas del club de beneficios con que cuenta la empresa por ejemplo). El desafío se encuentra fundamentalmente en lograr dilucidar como se puede lograr un apoyo entre los canales disponibles y generar una asignación de campañas más inteligentes.

1.3. Planteamiento del problema

Las campañas con las que cuenta la compañía se concentran fundamentalmente a través de llamadas por medio del canal Call Center, buscando transmitir hacia los clientes una oferta de una forma directa e interactiva. Este canal presenta la característica de ser relativamente invasivo, es decir, del rol importante que juega no solo el cliente en una comunicación sino que también el ejecutivo que representa a la empresa y que busca seducir al cliente a través de este canal. Además, existe el incentivo para concretar ventas de cualquier manera posible, por lo que se genera un trato a veces muy insistente de parte del ejecutivo por cerrar la transacción. Paralelamente, este canal al ser más directo, permite un intercambio de información muy superior a los demás, y de manera inmediata, lo que presenta ventajas importantes en ese sentido.

Canales como el E-Mail, si bien sigue siendo uno muy directo - al redirigir en forma inmediata hacia la sucursal on line o a un call inbound por medio del texto y los enlaces presente en el mensaje mismo -, presenta restricciones propias de espacio máximo que se puede, o reglas de negocio que lo pueden limitar (tamaño del mensaje, esquema de colores, etc.), pero al ser mucho menos invasivo que el Call, puede ser más atractivo como medio de información para ciertos clientes. El foco del proyecto está fundamentalmente en lograr extraer inteligencia respecto a la respuesta de los clientes al ser enfrentados a ofertas comunicadas por medio de distintos canales a través del testeo diferentes combinaciones de contactabilidad con los clientes de la compañía y así verificar si existe alguna de ellas que logre ser más efectiva que otra, al tomar en consideración las particularidades de los canales disponibles.

Desde el punto de vista del negocio, es interesante evaluar si alguna combinación es más efectiva en ventas que otra, es decir, si las ventas sobre los contactos generados, es más alta que las demás. También verificar si dichas combinaciones son más efectivas en términos de contactabilidad, vale decir, si los clientes presentan una tasa de apertura de emails mayor, un recibimiento de mensajes de texto mayor y una respuesta a llamadas superior. Por otro lado, es interesante lograr argumentar respecto a la viabilidad de reducir la utilización de canales más caros, es decir, utilizar otros canales remotos más económicos para lograr las mismas ventas. En este caso particular, el Call Center es el canal más caro de utilizar, debido no solo a la cantidad de recursos necesarios para su operación sino que además a que no se gestiona completamente en forma interna, sino que se externaliza a empresas que administran estos servicios, tanto en Chile como en otros países. Esto provoca que no se tenga control completo sobre las decisiones que se toman al momento de generar ofertas hacia los clientes y que clientes se contactan, lo que de alguna manera, motiva la utilización de canales más controlables.

A través de aplicación de primeras regresiones, utilizando técnicas de selección de variables

que sean relevantes respecto a la variable de decisión que se está considerando, es posible analizar canal a canal el rendimiento de cada uno, y entender además como distintas variables demográficas y transaccionales explican el comportamiento actual de los clientes frente a las distintas ofertas a las que son expuestos. La principal técnica a utilizar en el proceso de selección de variables corresponde al stepwise, debido lo exhaustivo en la búsqueda de variables que es y debido a que su implementación es conceptualmente simple de aplicar y de entender. Dado que actualmente en la compañía, solo se utiliza como principal vía de comunicación directa con el cliente el canal Call Center, solo con ese canal se podrá medir perfectamente la efectividad de las ventas que se generan. Esto debido a que para canales como el SMS y el E-Mail la generación de la transacción final no se genera en el mismo canal, sino que se redirecciona hacia otro canal (un Call receptor de llamadas en general), por lo que no es tan directa la trazabilidad de las transacciones motivadas por efecto de la acción directa sobre el cliente o bien por una motivación natural y proactiva de parte del cliente. Para estos canales, existen vías alternativas de medición, que se considerarán hasta cierto punto (aperturas de correos y entregas de mensajes), pero no se está considerando invertir mucho tiempo en esto, debido a que los insights relevantes que se desean obtener se desprenden de los diferentes niveles de ventas generados a través de las pruebas que se implementarán.

Los principales insights que se esperan encontrar son 2:

- Qué tasas de respuesta generan las distintas combinaciones aplicadas.
- Identificar si el orden de utilización de los canales tiene relevancia sobre los resultados de campañas.
- ¿Contactos reiterados son más efectivos que contactos únicos?
- Qué características de los clientes influyen de alguna forma la decisión y como esto se relaciona con alguna combinación de canales implementada.

La bibliografía de la línea de investigación del marketing directo menciona que ciertas respuestas de los clientes son esperables: responden de mejor manera a ciertos canales dependiendo de las características propias del cliente (personas de cierto rango etario desconfía de los canales remotos), responden de mejor manera mientras más contactados sean (más frecuentemente) pero solo hasta cierto punto (el beneficio marginal por contacto adicional es negativo cuando la cantidad de contactos tiende a infinito)[8, pp. 124–126].

La aplicación de un diseño experimental ad hoc a esto será la base sobre la que se espera obtener estos hallazgos, responder las preguntas antes planteadas, y validar o contradecir los puntos antes mencionados. Para no intervenir en la operación de la asignación de campañas actual, estas pruebas se generarán a nivel de piloto, considerando muestras de tamaño razonable que permitan generar inferencias y cuyos resultados posean significancia estadística para argumentar de manera clara respecto a uno u otro resultado y las diferencias que se obtengan de ellos.

Lo relevante del diseño se encuentra en la definición del rol de un canal particular, de cómo se le puede sacar mayor provecho a cada canal dentro de una comunicación particular con un cliente al ser usado en distintos órdenes y con una coherencia en cada instancia en que se genere un contacto con el cliente.

1.4. Objetivos

1.4.1. Objetivo general

Estimar efectividad de campañas multicanal considerando las preferencias específicas de cada cliente y la dinámica de múltiples contactos en una compañía de telecomunicaciones.

1.4.2. Objetivos específicos

- Identificar variables demográficas y transaccionales relevantes en la estimación de la probabilidad de éxito de un canal o secuencia de canales.
- Identificar y describir de manera exploratoria el desempeño de los canales usados en la compañía para la gestión de campañas.
- Encontrar insights de comportamiento de los clientes en un entorno multicanal por medio de un estudio experimental. (Respuesta a contactos previos, interacción de canales, roles de estos)
- Determinar la probabilidad de conversión a nivel de clientes para diferentes estrategias de contactos multicanal.

1.5. Alcances

1.5.1. Campañas históricas

Para la estimación del modelo de elección discreta, como así también para los análisis descriptivos sobre las campañas mismas, se considerará la data disponible de las campañas ejecutadas en el año 2014 entre los meses de Enero y Diciembre, por lo que se trabajará con 12 meses de data.

1.5.2. Diseño experimental

Canales

Para lograr los análisis de propensión antes mencionados, se consideraran los siguientes canales:

- SMS
- E-Mail
- Call Center

También existen otros canales remotos que se usan en la empresa, tales como SAT Push - mensajes multimedia que se despliegan al recibirlos el cliente - o mailing - insertos en boletas -, pero estos se descartarán a priori dado que aún no se encuentran implementados del todo en el parque de clientes (particularmente SAT Push), al alto costo que tienen para implementar (insertos en boleta) y al control más claro que se tiene por sobre los demás canales.

Por otro lado, se descartarán también otros canales de comunicación directa, como la tienda física o el canal sitio web, dado que en esos casos existen variables que son poco controlables para la evaluación e implementación de experimentos, y dado que en esos casos, el contacto es principalmente iniciado por el cliente.

Campañas

Debido a que la implementación de los modelos que se desean aplicar es intensiva en el requerimiento de datos, es importante considerar en el estudio solo a aquellas campañas sobre las cuales se tiene información suficiente de los clientes y así explicar por qué los resultados posiblemente difieren entre distintos tipos de clientes.

Por simplicidad se decidió acotar el estudio a una campaña particular de la compañía, la cual se aplicará por medio de distintas condiciones experimentales a considerar. Esto quiere decir que la oferta se mantendrá constante a lo largo de los distintos tratamientos en testeo.

1.6. Resultados esperados

- Una evaluación de la efectividad de cada canal por tipo de campaña a nivel de cliente (o grupo de clientes) por medio de un ranking que indique qué canal gatilla con mayor fuerza la propensión al consumo.
- Implementar un modelo experimental para la evaluación de contactos multicanal en secuencias temporales.
- Evaluación de los experimentos documentando las diferencias de efectividad de cada condición a través de test de hipótesis univariados.
- Un modelo que permita evaluar la probabilidad de conversión para distintas estrategias de contacto multicanal.
- Basado en la ejecución del modelo, proponer mejores prácticas de contactos multicanal (canales a activar, en que momento, y hacia clientes de qué características) para maximizar la propensión al consumo de cada cliente.
- Un procedimiento para la evaluación de futuras campañas multicontacto y la actualización de los parámetros iniciales del modelo en función de ello.

Capítulo 2

Marco Teórico

2.1. Multicanalidad en Marketing

Como señalan Neslin, Grewal et al (2006), la administración de clientes multicanal se entiende como el diseño, gestión, implementación, coordinación y evaluación de canales a través de los cuales las firmas interactúan con los clientes, con el objetivo de mejorar el valor del cliente por medio de la adquisición efectiva de estos [12, pp. 97]. Un término fundamental en dicha definición es la coordinación; la administración de una estrategia coordinada y que tenga sentido de cara al cliente es probablemente la dimensión más relevante en la construcción de estrategias multicanal, dado que existe el peligro latente de agobiar al cliente con comunicaciones diferentes entregadas por medio de distintos canales, o reiterar ofertas sin lograr una coherencia entre los distintos contactos, dejando en evidencia una clara desconexión entre las decisiones tomadas desde las áreas comerciales y de marketing con las necesidades y motivaciones particulares de los clientes.

No obstante, existen varios otros aspectos a considerar al momento de construir estrategias multicanal, como por ejemplo la evaluación individual de cada canal. La respuesta a la pregunta ¿cómo se debe considerar el desempeño individual de cada canal para hacerlo comparable con los demás? [1, pp. 9] no es algo fácil de responder, dado que hay factores importantes a considerar, tales como la atribución al aporte que los equipos administradores de marketing le dan a cada canal como generadores de ventas. Otro aspecto muy relevante a considerar es la asignación de recursos a cada canal, lo que de manera directa impacta sobre la efectividad de operación de cada uno de ellos. Dado que los recursos son limitados los beneficios marginales de un contacto por cada canal deben estimarse con claridad suficiente para que las asignaciones presupuestales puedan justificarse al momento de la evaluación ex post del desempeño de las campañas.

La aplicación de estrategias multicanal no se explica solo debido a los avances tecnológicos en las telecomunicaciones, que permiten cada vez más puntos de contacto con los clientes, sino que además como los comportamientos de estos ha ido modificándose en función de esos avances. El auge de los teléfonos celulares como herramientas del día a día es un fenómeno particular de esto, permitiéndole a los usuarios hoy en día ya no solo hablar a través de

una red de voz, sino que la navegación intensiva en datos, el envío de mensajería inmediata desplazando totalmente la necesidad de comunicaciones por medio de fax por ejemplo, y reemplazando una serie de dispositivos en el camino (relojes, calculadoras, despertadores e inclusive en muchos casos, computadores personales). Pero va más allá; las nuevas tecnologías de acceso web (fibra óptica) y los bajos costos de los computadores entregan esa vía adicional de contacto con los clientes. El desafío está en lograr generar estrategias inteligentes a través de este abanico de posibilidades, dando cuenta de las experiencias cada vez más personalizadas que requieren los clientes.

De la mano de lo anterior, la utilización de distintos canales no es útil sencillamente como medios de venta de productos o servicios, sino además como medios de retención de clientes por medio de la experiencia que se le entrega a estos. La implementación de una sucursal virtual por medio de un sitio web o una aplicación de Smartphone por ejemplo hacen sentido desde la perspectiva del cliente como una forma de satisfacer sus necesidades y requerimientos de post-venta sin necesidad de asistir a una sucursal física para hacerlo. Estudios además muestran que la aplicación e implementación de comunicaciones por medio de distintos canales influye sobre el lifetime value (LTV) de los clientes [16, pp. 112–114]. Esto justifica la inversión en estas estrategias, habilitando cada vez más vías de comunicación entre la empresa y sus abonados, lo que de alguna manera impacta no solo en la experiencia de los clientes, sino en la capacidad de retención de los mismos por parte de las compañías, aspecto importantísimo en el mercado de las telecomunicaciones producto no solo de la competencia, sino del entorno supeditado a las regulaciones del mercado mencionadas en secciones anteriores, que ha generado (en el caso de Chile al menos) una reducción en las líneas de negocio en las cuales las Telco son partícipes. Siendo riguroso, los autores Nelsin (2006) muestra por medio de un modelo matemático que a medida que los canales se van integrando (generando comunicaciones más coherentes y conexas entre sí), el lifetime value se incrementa; pero, al aumentar esa integración, hay un aumento de costos, por lo que se debería lograr encontrar un punto en el cual se maximiza el LTV a cierto nivel de integración, siendo ese punto el escenario ideal para la compañía en términos económicos.

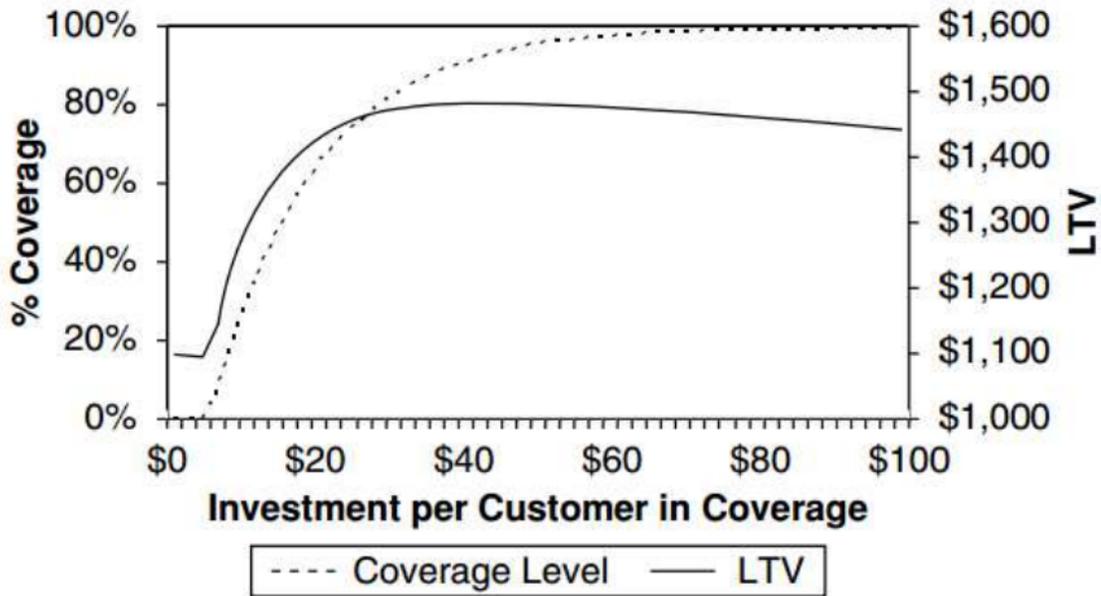


Figura 2.1: Ejemplo de modelo de integración VS LTV (Fuente: Paper número [12, pp. 99])

Los consumidores además muestran un comportamiento más atractivo desde la perspectiva de las empresas cuando se están inmersos en un ambiente multicanal. Esto debido a las compras cruzadas que se pueden generar potencialmente, entendiéndose este concepto como la participación de parte de los clientes en transacciones correspondientes a distintas categorías de productos [10, pp. 102]. Es muy probable por ejemplo que ciertos clientes no muy instruidos tecnológicamente, prefieran asistir directamente a una sucursal para recibir asesoramiento respecto a qué dispositivo y plan móvil es más ad hoc a sus necesidades, verificando in situ aspectos como el tamaño de la pantalla, velocidad, compatibilidad de aplicaciones, etc; mientras que para un servicio de upselling o la contratación de un producto de la categoría fija (banda ancha, TV), un contacto por medio remoto como una llamada o una atención web es quizás más cómodo tanto para el consumidor como para la empresa, y esa apertura de posibilidades es la que al final, permite la máxima generación de valor por medio de un cliente específico, debido a que se logra generar una comunicación más integral con los clientes para satisfacer sus requerimientos individuales, y poder responder de manera cómoda frente a nuevas solicitudes de parte de los clientes, generando una entrega de servicios integral de cara al cliente.

2.2. Rol del canal

Una definición importante en un proceso comunicativo es la identificación de cual es la forma en la cual se le puede sacar el mayor provecho a los canales individuales, debido a las características y atribuciones de cada uno de ellos. Se pueden identificar principalmente, 3 características que definen el rol de un canal en una comunicación:

- Su grado de **invasividad**.

- Su grado **interactividad**.
- Su grado **accionabilidad**.

El Call Center es un canal altamente invasivo, interactivo y accionable dado que requiere de una complicidad directa de parte de los receptores de la comunicación para alcanzar la eficiencia necesaria del canal. A su vez, las interacciones que se producen por medio de este canal, permiten la transferencia no solo de información, sino que de la ejecución de acciones comerciales directas, permitiéndose la contratación de servicios de manera rápida, replicando en parte la experiencia que se puede obtener a través de la atención en sucursal.

El SMS se encuentra en otro extremo del espectro, en el cual las oportunidades de acciones comerciales prácticamente son inexistentes - al menos de manera directa - sino que las restricciones propias del canal (cantidad de caracteres soportados en cada comunicación) permiten principalmente la entrega de información de manera unidireccional. Considerando lo anterior, se puede proponer un marco conceptual respecto a la accionabilidad de los canales, una escala de Información-Acción en el cual se puede posicionar cada canal de marketing directo a disposición las empresas como se visualiza en la figura a continuación.



Figura 2.2: Escala de información-acción de canales

Esto considera que los canales de marketing directo, por un lado tienen distintos niveles de invasividad sobre los clientes, al requerir para su operación, más atención de parte del cliente e inclusive una interacción inmediata (como es en el caso del Call Center). La problemática que esto trae, es que muchas veces uno de los pecados en los que se incurre es en el contactar a los clientes en aquellos momentos en los cuales aún no están en un *mindset* preparado para aceptar alguna oferta que se le esté haciendo, por más conveniente que esta sea, sino que se cierran debido a muchos factores posibles: desinterés en general por las ofertas que reciben, atención a otros temas del momento (laborales o de otra clase, etc.), razón por la cual, la utilización de canales muy invasivos puede ser una mala estrategia como punto de entrada.

Del mismo modo, los canales se pueden distinguir mediante un rol de información - acción descrito en la figura anterior. Si bien canales de texto como el SMS y el E-Mail son herramientas que entregan un mensaje rápido y en forma concisa, ellos no permiten una concreción inmediata ni generar una acción tendiente a la contratación de servicios, al menos en forma automática. El E-Mail en este apartado supera al SMS, puesto que da acceso inmediato a enlaces que pueden ayudar al cliente no solo a buscar más información respecto de las ofertas que se le presentan, sino que también a la adquisición del producto ofrecido, pero muchas veces puede ser un medio poco efectivo, debido a la gran cantidad de correo Spam al que son

expuestos los consumidores.

El SMS por su parte, si bien no permite un paso adicional inmediato de comunicación con la empresa, es un mecanismo más invasivo, dado que el cliente lo recibe en forma inmediata y su visualización es más rápida. Desde la perspectiva del cliente, es un canal más pasivo, dado que el cliente sencillamente lo recibe y en función de ello decide si le interesa o no la oferta, pero el paso posterior depende mucho más del cliente mismo.

Una pregunta importante que este marco busca plantear es: ¿Cuál es el rol de los canales? [1, pp. 5-9] Se pueden identificar principalmente 3 roles relevantes:

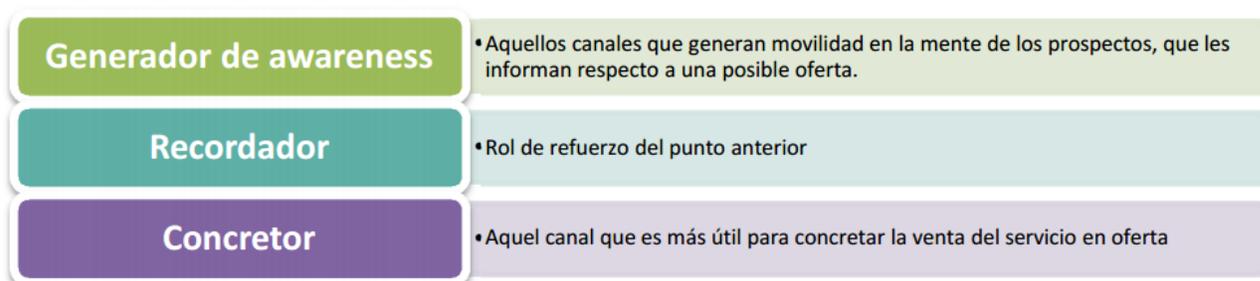


Figura 2.3: Roles fundamentales canales

Claramente de acuerdo a lo antes mencionado, ciertos canales son más integrales, como por ejemplo el Call Center, el cual entrega inmediatamente una ventana para que el cliente acepte o no un servicio. Otros funcionan naturalmente como generadores de awareness, como el SMS o incluso el E-Mail, presentando en la mente de los clientes esa idea, esa "semilla" de una posible oferta.

Aun así, lo anterior no obsta a que por ejemplo ciertos canales puedan asumir los 3 roles antes descritos (particularmente el Call Center opera de buena manera en los 3 roles descritos). El punto relevante está en identificar qué rol es mejor para un canal particular y como ello se puede utilizar para la implementación de estrategias multicanal que hagan sentido y que generen la mayor contactabilidad y ventas posibles con los recursos a disposición. Con todo lo anterior los 3 canales evaluados presentan la siguiente evaluación a priori:

| Canal | Invasividad | Interactividad | Accionabilidad |
|-------------|-------------|----------------|----------------|
| Call Center | Alta | Alta | Alta |
| E-Mail | Baja | Media | Media |
| SMS | Media | Baja | Baja |

Tabla 2.1: Nivel de invasividad - interactividad - accionabilidad por canal

2.3. Modelos de propension

Las regresiones son herramientas matemáticas bastante útiles, dado que a partir de estimaciones basadas en situaciones y comportamientos históricos, permite medir por un lado interacciones entre distintos atributos, y por otro lado, anticipar ocurrencias futuras sujeto al conocimiento ex ante de algún valor de una variable particular, es decir, permite posicionarse en escenarios hipotéticos, lo que es algo particularmente interesante, dado que permiten generar predicciones respecto a una variable de interés.

Los modelos más sencillos corresponden a las regresiones lineales, es decir, aquellos en los que existe una relación lineal entre una variable de interés: la variable dependiente (DV), y las variables que en alguna medida tienen influencia en el comportamiento de la variable dependiente: variables independientes (IV). Los modelos más sencillos asumen que la media poblacional ($E(Y)$) de la DV cambia a tasa constante y en forma lineal a medida que la IV aumenta o disminuye. Las observaciones de la DV se asumen como una ejecución aleatoria donde la media de dicha variable aleatoria corresponde a la media poblacional (desconocida).

$$Y = \beta_0 + \beta_1 X$$

Los modelos de elección discreta, los que corresponden a situaciones en la cual los individuos se exponen a un set definido y discreto de alternativas posibles entre los cuales debe efectuar una decisión. El set de alternativas debe tener ciertas características fundamentales para que el modelamiento sea correcto de acuerdo a Kenneth Train [15, pp. 15–17]:

1. El set de alternativas debe ser **mutuamente exclusivo**, lo que significa que cada alternativa posible para el cliente no se cruza con otra alternativa, es decir, elegir una opción implica no elegir ninguna otra alternativa.
2. El set de alternativas debe ser **exhaustivo**, todas las alternativas posibles deben ser consideradas, es decir, el cliente necesariamente debe hacer una elección. En este caso particular, la opción *no comprar* debe ser incorporada para considerar el hecho de que los clientes pueden tomar la decisión de no aceptar la oferta comercial a la que se ven expuestos.
3. Las alternativas deben ser **finitas**.

De acuerdo a cada caso particular, se pueden tomar ciertas consideraciones para la definición de las alternativas de manera que sean mutuamente exclusivas y exhaustivas y de esa manera asegurar que el modelamiento se acomode a las restricciones consideradas. La tercera condición en general cuando se habla de productos y servicios se da, por lo que no existen problemas en su cumplimiento para un modelo de esta clase.

Estos modelos operan bajo la asunción de que los individuos se comportan de forma racional, de acuerdo a la maximización de utilidades, por lo que los clientes deberían siempre seleccionar aquellas opciones que más beneficios le traigan a su utilidad. Si bien este es un supuesto fuerte, la formulación sugiere una separación de la utilidad en 2 elementos distinguibles (utilidad observable y utilidad no observable) que de alguna manera logra explicar

de manera clara como opera el raciocinio de los individuos al tomar decisiones. Así, la probabilidad de que un individuo n seleccione una alternativa i sobre una j dependerá de las utilidades que el cliente perciba de esas opciones:

$$\begin{aligned} P_{ni} &= \text{Prob}(U_{ni} > U_{nj} \quad \forall j \neq i) \\ &= \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}) \end{aligned}$$

Como se mencionó antes, si bien pueden existir muchos factores disponibles para el análisis que se pueden observar de los clientes que en parte pueden explicar qué utilidad le genera a un individuo particular una alternativa específica (V_{ni}, V_{nj}) , existen otros factores externos no observables que también pueden estar influenciando de alguna manera la utilidad que le genera a un cliente una alternativa $(\varepsilon_{ni}, \varepsilon_{nj})$. Dado que este último aspecto se desconoce, se asume que se comporta de manera aleatoria de acuerdo a una determinada distribución de probabilidades. La fórmula anterior, caracteriza una distribución acumulada, la probabilidad de que la diferencia entre las partes aleatorias de las utilidades sea menor a la diferencia entre las partes determinísticas de las utilidades. La definición general entonces de esta clase de modelos de elección se puede caracterizar de acuerdo a lo siguiente:

$$\begin{aligned} P_{ni} &= \text{Prob}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} + V_{nj}) \\ &= \int_{\varepsilon} \text{Indicatriz}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} + V_{nj}) f(\varepsilon_n) d\varepsilon_n \end{aligned}$$

Uno de los modelos más sencillos de implementar e interpretar es la regresión logística, la cual corresponde a un caso particular de la formulación anterior, asumiendo que los términos de error ε_{nj} distribuyen valor extremo tipo I. esta distribución se puede describir de acuerdo a lo siguiente:

$$\begin{aligned} f(\varepsilon_{nj}) &= e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \\ F(\varepsilon_{nj}) &= e^{-e^{-\varepsilon_{nj}}} \end{aligned}$$

La fórmula anterior de la probabilidad de que un individuo seleccione una alternativa i para un término de error dado, se puede expresar de la siguiente manera al reordenar términos:

$$P_{ni} | \varepsilon_{ni} = \prod_{j \neq i} e^{-e^{-(\varepsilon_{ni} + V_{ni} - V_{nj})}}$$

Al generalizar para cualquier ε_{ni} , se obtiene:

$$P_{ni} = \int \left(\prod_{j \neq i} e^{-e^{-(\varepsilon_{ni} + V_{ni} - V_{nj})}} \right) e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}} d\varepsilon_{ni}$$

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$

Que corresponde a la formulación clásica de este modelo. Se puede interpretar como una suerte de "casos favorable dividido por casos totales", donde en el numerador se encuentra la alternativa i , correspondiente a la probabilidad que interesa obtener, y en el denominador, todo el set de alternativas disponibles. La parte determinística de las utilidades que genera cierta alternativa en general se describe como una combinación lineal de las variables, en donde se estiman parámetros que ponderan cada variable:

$$V_{ni} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k = \beta' X$$

Los parámetros del modelo se logran estimar a través de la maximización de la función de verosimilitud. La función de verosimilitud expresa los valores de los parámetros en función de los valores fijos de la variable respuesta, en oposición a la probabilidad P_{ni} la cual expresa la probabilidad de ocurrencia de la variable dependiente sujeto a parámetros fijos [15, pp. 19–22].

$$L(\beta|y) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{(y_i-1)}$$

$$\ln(L(\beta|y)) = \sum_{\{y_i=1\}} \ln(\pi_i) + \sum_{\{y_i=0\}} \ln(1 - \pi_i)$$

Los valores de β que maximizan esta función anterior son los que se asumen como las estimaciones de los efectos que cada variable tiene sobre la variable dependiente correspondiente.

En estas últimas expresiones se expresa la no linealidad antes descrita, dado que la variable dependiente ya no toma un valor real, sino que binario (o discreto de acuerdo al número de posibles respuestas) dependiendo de las respuestas de los individuos. Este modelo antes descrito, se puede categorizar como uno "intrínsecamente lineal", dado que a partir de ciertas transformaciones, se podría lograr recuperar un modelo lineal (la función de utilidad en este caso), particularmente la definición de probabilidad de elección, la cual se construye a partir de funciones lineales (V_{ni}).

2.4. Test de hipótesis

Los test de hipótesis son herramientas que en general son muy fáciles de aplicar y concluir en función de los resultados obtenidos, lo que los convierte en algo que se debe tener siempre a la mano al momento de generar aplicaciones estadísticas y especialmente, de evaluarlas.

Aún más, son un grupo de técnicas clave para la inferencia estadística, entregando un procedimiento de evaluación de resultados, dado que permiten tener control sobre la probabilidad de decidir en forma incorrecta respecto a cierto output.

Existen variados tests, cada uno con un objetivo específico y con restricciones de aplicación propias de cada problema en donde se requieran. Por tanto, es fundamental el tener claridad respecto a qué test es ad hoc a cada situación particular, puesto que de lo contrario, las conclusiones que se obtengan pueden ser falsas, invalidando cualquier aseveración que se haga.

Una clasificación principal que se le puede dar a los distintos tests de hipótesis existentes es:

- **Test a una muestra:** donde se compara una muestra versus un parámetro conocido ex ante, vale decir, se busca validar un comportamiento basado en suposiciones teóricas respecto a un grupo de observaciones. Un ejemplo específico de esto es el test de significancia individual de parámetros en un modelo predictivo
- **Test a dos muestras:** para la comparación de dos muestras disjuntas y la verificación de su similaridad o diferencias entre ellas.

Este último tipo de test es particularmente necesario al momento de comparar resultados de aplicaciones experimentales, dado que en general, dichos resultados se evalúan en forma pareada, es decir, un grupo tratado versus un grupo de control (ausente de cualquier intervención adicional) o bien 2 grupos tratados entre sí, y así lograr concluir respecto a qué modificación sobre cierto sistema genera una diferencia sobre la variable dependiente en estudio. Básicamente, el test logra argumentar estadísticamente si un resultado es diferente a otro, lo que le da a las conclusiones un sustento fuerte identificando una tendencia en el comportamiento de las observaciones.

2.5. Diseño experimental

En determinadas situaciones de un proceso investigativo o de aplicación de metodologías sobre un problema real, no es suficiente el consultar la bibliografía a disposición en la academia, dado que la línea investigativa en estudio no se encuentra explorada a cabalidad, es decir, no se han generado resultados que logren concluir respecto a una interrogante específica de interés para algún mercado, negocio, etc, por lo que no es posible tomar resultados previos y transferirlos a nuevas aplicaciones. La construcción y ejecución de un diseño experimental [7, pp. 1–5, 133–135] en estos casos es particularmente útil y necesario, dado que permite de alguna manera generar acercamientos hacia la respuesta de estas interrogantes y en definitiva aportar sobre la línea investigativa en cuestión. Más aún, la relevancia fundamental de la aplicación de un experimento radica en la causalidad que se puede visualizar al implementar una modificación controlada sobre un sistema y así medir si al aplicar ese cambio sobre el sistema, se produce un efecto observable y cuantificable.

Los diseños experimentales no son únicos, sino que es posible a través de muchas maneras

buscar obtener un resultado particular, es decir, por medio de distintos diseños, selecciones de muestras o asignaciones de tratamientos. Es parte importante del trabajo de construcción, el lograr dilucidar el diseño ad hoc al problema en cuestión de manera que los resultados logren responder las interrogantes y los objetivos planteados ex ante a la implementación del diseño. A nivel agregado, se pueden diferenciar 2 clases de diseños experimentales: **estudio experimental controlado**, en el cual las intervenciones y las asignaciones de las mismas están definidas por el investigador de manera clara y definida; y por otro lado, **estudios observacionales**, en los cuales el diseño propone sencillamente los parámetros a observar de una determinada población, sin que exista una intervención directa de parte del experimentador.

Por otro lado, la detección de factores mediadores y confounders [11, pp. 175] en la definición del diseño experimental es un punto importante, dado que la ignorancia de estas posibles contingencias puede resultar obviamente en malas interpretaciones de los resultados obtenidos como consecuencia de los distintos tratamientos implementados, invalidando en parte las conclusiones que se desprenden de la ejecución del diseño.

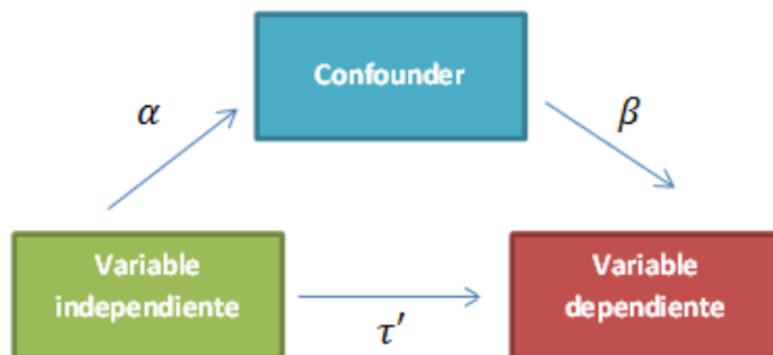


Figura 2.4: Efecto mediador

Básicamente, la relación existente entre variable dependiente e independiente con la existencia de un mediador queda descrito como se muestra en el esquema anterior, existiendo dos efectos interviniendo: aquel generado en forma directa por la interacción entre las variables - τ' - y aquel generado por la existencia de este tercer factor que interviene de manera no directa en el efecto de las variables independientes sobre la variable objetivo - $\alpha\beta$.

En contraste, el confounder queda descrito como una tercera variable relacionada a dos factores que acentúa o reduce la relación entre estos. La reducción del efecto del confounder permite detectar de manera directa entre variables independiente y dependientes.

$$\textit{Efecto total} = \alpha\beta + \tau'$$

Es importante a su vez, reconocer que los diseños experimentales presentan cierta imperfección que es en realidad imposible de evitar y tiene que ver con los factores externos que

no son parte del diseño. Esto es el error experimental, el cual se atribuye en general a todos aquellos factores que no son parte del estudio, y que pueden generar ciertos impactos sobre los resultados de la variable objetivo de interés, que no son detectables a priori (dado que no se incluyen en el diseño). También, este error se puede deber a malas definiciones del propio diseño, por lo que de igual forma, existe una componente interna que se debe controlar y en consecuencia, la construcción y definición de un buen diseño puede aportar a que ese error que se genera se vea reducido al mínimo posible, por medio de buenos sampleos, buena definición de tratamientos y condiciones a las que son expuestos los individuos, etc.

Finalmente, un problema fundamental que se busca solucionar por medio de un diseño experimental tiene que ver con las relaciones causales que se pueden desprender de las distintas variables en observación. No es muy complejo como primera mirada lograr observar las correlaciones entre pares de variables sencillamente al examinarlas de manera cualitativa, pero el lograr concluir respecto a la causalidad existente (qué variable influye sobre la otra de manera directa), es más complejo y no se logra solamente a través de un análisis exploratorio. Como define el análisis clásico de John Stuart Mill, una relación causal existe si: Una causa precede a un efecto; causa y efecto están relacionados; y si no es posible encontrar alternativas que expliquen el efecto generado más allá que la causa identificada [14, pp. 6–7]. Los diseños experimentales, al ser controlados, se hacen cargo de lo anterior al buscar establecer condiciones en las cuales los efectos se generan principalmente debido a determinados tratamientos (causas) definidos a priori, por lo que son buenas herramientas para encontrar relaciones causales en estudios de toda índole, en contraste a una regresión por ejemplo la que es capaz de detectar de manera directa la correlación entre variables, pero no la causalidad que existe de manera subyacente.

En el caso particular que se estudia en este proyecto, la causalidad que interesa detectar es principalmente si las distintas combinaciones de canales y formas de comunicación probadas generan como respuesta la compra o aceptación del servicio que se le está ofreciendo a los clientes y en qué magnitud esto ocurre, esto último medido como una proporción, lo que permite identificar y comparar distintas combinaciones probados.

2.6. Análisis de experimentos

La ejecución de un diseño experimental debe ir de la mano de una metodología de análisis ad hoc que se haga cargo de explicar con argumentos sólidos si la aplicación de un tratamiento o condición a probar genera efectivamente alguna modificación o impacto sobre la operación de un sistema o bien debido a factores externos no considerados dentro del diseño, que en general se consideran como factores aleatorios. No basta con hacer la mera comparación entre resultados de los distintos tratamientos, dado que de contarse con diferencias muy pequeñas, el efecto generado se podría explicar debido a factores externos al estudio, es decir, a elementos que se consideran aleatorios en el contexto experimental.

Lo que se busca es a partir de herramientas estadísticas, confirmar el efecto de un tratamiento particular, más allá de sencillamente concluir basado en la diferencia entre 2 resultados comparados (grupo tratado vs grupo de control).

2.6.1. Análisis de poder estadístico

El paso inicial tiene que ver con la definición y diseño de los experimentos a realizar, particularmente respecto a los tamaños muestrales requeridos en cada uno de los tratamientos a testear. Esto dado que la mala definición del tamaño muestral requerido puede generar que los resultados del estudio no sean conclusivos debido a la significatividad del test por medio del cual se evalúan las diferencias entre resultados de la variable dependiente.

El análisis antes mencionado se logra a través del análisis de poder a priori [6, pp. 176], el cual es una metodología eficiente para controlar el poder estadístico previo a la implementación de una experiencia experimental. Es un método fácil de implementar y de evaluar por medio de distintos softwares y herramientas que permiten su construcción. En este tipo de análisis, el tamaño muestral se obtiene como una función dependiente de varios parámetros:

- El **nivel de poder estadístico** esperado que se quiere alcanzar con el experimento, definido como $(1-\beta)$, el cual se define como la probabilidad con la cual cierto test rechaza correctamente la hipótesis nula cuando la hipótesis alternativa es verdadera, o bien, la probabilidad de aceptar correctamente la hipótesis alternativa. En términos más simples, el poder denota la capacidad del test de detectar un efecto, vale decir, la capacidad que tuvo el experimento de generar un efecto sobre un grupo tratado.
- **Nivel de significancia** α , el cual hace referencia a la probabilidad de que la conclusión considerada esté errónea, vale decir, sobre qué casos, la asignación de una hipótesis del test será incorrecta. El trade off que se da entre la asignación de un nivel de significancia, impacta directamente sobre el riesgo del error tipo I y el error de tipo II, por lo que la selección de un α apropiado es muy relevante.
- **Tamaño del efecto a ser detectado con probabilidad $(1-\beta)$** , vale decir, el resultado de la variable dependiente que se espera obtener tras la definición. En el caso particular de análisis de test de proporciones (tasas de respuesta por ejemplo), la variable dependiente considera la tasa de cada grupo, y el efecto corresponde a la diferencia observada entre cada uno de los grupos. Esto es ex ante difícil de estimar con precisión, dado que se basa en asunciones que pueden no ser correctas, dado que el efecto se puede sub estimar o sobre estimar, por lo que una práctica útil es la de generar análisis para un conjunto de distintos efectos, logrando obtener así un diagnóstico exhaustivo respecto a los posibles resultados a obtener.

$$Q = f((1 - \beta), \alpha, \gamma)$$

En general, una lógica fácil de seguir al momento de identificar el tamaño muestral necesario para obtener un cierto nivel de poder estadístico, va de la mano con el efecto a obtener, es decir, mientras mayor sea el efecto del diseño experimental (mientras más altas sean las diferencias entre grupo tratado y grupo de control), menor será el Q necesario por muestra para alcanzar un determinado poder estadístico. Esto es así dado que frente a diferencias demasiado pequeñas entre los resultados de grupo tratado versus grupo de control, el tamaño muestral necesario será muy alto, dado que dichas diferencias se pueden atribuir a factores externos (considerados aleatorios) que no dependen del tratamiento mismo, por lo que se concluiría que el tratamiento no tuvo efecto alguno.

Por otro lado, ex post a la aplicación del diseño experimental, y cuando ya se cuenta con los resultados de los distintos tratamientos implementados, se puede generar un análisis de poder alternativo al antes mencionado, considerando ya no suposiciones de efectos, sino el efecto real que se ha obtenido tras la aplicación de un determinado test. Este es el **análisis de poder post hoc**, el cual opera de manera conceptualmente idéntica al análisis a priori, modificando básicamente el rol de los parámetros en consideración. En este análisis, lo que se busca obtener como output es exactamente el poder estadístico alcanzado, como función del nivel de significancia, el efecto generado como consecuencia de la intervención realizada, y los tamaños muestrales utilizados finalmente.

$$(1 - \beta) = f(\alpha, \gamma, Q)$$

Este análisis antes descrito, logra validar los resultados obtenidos y así concluir sobre la capacidad del diseño experimental de detectar efectos atribuidos en forma directa a los tratamientos generados o en contraste, al azar o a factores externos al diseño experimental.

2.6.2. Análisis de test de hipótesis

Como se mencionó antes, los test de hipótesis fundamentalmente permiten generar conclusiones respecto a la evaluación de un determinado estadístico de prueba que se contrasta con un valor tabulado para concluir respecto a un resultado en forma estadísticamente significativa.

Esto permite en el caso particular de los experimentos, el comparar tratamientos y asegurar con una determinada confianza, si el efecto generado tras la implementación de un diseño experimental, si el efecto se puede atribuir de manera correcta a la implementación de un determinado set de condiciones experimentales caracterizadas en cada tratamiento construido.

Capítulo 3

Metodología

3.1. Análisis exploratorio de campañas anteriores

La empresa sobre la cual se aplica el estudio trabaja la administración del marketing directo de acuerdo a un esquema de campañas, las cuales mes a mes definen una oferta específica, asociada a la categoría de productos al cual se está gestionando: móvil o fijo. En este proyecto en particular, el análisis se acota al mundo móvil, dado que si bien desde la perspectiva del negocio hace sentido atacar ambos segmentos y Telefónica es el líder respecto a número de abonados en el mercado móvil (lo que quizás hace inferir que no hay espacio de crecimiento en ese mercado), el mercado fijo en general es más complejo de explorar, principalmente debido a que la decisión de adquisición de servicios (Internet, Televisión de cable) depende de factores menos observables directamente y de requerimientos y preferencias familiares que no son fáciles de inferir desde la data. En el mercado móvil, los servicios son más personales, dado que se ajustan en general a necesidades individuales, por lo que tiene sentido el hecho de que el hallazgo de patrones se tienda a ajustar a variables y características del usuario mismo con las que se cuenta en la empresa. Además, el foco de la compañía se encuentra en 2 aristas fundamentales: incrementar la cantidad de abonados y mejorar los servicios contratados por los clientes. Este último punto hace alusión al hecho de generar acceso a los clientes hacia productos de mayor valor, y por tanto poder generar una extracción de ese valor hacia la compañía, particularmente en los servicios móviles, en donde aún existe espacio para crecer dada la naturaleza del mercado.

El análisis exploratorio y analítico de las campañas anteriores con las que se cuenta, es relevante dado que permite hacer inferencias primero respecto a qué campañas son las que a priori son más relevantes para la compañía, debido a las tasas de efectividad propia que presentan (entendiéndose esto como la cantidad de ventas generadas tras un contacto efectivo con los clientes), y segundo respecto a las interacciones propias que se pueden extraer entre las distintas variables demográficas disponibles. Si bien esto último no permite un análisis tan profundo como es el que se puede lograr a través de una regresión, sí es útil desde la perspectiva de la comprensión de la data y para lograr una mirada general a las distintas respuestas que muestran las campañas dependiendo del tipo de cliente que se evalúe. Se considerará la data del pasado año 2014 para los análisis correspondientes.

Adicionalmente, en la búsqueda de generar una descripción más acabada de las campañas históricas de la compañía, es interesante la ejecución de regresiones para poder medir de manera más clara no solo si las variables a disposición ayudan a explicar si el cliente acepta o no una oferta comercial y así identificar además en qué magnitud se da dicha interacción y poder por consiguiente, concluir respecto a qué campañas son más propicias para un determinado tipo de cliente. Dado que la variable a explicar no es una continua, sino una binaria, las regresiones lineales no son la estrategia más natural a seguir, sino regresiones que consideren este tipo de variables.

3.1.1. Análisis y selección de datos

Las fuentes de datos con las que se cuenta para los análisis pertinentes son principalmente dos. La primera de ellas, corresponde a las distintas campañas que ha generado la compañía en los últimos meses, data que muestra tanto la nomenclatura específica a la que corresponde cada campaña, como así también los clientes específicos a los que se envió a dicha campaña y si estos respondieron en forma positiva a la misma. Se cuenta principalmente con la data de las campañas ejecutadas mediante los canales Call Center e E-Mail, mientras que para el canal SMS solo se posee información respecto a las tasas de recepción de los mensajes.

La segunda gran fuente de datos con que se cuenta, corresponde a la cartera de clientes de la empresa, el cual concentra una gran cantidad de variables transaccionales y demográficas de cada cliente, el cual se actualiza mes a mes para dar cuenta del comportamiento de cada cliente en un periodo particular. Frente al gran número de atributos con las que se cuenta, la eliminación de variables se hace necesaria no solo para reducir la dimensionalidad intrínseca de la base de datos, sino que además para enfocar los esfuerzos solo en aquellas variables que son más explicativas y tienen mayor valor.

En este caso, se usará un método de wrapper, que si bien no es el más eficiente, dado que requiere generar cálculos incrementalmente más complejos al ir introduciendo variables al análisis, si es más exhaustivo al buscar variables relevantes, dado que depende del clasificador que se use. En este caso, se utilizará el clasificador de la regresión logística (ver sección 3.1.6 *Regresión Logística*). Existen 2 modalidades fundamentales de este tipo de métodos:

- **Forward selection**, el cual considera comenzar entrenando un modelo sin variables y compararlo versus otro al ir agregando variables una a la vez y así determinar qué modelo es el que mejor se desempeña (en general dicha comparación se realiza de acuerdo a una métrica estándar, en este caso el criterio de información de Akaike AIC).
- **Backward elimination**, el cual considera una metodología análoga a la anterior, pero en lugar se comienza con un modelo entrenado con todas las variables disponibles y comparar eso versus modelos con una variable menos por vez.

El enfoque que considera la combinación de ambos, o sea la *Bidirectional elimination* es la metodología seleccionada, dado que permite una selección bastante exhaustiva de las variables al considerar prácticamente todas las combinaciones de modelos que se pueden generar con un set determinado de atributos disponibles al ir, en cada iteración, considerando si la agregación o eliminación de una determinada variable es lo que más información le agrega al modelo.

3.1.2. Tratamiento de *valores atípicos y valores faltantes*

La implementación de los métodos wrapper para la selección de variables implica un desafío adicional, y es que previo a la implementación de estos, se requiere de una corrección previa de la data. Fundamentalmente, esto se da por los valores faltantes y los valores atípicos fuera de rango (o outliers).

Se entenderá por valores atípicos a aquellos casos que se escapan del rango intercuartil de una variable particular, siendo el rango intercuartil la diferencia entre la posición donde comienza el cuarto cuartil de la variable en cuestión y la posición donde termina el primer cuartil. El criterio general para definir un valor atípico es verificar si se encuentra fuera de los extremos del rango intercuartil ponderado por 1.5. Dependiendo de la cantidad de valores atípicos presentes en cada variable, se decidirá si es más propicio eliminar esos casos, o reemplazarlos aplicando alguna técnica de imputación.

Los valores faltantes corresponden efectivamente a eso: valores que no se encuentran en la data correspondiente a un atributo particular. Esto se puede deber a un número de factores: error al cambiar el formato de una tabla, o en la recopilación misma de la data por parte de un ejecutivo de ventas, malas prácticas en la administración del warehouse, etc. Nuevamente, se analizará caso a caso las variables para determinar el curso de acción a implementar.

Las imputaciones de variables, o reemplazo valores faltantes o fuera de rango, son útiles dado que permiten corregir en parte estas situaciones. No obstante, no son 100 % confiables, dado que pueden inducir errores en la data. Además, el proceso en ocasiones puede ser demasiado engorroso para que se justifique la aplicación de imputaciones a todas las variables que presenten problemas de la especie anterior. Por estas razones, solo se aplicarán técnicas de imputación en aquellos casos en donde esta sea más bien directa (por ejemplo, encontrar la edad de los individuos por medio del rut).

3.1.3. Análisis descriptivo

El análisis descriptivo de variables busca básicamente identificar inicialmente el tipo de variable con el que se cuenta y, posteriormente, construir medidas agregadas que logren resumir la data de manera de entender de manera rápida y en forma sencilla que describe la data disponible.

El análisis contrasta fundamentalmente de acuerdo al tipo de variable:

- **Variables categóricas**, los análisis se reducen fundamentalmente a uno de frecuencias y de porcentajes, vale decir, la cantidad de repeticiones que una categoría específica de una variable se repite y cuál es la proporción de la misma versus el total. Para análisis de más de una variable, la aplicación de tablas de contingencia es particularmente útil, puesto que permiten generar un cruce inicial entre las variables, e identificar ciertas relaciones iniciales que se puedan dar entre los individuos.

Más aún, los porcentajes relativos entre variables permiten el cálculo de intervalos de confianza de manera simple para argumentar de manera más clara respecto a las

diferencias que podrían aparecer entre variables a través de una simple inspección.

| | | GSE | | | | |
|--------|-----------|------|-----|----|----|-----|
| | | ABC1 | C2 | C3 | D | E |
| Genero | Masculino | 100 | 150 | 20 | 30 | 100 |
| | Femenino | 123 | 200 | 50 | 40 | 52 |

Tabla 3.1: Ejemplo de tabla de contingencia

La tabla de contingencia de ejemplo antes descrita, muestra el análisis de la variable genero versus la variable GSE. A priori se observa que en la muestra seleccionada, el género femenino se encuentra más homogéneamente distribuido en los distintos grupos socioeconómicos, mientras que en el género masculino, existe más variabilidad.

Este análisis se puede complementar aplicando tests de independencia de variables, como el test chi-cuadrado de independencia, y lograr así determinar si efectivamente las variables en análisis presentan cierta correlación y así decidir si se justifica el eliminar una de ellas del análisis.

- **Variables cuantitativas**, los análisis se enfocan principalmente en el resumir la data. Para ello, indicadores como los histogramas, la media, varianza y mediana son útiles para de esa forma, observar como la data se distribuye a través de la data.

Otros análisis más acabados por ejemplo corresponden a la obtención de boxplots, o diagramas de caja, los cuales ayudan a comprender de manera más clara como se distribuye la variable en análisis, y además identificar aspectos como el rango de la variable, los outliers, entre otros (como se describe en la sección *3.1.2 Tratamiento de valores atípicos y valores faltantes*).

3.1.4. Extracción de dinámicas

Si bien las variables por sí solas son un buen punto de partida para buscar patrones que logren caracterizar a cierta variable dependiente, el análisis puede enriquecerse a través de la construcción de nuevos inputs, derivados de las variables con las que se cuenta y de esa forma, obtener inputs más relevantes que ayuden a generar una predicción.

Para aquellas variables de tráfico de los clientes, se decidió seguir este enfoque al considerar como elemento relevante no solo los comportamientos del propio mes en el cual se está analizando la propensión a la adquisición de un servicio, sino que también el comportamiento de los meses anteriores, y de esa forma buscar detectar si ciertas tendencias propias de cada variable aportan a la predictibilidad. Tomando esto en consideración, los nuevos inputs que se proponen son:

- Media de cada variable.
- Desviación estándar de cada variable.
- Aumento de cada variable.
- Disminución de cada variable.

Lo anterior, considerando la historia de los últimos seis meses anteriores de cada cliente.

Por supuesto que la media considera principalmente, el identificar un comportamiento agregado de los últimos meses, sin mirar las variaciones periodo a periodo que se puedan generar, sino sencillamente resumir sin ir más lejos. Por su parte, la desviación estándar busca complementar lo anterior al explicar en parte, la variabilidad del comportamiento de los individuos. No obstante, esto no permite identificar de manera clara, en qué forma el consumo de los clientes se ve modificado en el tiempo. La duda que aparece es ¿Hubo un aumento significativo en las variables correspondientes a cada cliente en los últimos 6 meses? ¿Hubo una disminución significativa? ¿Se mantuvo estable? La tercera variable antes propuesta busca ser un aporte en esa línea.

Se entenderá como un aumento a variaciones superiores al 30 % mes a mes dentro de los 6 meses de análisis, y el mismo enfoque a la inversa para las disminuciones. En la sección siguiente se describirá en forma explícita el tipo de variables que se construyeron para lograr definir aumentos o disminuciones que se desea extraer.

Si bien a priori no se tienen argumentos suficientes para justificar si la extracción de estas tendencias generará un aporte importante a la calidad predictiva del modelo que se desea construir, es prudente explorar cómo este tipo de elementos puede influir en los clientes, puesto que en general en la industria de los servicios, los procesos decisivos son más largos y dependen de tanto la satisfacción con el servicio mismo como así también con la variación en las necesidades de los clientes.

3.1.5. Construcción de inputs adicionales

Los inputs que se identifican como relevantes de construir para lograr identificar las tendencias en la data son los siguientes:

- **Diferencia entre mes $i+1$ y mes i** , la cual básicamente es la diferencia entre el valor de ambos meses para la variable correspondiente.
- **Cambio entre mes $i+1$ y mes i** , la cual opera como una binaria que toma valor 1 si la diferencia al alza es superior al 30 % del mes i , y valor -1 si la diferencia a la baja es superior al 30 %.
- **Tendencia variable**, la cual es única para cada cliente, y toma valor 1 si el cambio agregado en los últimos 6 meses es mayor a 0 y toma valor -1 si ocurre lo contrario.

Finalmente, separando esta última variable en 2 binarias "aumento" "disminución" de cada variable correspondiente, se obtiene un estimado del patrón al alza o a la baja que presentan las variables. Se decidió así puesto que en la realidad, existen 3 posibles tendencias que se pueden dar: aumento, descenso, estabilidad, siendo este último el caso en el que en los 6 meses de análisis no se presenten alzas ni bajas importantes en las variables. (En el anexo se observa un ejemplo práctico de la construcción de los inputs).

3.1.6. Regresión logística

Con el marco antes descrito en la sección anterior, se seguirán dos enfoques para generar el análisis y describir como las variables disponibles afectan la decisión de los individuos:

1. **Un modelo agregado**, asumiendo que la decisión de los clientes es binaria, es decir, la decisión es comprar o no comprar (o suscribirse o no a un servicio de telefonía móvil). Aquí se asumirá que todas las campañas a las que se ven expuestos los clientes corresponden a un solo gran grupo de opciones, dado que en general, los clientes que prospectan para una campaña específica, no prospectan para otra, por lo que este enfoque tiene sentido desde el punto de vista del negocio.
2. **Modelos desagregados por campaña**, vale decir, estimar una regresión logística para cada campaña. Esto debido a que en cada campaña, pueden existir diferencias particulares respecto a cómo los clientes toman la decisión de acceder a la oferta y como se producen las interacciones de las variables explicativas con la variable objetivo, y de esa forma obtener una caracterización más detallada del proceso decisivo de los clientes.

La variable dependiente corresponde a una binaria que toma valores 1 o 0 dependiendo de si los clientes aceptan la oferta comercial que se les está haciendo, es decir, si los esfuerzos de marketing directo rindieron frutos y lograron el objetivo final de generar una venta de un servicio de telefonía móvil:

$$y_n = \begin{cases} 1 & \text{si cliente } n \text{ compra producto} \\ 0 & \text{si no} \end{cases}$$

La metodología clásica de analizar e interpretar los resultados generados tras la estimación de estos modelos se describe en la sección siguiente *3.1.7 Descripción de interacción de variables*, metodología conceptualmente sencilla, existen ciertas distinciones importantes a considerar al momento de examinar el output generado por el software de análisis estadístico utilizado.

Un detalle adicional que es importante considerar, es que dado el desempeño propio de las campañas, la cantidad de casos favorables es bastante menor a la cantidad de casos no favorables. Esto se visualiza con mayor detalle en la sección de análisis de resultados, donde se ve que en términos agregados, las campañas por canal Call Center, presentaron una tasa de efectividad entre 1.6 % y 2.4 % para las campañas a clientes de la compañía, mientras que para clientes que no pertenecen a la compañía, la efectividad se encuentra en torno al 0.3 % - 0.8 %. Lo primero quiere decir que de cada 100 contactos efectivos generados con los clientes, se generan entre 1.6 y 2.4 altas de un producto para los clientes de la compañía.

Un enfoque pertinente para atacar esta situación es la que proponen King y Zeng (2001) [9, pp. 143–145], el cual se hace cargo a través de la introducción de factores de ajuste sobre las estimaciones de los parámetros, dependientes de la proporción de éxitos en la muestras consideradas. Los autores mencionan que esta metodología tiene sentido cuando se tiene por ejemplo una efectividad de *éxitos* en torno al 0.3 %, orden de magnitud similar a las efectividades observadas en campañas enfocadas en clientes de la competencia de acuerdo a lo antes mencionado.

Asumiendo que se conoce la proporción de éxitos.^{en} la población τ , se debe seleccionar una muestra para la estimación del modelo, cuya proporción de éxitos \bar{y} depende del muestreo generado. A partir de estos indicadores, se construyen 2 pesos que se incorporan a la función de verosimilitud:

- $w_1 = \frac{\tau}{\bar{y}}$
- $w_0 = \frac{(1-\tau)}{(1-\bar{y})}$

Con esto, los autores proponen una nueva estimación de máxima verosimilitud ajustada:

$$\ln(L_w(\beta|y)) = w_1 \sum_{\{y_i=1\}} \ln(\pi_i) + w_0 \sum_{\{y_i=0\}} \ln(1 - \pi_i)$$

Formulación que le asigna mayor peso a aquel grupo de respuestas que presenta menos casos (y que por ende, podría no expresarse con la fuerza suficiente para ser representado de buena manera por el modelo), logrando describir de manera más eficiente las posibles respuestas de los individuos, y así generar mejores predicciones.

3.1.7. Descripción de interacción variables

Como ya se ha descrito en la sección anterior, la variable dependiente que interesa describir es la respuesta positiva o negativa de un cliente cuando este se ve enfrentado a una oportunidad de contratación de un servicio de telefonía móvil, por tanto, la situación corresponde a una decisión binaria.

En general, los outputs que se generan al aplicar la estimación de un modelo de regresión logística que se ajusta a las características descritas en la sección anterior, son los siguientes:

- Estimador del coeficiente.
- Error estándar.
- Estadístico z de Wald.
- P valor test z de significatividad individual.

Los coeficientes principalmente indican en que magnitud las variables influyen la realización de la variable dependiente binaria. El signo de estos explica en qué dirección se manifiesta esa influencia, interpretándose así directamente y de manera exploratoria como cada variable independiente se relaciona con la variable objetivo.

El error estándar se puede interpretar como la precisión con la cual el modelo estimó el valor desconocido del coeficiente en cuestión. La relevancia de este valor es que permite el cálculo del estadístico z, el cual permite verificar si el coeficiente es significativo, y si se debe considerar dentro de la estimación del modelo como un término relevante.

El estadístico z se calcula sencillamente dividiendo el valor del estimador del coeficiente por su error estándar. Este valor se contrasta contra una distribución normal $0,1$, obteniéndose un test de hipótesis cuyo contraste considera las hipótesis $H_0: \beta_0 = 0$; $H_1: \beta_0 \neq 0$. Por tanto, el coeficiente tendrá significancia y por lo tanto tendrá valor su incorporación en el modelo cuando el test rechace la hipótesis nula. Por último, la columna final muestra con qué confianza estadística el parámetro es significativo. Tres asteriscos señalan una significatividad del 100 %, es decir, el p-valor es muy cercano a cero; 2 asteriscos muestran una significatividad del 99 %, o sea un p-valor inferior a 0.01 y así sucesivamente. Depende de cada situación particular el nivel de significatividad aceptado, pero en general, un 95 % se considera como un buen nivel.

Finalmente, otro elemento del output que vale la pena destacar se relaciona con la desviación nula y la desviación residual del modelo. La primera señala qué tan bien se desempeña el modelo sin la incorporación de variables, vale decir, un modelo estimando solo el intercepto, mientras que la desviación residual muestra el desempeño del modelo con las variables incorporadas.

3.2. Diseño experimental

El diseño experimental que se desea implementar, considera primero y principalmente, la ejecución de estímulos comunicativos hacia los clientes de la compañía por medio de los canales de comunicación fundamentales de marketing directo con los que se cuenta:

- Call Center
- SMS - Mensaje de Texto
- E-Mail

A través del testeo de las distintas combinaciones que se pueden generar entre ellos, se busca definir qué rol tiene cada canal particular en una interacción con los clientes, y además, como se pueden complementar para así lograr mejorar la efectividad propia de las campañas, es decir, lograr mejores tasas de contactabilidad y tasas de ventas.

Con el objetivo de que la única variable que se modifique a través de todas las pruebas sea solamente el uso de canales en una instancia comunicativa hacia el cliente, se definió mantener la oferta estática, considerando la campaña de "Segunda línea personas con la que cuenta la empresa, la cual considera el ofrecimiento de un nuevo producto móvil (dispositivo + plan) a clientes que ya poseen un producto con la compañía. De esa manera, todos los clientes expuestos al diseño experimental, contarán con la misma oferta fundamental, y solo se modificará la configuración de contactabilidad seleccionada.

3.2.1. Selección de prospectos

La campaña de "Segunda línea personas" considera ciertas condiciones del negocio para determinar cuáles clientes son los que prospectan para dicha campaña. Estas condiciones

consideran por ejemplo la cantidad de líneas con las que actualmente cuenta el cliente, o que no registre como dado de baja en el algún momento reciente. A partir del cumplimiento de estas condiciones iniciales, se tiene la base total desde donde se genera la selección de los clientes que se asignarán a las distintas condiciones.

Para ambas etapas del diseño experimental, se generó un muestreo aleatorio simple para la asignación de clientes a cada una de las condiciones experimentales, buscando generar muestras lo más similares entre sí y de esa manera evitar la introducción de variables externas que puedan ensuciar los resultados de los distintos tratamientos. Esto asegura además independencia entre las muestras, vale decir, que el resultado de un individuo perteneciente a un tratamiento no está correlacionado con el resultado de otro individuo cualquiera en otro tratamiento [5, pp. 4–9].

Si bien el muestreo aleatorio simple teóricamente reduce cualquier sesgo de muestreo que se pueda generar tras la selección de los individuos, aun así, puede darse que las muestras no sean tan similares como se desea (un muestreo con una edad promedio muy distinta a otra por ejemplo, la proporción de regiones de vivienda de los clientes, etc.), por lo que la asignación introduciría cosas que no se quieren introducir al diseño experimental. Para evitar esto, se aplicará un re-muestreo aleatorio", el cual considera una repetición del muestreo aleatorio simple en aquellos casos en los cuales no se cumplan ciertas condiciones entre las muestras correspondientes a cada tratamiento. Si bien esto no es estrictamente necesario, se considera como un elemento adicional para mantener una alta similaridad en las muestras.

Las variables a considerar son las siguientes:

- Edad
- Región
- Antigüedad línea
- Promedio minutos de salida (últimos 6 meses)

Los criterios corresponden básicamente a la aplicación de tests estadísticos que muestren si tanto las medias (test t-student para 2 muestras independientes) y las varianzas (test F para dos varianzas) de las distintas variables son iguales muestra a muestra. Para el caso particular de la variable Región", se aplicará un test de proporciones (test Chi cuadrado para dos proporciones).

3.2.2. Asignación de condiciones y tamaño muestral

El diseño experimental propuesto, considera la implementación de 2 etapas de testeo, cada una aplicada en meses distintos, Enero y Marzo de 2015, para lograr identificar las respuestas de los clientes durante situaciones similares (misma oferta, mismos filtros para determinar prospectos), pero con las distinciones particulares producto del mes en el que se encuentran: Enero es un mes de menor tráfico (provocado por el periodo vacacional), mientras que Marzo representa de mejor manera la operación del resto del año. Se hace necesaria la aplicación de dos etapas de experimentación principalmente debido al contexto anómalo en el que se ejecuta la primera etapa durante el mes de Enero, en el cual las ventas son inferiores al resto

del año debido a la estacionalidad natural producto del periodo de vacaciones.

Marzo es un mejor representante del resto del año, por lo que los resultados obtenidos en ese periodo tienen mayor validez como argumento de conclusión. Febrero se descarta debido a que el efecto de las vacaciones es aún más marcado que en Enero y más importante, debido a que se requiere de un periodo de aprendizaje de la primera etapa antes de implementar la segunda.

De esta manera, el objetivo de la primera etapa es lograr un entendimiento general del comportamiento de los clientes bajo este marco multicanal y además aprender a través de la experiencia de las posibles complicaciones y errores que se pueden generar al aplicar el diseño experimental, mientras que la segunda etapa.

Cada etapa considera además distintos bloques de testeo, cada uno de ellos buscando la validación de distintas hipótesis particulares (Ver sección 4.2.2 *Hipótesis a testear*).

Enero 2015

El primer bloque de Enero busca validar hipótesis de multicontactabilidad, vale decir, verificar como responden los clientes frente a contactos reiterados por un mismo canal versus contactos individuales (este último el método que corresponden al manejo típico de las campañas en la empresa), limitando la construcción de tratamientos a solo 2 instancias de contacto por tratamiento. Con esto, se construyen 7 tratamientos descritos en la tabla a continuación.

| | Multicontacto | |
|---------------|-----------------------|-------------|
| | Instancia 1 | Instancia 2 |
| Tratamiento 1 | E-Mail | E-Mail |
| Tratamiento 2 | E-Mail | |
| Tratamiento 3 | SMS | SMS |
| Tratamiento 4 | SMS | |
| Tratamiento 5 | Call Center | Call Center |
| Tratamiento 6 | Call Center | |
| Tratamiento 7 | Control: No contactar | |

Tabla 3.2: Tratamientos testeados Enero - Multicontacto

Lo interesante de lo anterior, es que al comparar por ejemplo los resultados de los tratamientos 1 contra el 2, el 3 contra el 4 y el 5 contra el 6, se puede explorar el efecto adicional que se genera por un segundo contacto prácticamente idéntico al primero.

Una limitante que puede aparecer de este primer bloque, es que efectos negativos de mediano y largo plazo sobre los clientes no se pueden extraer directamente de los resultados, al menos no inicialmente. Esto debido a que debido a la sensibilidad particular de cada cliente, dos contactos en forma reiterada en un espacio temporal reducido puede generar sobre ellos

cierto rechazo, lo que puede causar - en un caso extremo - que un cliente dé de baja el servicio que ya tiene contratado con la compañía.

Se puede considerar que en caso de dos instancias de contacto con un determinado cliente, ciertos efectos negativos se podrán identificar cuando la decisión de contratación del servicio ocurra en un momento muy posterior a la primera instancia de contacto, o sea, en aquellos casos en que al recibir un segundo contacto, el cliente aún se encuentra en la toma de decisión provocado por la primera instancia de contacto. Todo lo anterior aplica dependiendo de cómo y en qué momento los clientes se vean motivados a tomar la decisión.

El segundo gran bloque de la primera etapa de testeo considera ya la aplicación de condiciones multicanal, al combinar los 3 canales disponibles en todas las posibles variaciones que se puedan generar, nuevamente, limitado a solo 2 instancias de contacto. Este, busca validar hipótesis de interacción de los distintos canales al explorar si la posición particular de un canal como así también el antecesor particular del mismo provoca comportamientos diferentes a través de los clientes en examinación. Los 6 tratamientos que se construyen con esto se muestran en la tabla a continuación:

| | Multicanal 2 contactos | |
|----------------|------------------------|-------------|
| | Instancia 1 | Instancia 2 |
| Tratamiento 8 | E-Mail | Call Center |
| Tratamiento 9 | Call Center | E-Mail |
| Tratamiento 10 | SMS | Call Center |
| Tratamiento 11 | Call Center | SMS |
| Tratamiento 12 | SMS | E-Mail |
| Tratamiento 13 | E-Mail | SMS |

Tabla 3.3: Tratamientos testeados Enero - Multicanal 2 contactos

Marzo - Abril 2015

Dada la historia que se tiene de las distintas campañas realizadas y más importante, a las limitaciones de cantidad de registros y tratamientos que se pueden gestionar en paralelo, no se considera pertinente repetir exactamente los mismos tratamientos implementados en Enero en la ejecución de Marzo, particularmente de aquellos tratamientos donde solo se usa el canal SMS o el canal E-Mail (canales con bajas tasas históricas de efectividad). Además, dado el estado actual de las campañas, uno de los baseline contra los cuales es más interesante comparar es aquella combinación que solo utiliza el Call Center como canal de comunicación, el cual es el más usado en las campañas de la compañía.

Tomando en consideración lo antes descrito, los resultados parciales de la ejecución de la etapa de Enero del diseño experimental, como así también las restricciones de cantidad de prospectos con los que se cuentan para esta segunda etapa de experimentación, se definieron los siguientes bloques de experimentación:

| | Multicontacto | |
|---------------|-----------------------|-------------|
| | Instancia 1 | Instancia 2 |
| Tratamiento 1 | Control: No contactar | |
| Tratamiento 2 | E-Mail | E-Mail |
| Tratamiento 3 | SMS | SMS |
| Tratamiento 4 | Call Center | Call Center |
| Tratamiento 5 | Call Center | |

Tabla 3.4: Tratamientos testeados Marzo - Multicontacto

| | Multicanal 2 contactos | |
|----------------|------------------------|-------------|
| | Instancia 1 | Instancia 2 |
| Tratamiento 6 | E-Mail | Call Center |
| Tratamiento 7 | Call Center | E-Mail |
| Tratamiento 8 | SMS | Call Center |
| Tratamiento 9 | Call Center | SMS |
| Tratamiento 10 | SMS | E-Mail |
| Tratamiento 11 | E-Mail | SMS |

Tabla 3.5: Tratamientos testeados Marzo - Multicanal 2 contactos

| | Multicanal 3 contactos | | |
|----------------|------------------------|-------------|-------------|
| | Instancia 1 | Instancia 2 | Instancia 3 |
| Tratamiento 12 | SMS | E-Mail | Call Center |
| Tratamiento 13 | E-Mail | SMS | Call Center |
| Tratamiento 14 | SMS | SMS | Call Center |
| Tratamiento 15 | E-Mail | E-Mail | Call Center |
| Tratamiento 16 | SMS | Call Center | Call Center |
| Tratamiento 17 | E-Mail | Call Center | Call Center |
| Tratamiento 18 | SMS | SMS | E-Mail |
| Tratamiento 19 | E-Mail | E-Mail | SMS |
| Tratamiento 20 | E-Mail | Call Center | E-Mail |
| Tratamiento 21 | Call Center | E-Mail | SMS |
| Tratamiento 22 | Call Center | SMS | E-Mail |
| Tratamiento 23 | Call Center | E-Mail | E-Mail |
| Tratamiento 24 | Call Center | E-Mail | Call Center |
| Tratamiento 25 | Call Center | Call Center | E-Mail |

Tabla 3.6: Tratamientos testeados Marzo - Multicanal 3 contactos

Al introducir un tercer bloque de experimentación multicanal en 3 contactos, se espera lograr responder con aun más argumentos (de los que puede entregar el segundo bloque) a las hipótesis respecto a los roles del canal en la operación comunicativa hacia los clientes. Si bien, es deseable probar todas las posibles combinaciones que se pueden generar a partir de los 3 canales en testeo (27 combinaciones en total), la restricción de recursos impide que ello

sea posible, tanto en la cantidad de individuos que se pueden considerar dentro del diseño experimental como así también por el alto costo de recursos durante la operación misma de los contactos.

Para la determinación de los tamaños muestrales necesarios para que cada condición experimental alcance un determinado poder estadístico, y por tanto, que las conclusiones finalmente tengan una base estadística sólida, se implementará un análisis a priori de poder. Este, básicamente considera un test z de proporciones, al cual a través de ciertos inputs, entrega el tamaño muestral necesario para alcanzar un determinado poder estadístico. Por medio del software *G-Power* esto se puede lograr de manera rápida, al entregarle los inputs: proporción esperada grupo 1, proporción esperada grupo 2, poder estadístico deseado y significatividad estadística. Un ejemplo de aplicación para un baseline (proporción 1) de 2% se puede visualizar a continuación:

| | | | Tamaño muestral necesario (por muestra) | | | | | | | | |
|---|-----------------|---|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Tasa de respuesta Grupo Control (baseline) p1 | Efecto absoluto | Tasa de respuesta Grupo Tratado (efecto) p2 | 50% poder | 55% poder | 60% poder | 65% poder | 70% poder | 75% poder | 80% poder | 85% poder | 90% poder |
| 2.00% | 0.10% | 2.10% | 154254 | 174676 | 196725 | 220887 | 247864 | 278719 | 315206 | 360567 | 421971 |
| | 0.20% | 2.20% | 39484 | 44712 | 50355 | 56540 | 63445 | 71343 | 80682 | 92292 | 108009 |
| | 0.30% | 2.30% | 17957 | 20335 | 22901 | 25714 | 28854 | 32446 | 36693 | 41973 | 49121 |
| | 0.40% | 2.40% | 10331 | 11698 | 13175 | 14793 | 16599 | 18666 | 21109 | 24146 | 28258 |
| | 0.50% | 2.50% | 6759 | 7653 | 8619 | 9678 | 10859 | 12211 | 13809 | 15796 | 18486 |
| | 0.60% | 2.60% | 4796 | 5430 | 6116 | 6866 | 7705 | 8664 | 9798 | 11207 | 13116 |
| | 0.70% | 2.70% | 3598 | 4074 | 4588 | 5152 | 5781 | 6500 | 7351 | 8408 | 9840 |
| | 0.80% | 2.80% | 2812 | 3184 | 3586 | 4026 | 4518 | 5080 | 5745 | 6571 | 7690 |
| | 0.90% | 2.90% | 2267 | 2567 | 2891 | 3246 | 3642 | 4095 | 4631 | 5297 | 6199 |
| | 1.00% | 3.00% | 1873 | 2121 | 2388 | 2681 | 3009 | 3383 | 3826 | 4376 | 5121 |
| | 1.10% | 3.10% | 1578 | 1787 | 2012 | 2259 | 2535 | 2850 | 3223 | 3687 | 4314 |
| | 1.20% | 3.20% | 1351 | 1530 | 1723 | 1935 | 2171 | 2441 | 2760 | 3157 | 3694 |
| | 1.30% | 3.30% | 1173 | 1328 | 1496 | 1679 | 1884 | 2118 | 2396 | 2740 | 3206 |
| | 1.40% | 3.40% | 1030 | 1166 | 1313 | 1475 | 1654 | 1860 | 2103 | 2406 | 2815 |
| | 1.50% | 3.50% | 914 | 1034 | 1165 | 1308 | 1467 | 1649 | 1865 | 2133 | 2496 |
| | 1.60% | 3.60% | 817 | 925 | 1042 | 1170 | 1312 | 1475 | 1668 | 1908 | 2233 |
| | 1.70% | 3.70% | 736 | 834 | 939 | 1054 | 1182 | 1329 | 1503 | 1719 | 2012 |
| | 1.80% | 3.80% | 668 | 756 | 852 | 956 | 1073 | 1206 | 1364 | 1560 | 1825 |
| | 1.90% | 3.90% | 610 | 690 | 777 | 872 | 979 | 1100 | 1244 | 1423 | 1665 |
| | 2.00% | 4.00% | 559 | 633 | 713 | 800 | 898 | 1009 | 1141 | 1305 | 1527 |
| | 2.10% | 4.10% | 516 | 584 | 657 | 738 | 828 | 930 | 1052 | 1203 | 1407 |
| | 2.20% | 4.20% | 477 | 540 | 608 | 683 | 766 | 861 | 974 | 1113 | 1303 |
| | 2.30% | 4.30% | 444 | 502 | 565 | 634 | 712 | 800 | 905 | 1035 | 1210 |
| | 2.40% | 4.40% | 414 | 468 | 527 | 592 | 664 | 746 | 844 | 965 | 1129 |
| | 2.50% | 4.50% | 387 | 438 | 493 | 553 | 621 | 698 | 789 | 902 | 1056 |
| | 2.60% | 4.60% | 363 | 411 | 463 | 519 | 583 | 655 | 740 | 847 | 990 |
| | 2.70% | 4.70% | 342 | 387 | 435 | 489 | 548 | 616 | 697 | 796 | 932 |
| | 2.80% | 4.80% | 322 | 365 | 411 | 461 | 517 | 581 | 657 | 751 | 879 |
| | 2.90% | 4.90% | 305 | 345 | 388 | 436 | 489 | 549 | 621 | 710 | 831 |
| | 3.00% | 5.00% | 289 | 327 | 368 | 413 | 463 | 521 | 588 | 673 | 787 |

Tabla 3.7: Análisis a priori de poder estadístico

Como se observa, fijando el efecto absoluto o mejor, las diferencias absolutas en torno al 1% (considerando un $p1=2\%$), para alcanzar un poder estadístico en torno al 75%, se requiere de un tamaño muestral cercano a 3.000. Se definió que para esta clase de experimentos, ese poder es suficiente para justificar diferencias entre los distintos tratamientos, por lo que se

definió que el tamaño de las muestras a utilizar será de 3.000 sujetos por tratamiento.

3.2.3. Hipótesis a testear

Las hipótesis que se busca testear a partir de lo anterior, se correlacionan en forma directa con cada uno de los bloques de experimentación. Es decir, cada bloque particular apunta a responder planteamientos e ideas propias relacionadas con los tratamientos correspondientes.

Para el bloque **MULTICONTACTO**, las hipótesis que se plantean son las siguientes:

***H1:** Contactar a clientes por medio de multi-contactos genera mejores tasas de respuesta que contactos únicos por medio de un canal específico.*

***H2:** El marketing directo genera tasas de respuesta que justifiquen la implementación de esta clase de estrategias.*

Para la primera hipótesis, la comparación de los tratamientos 1 - 2, 3 - 4 y 5 - 6 de la etapa de experimentación de Enero son relevantes, dado que lo que interesa primordialmente, es identificar las diferencias entre los resultados de contactar una versus más de una vez a los clientes. Esta hipótesis se puede testear además agregando tratamientos en los que se tiene solo una instancia de contacto versus aquellos en los que se consideran 2, es decir, considerando en conjunto los tratamientos 1, 3 y 5 y compararlos con el 2, 4 y 6.

Por su parte, el objetivo de la hipótesis número 2 es lograr comparar el desempeño de las campañas de marketing directo, expresado en los tratamientos 1 a 6, versus el tratamiento número 7, que considera un grupo aislado de clientes en los cuales no existe estímulo alguno, sino que cualquier respuesta o intención de parte de este grupo a la adquisición de un servicio, se puede atribuir a otros factores externos (como por ejemplo el boca a boca, o el marketing masivo).

No obstante, todo lo anterior, aún no da cuenta de cómo los distintos canales disponibles pueden complementarse y generar sinergias importantes que justifiquen la implementación de campañas coordinadas a través de los canales. Las hipótesis del bloque **MULTICANAL EN 2 CONTACTOS** atiende hipótesis que van en esa línea:

***H3:** La combinación de canales genera mejores tasas de respuesta que las tasas de respuesta individuales de los canales utilizados.*

***H4:** El orden en que se utilizan los canales tiene influencia en la respuesta de los prospectos, debido al rol particular que cada canal tiene en una interacción con los clientes.*

La hipótesis **H3** busca identificar esas sinergias que pueden aparecer entre los distintos canales, proponiendo un planteamiento similar al que señala la hipótesis **H1**, pero con la distinción de que en este caso se están considerando ya no un mismo canal repetido, sino combinaciones, razón por la cual se espera que existan diferencias con los resultados obtenidos en el bloque **MULTICONTACTO** del diseño experimental.

La hipótesis número 4 por otro lado, plantea la identificación no solo de la cantidad y combinación de los canales, sino que también cual secuencia u forma de ordenar los canales en una interacción con los clientes, es más efectiva y genera por tanto mejores tasas de contactabilidad y de ventas. Esto se relaciona íntimamente con la definición del rol de canal que es de interés principal para el desarrollo de esta memoria. Por ejemplo, si se considera la combinación que va desde el canal menos invasivo al más invasivo como aquella definida por: **SMS - E-Mail - Call Center**, se esperaría que esta combinación tenga un mejor desempeño que otra que considere los mismos tres canales pero en otro orden.

Respecto al tercer bloque de experimentación considera la implementación de tratamientos **MULTICANAL EN 3 CONTACTOS**, las hipótesis que se plantean van en una línea similar a lo ya mencionado, planteando proposiciones tanto de la cantidad de instancias de contacto generadas como así también respecto a qué estrategias parecieran ser particularmente mejor, es decir, cuales canales funcionan mejor en qué orden.

H5: La combinación de canales durante 3 instancias de contacto genera mejores tasas de respuesta que aquellas obtenidas mediante solo 2 instancias de contacto.

H6: Los 3 canales considerados tienen roles definidos en un proceso comunicativo. El orden que mejores tasas de respuesta genera es aquel que va desde canales menos invasivos a aquellos más invasivos.

La hipótesis número 5 apunta fundamentalmente a lo ya mencionado, lo que se puede examinar al comparar en conjunto los resultados del tercer bloque de experimentación versus el primero y segundo bloque de experimentación en forma agregada. La hipótesis número 6 por su parte apunta directamente a la identificación de aquellos roles más relevantes para cada canal particular, lo que dará insights respecto a cuales estrategias son más relevantes y más efectivas. La segunda parte de esta hipótesis es mucho más direccional y señala, en forma indirecta, que no debería ser raro esperar que las mejores estrategias de contactabilidad son aquellas que van en aumento en la invasividad de los propios canales: **E-Mail - SMS - Call Center** corresponde a la combinación que de manera más directa cumple este punto de acuerdo a las consideraciones ya mencionadas.

3.2.4. Análisis de resultados

La gran interrogante que se debe resolver al momento de analizar las diferencias obtenidas tras la ejecución de un diseño experimental es: ¿Las diferencias se dan efectivamente debido a la intervención propuesta a través del diseño experimental o más bien debido al azar? Los análisis descritos a continuación buscan lograr responder a esa interrogante y entregar argumentos estadísticos sólidos para lograr identificar diferencias entre los distintos tratamientos en testeo.

Test de comparación de proporciones

La primera mirada que se puede implementar para la comparación de 2 proporciones, es por medio de un sencillo test z de 2 proporciones, el cual permite el contraste de tanto hipótesis a 2 colas como a una cola, dependiendo de cómo se comparen las distintas proporciones en cuestión. Existen por supuesto, condiciones que este test requiere para que su aplicación tenga sentido y ser una metodología de análisis válida:

- El muestreo de cada población debe hacerse a través de un muestreo aleatorio simple.
- Las muestras son independientes.
- Las muestras presentan al menos 10 éxitos.
- Cada población es al menos 10 veces tan grande como la muestra.

En el caso particular que esta memoria examina, cada tratamiento corresponde a una muestra tomada en forma aleatoria y de forma independiente, es decir, individuos asignados a cada tratamiento son distintos a los de otros asignados a otros tratamientos (lo que se asegura a través del código del cliente), por lo que las primeras 2 condiciones son cumplidas. La tercera condición por otro lado, depende en parte de cómo se desempeñe cada tratamiento. Si bien, el análisis de campañas previas muestra que para las campañas consideradas equivalentes a la que se implementó en el diseño experimental las tasas de efectividad se encuentran entre 1% y 2% (vale decir, por cada 100 contactos efectivos, se logran entre 1 y 2 ventas de servicios), ello aplica para el canal Call Center. Nada asegura a priori que los demás canales se comportarán al menos así, por lo que la tercera condición que requiere este test, puede no cumplirse dependiendo de cómo operen los distintos tratamientos experimentales. No obstante lo anterior, el test de diferencia se puede implementar a través de lo siguiente (considerando solo pares de proporciones):

Sean p_1 = tasa de efectividad grupo de tratamiento 1
 p_2 = tasa de efectividad grupo de tratamiento 2

$$H_0: p_1 = p_2$$

$$H_1: p_1 \neq p_2$$

En este caso particular, la hipótesis es a dos colas dado que no se puede afirmar a priori si un tratamiento específico será estrictamente superior o estrictamente inferior a otro, por lo que es preferible para mantener un análisis más general, el aplicar un test de esta clase.

El estadístico de contraste ' z ' que se construye, distribuye como una normal de media 0 y varianza 1. Este se genera a partir de lo siguiente:

- **Pooled sample proportion ' p ' (proporción agregada de la muestra):** Este elemento corresponde a un paso previo para el cálculo del error estándar de la distribución de muestreo (donde n_1 y n_2 corresponden a los tamaños muestrales correspondientes):

$$p = \frac{(p1 * n1 + p2 * n2)}{n1 + n2}$$

- Error estándar 'SE':

$$SE = \sqrt{p * (1 - p) * \left[\left(\frac{1}{n1}\right) + \left(\frac{1}{n2}\right)\right]}$$

- Estadístico de contraste 'z': Con lo anterior, se construye el estadístico de prueba, que se contrasta con una normal (0,1):

$$z = \frac{(p1 - p2)}{SE}$$

Contrastando los resultados que se pueden generar con lo anterior entre todas las muestras disponibles, se logra argumentar estadísticamente respecto a las diferencias existentes entre los tratamientos.

Análisis post-hoc de poder

Adicionalmente, un análisis que adquiere relevancia tras la implementación de los distintos tratamientos del diseño experimental, tiene que ver con la verificación del poder estadístico que alcanzará un determinado resultado.

El análisis de poder a priori que se genera previo a la implementación de un diseño experimental, permite algo muy importante que forma parte integral del diseño y tiene que ver con el tamaño muestral necesario para alcanzar un determinado poder estadístico. En contraste, el análisis post-hoc, permite verificar una vez que el experimento ya ha sido implementado, el poder estadístico real con el que se pueden hacer ciertas afirmaciones respecto a las diferencias entre los resultados de los tratamientos.

Al igual que en el análisis a priori, través del software *G-Power* es sencilla la implementación de este análisis, sencillamente generando como parámetro un test z de proporciones, y estableciendo el test como uno de diferencias de proporciones, y finalmente estableciendo el test como uno post-hoc.

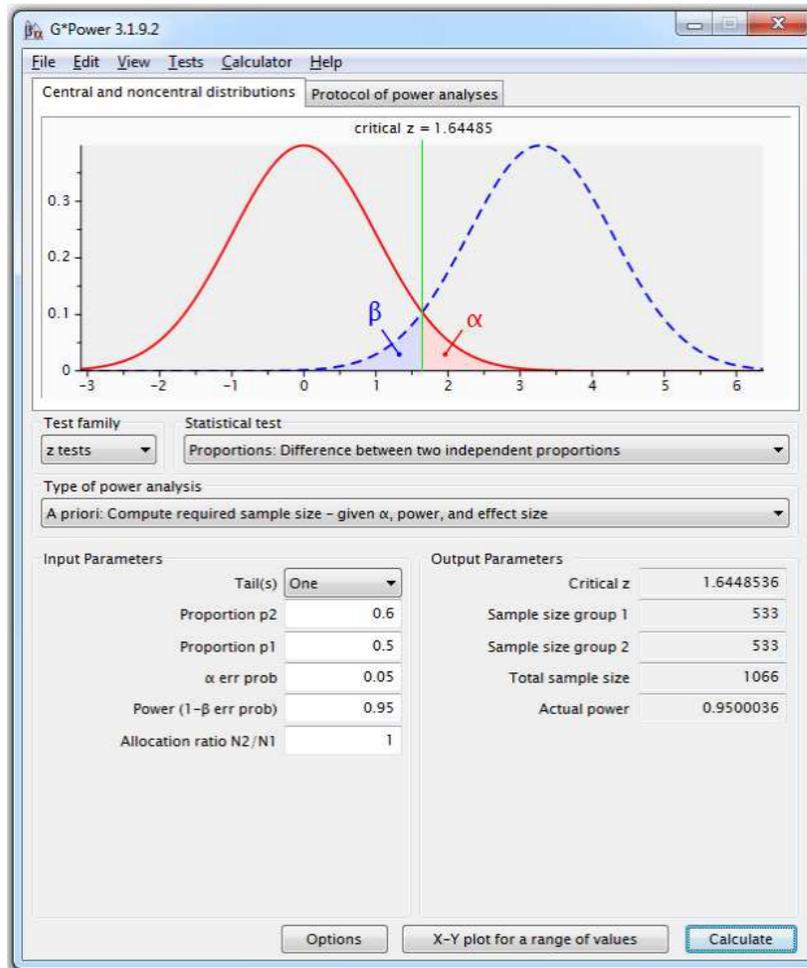


Figura 3.1: Representación gráfica del test post-hoc(muestras de 3.000 individuos)

La figura anterior muestra cómo se visualiza el test, donde la línea roja representa a la proporción 1 mientras que la azul la proporción 2. Las líneas verdes muestran los puntos críticos del test, que dependen de la significatividad que se establezca. El área roja muestra la probabilidad de error tipo I, mientras que el área azul muestra la probabilidad del error tipo II que tiene el test, lo que a la vez muestra en forma indirecta, el poder al que se accede:

$$\text{Statistical Power} = (1 - \beta)$$

Definición de éxito

Para la medición de la efectividad total de la aplicación de cada uno de los tratamientos en testeo, se define que la unidad de medida sean las transacciones en que se genere un ingreso positivo hacia la compañía, es decir **altas de línea**. Se dejarán fuera transacciones del tipo *upselling* (mejoras de plan) y del tipo *sva* (servicios de valor agregado) dado que estas no corresponden a transacciones relevantes a la campaña en estudio.

Además, para asegurar que se capture la mayor cantidad de éxitos posibles como consecuencia de las comunicaciones ejecutadas hacia los clientes, se ha definido considerar transacciones registradas hasta 5 días después de la última instancia de contacto en cada tratamiento. Esta definición proviene netamente de la experiencia y el consejo del área de Desarrollo de mercado de la compañía.

Capítulo 4

Resultados

4.1. Análisis de campañas anteriores

El análisis busca en primer lugar, se busca generar un análisis de efectividad de campañas, desagregando por ciertas variables atinentes a cada campaña, como por ejemplo el call específico que genera la llamada - lo que permite verificar si un call presenta mayor tráfico que otro y la efectividad específica - el tipo de campaña y por último la campaña específica. Esto último es relevante debido a que cada campaña concentra ofertas similares, cuyo desempeño frente al cliente presenta similitudes.

En segundo lugar, con el objetivo de extraer interacciones entre las distintas campañas y las variables transaccionales y demográficas con las que se cuenta, se genera un análisis regresivo, particularmente a través de una regresión logística balanceada, dadas las características de la variable objetivo relevante para este problema particular. Fundamentalmente, el análisis se separa en 2: uno agregado, considerando todas las campañas en forma agrupada, y por otro lado, un análisis desagregado por campaña, considerando las diferencias tanto en el público al cual están dirigidas las campañas como la oferta misma.

4.1.1. Análisis de efectividad

La medición de las tasas de efectividad se hace directamente calculando contactos efectivos dividido por contactos totales, obteniéndose así la proporción de veces que las campañas individuales logran su objetivo final de generar la contratación de un servicio de telefonía. Además, se propone es la desagregación de este análisis a partir de las distintas variables demográficas con las que se cuenta, para lograr esa identificación inicial respecto a cómo distintos clientes responden a distintas campañas, e inclusive lograr verificar si algún Call Center específico se desempeña mejor que los demás (agregando una componente de gestión propia de la compañía al análisis).

Para este análisis, se consideran las comunicaciones generadas el pasado año 2014, entre

los meses de enero y diciembre. Si bien es posible generar un análisis tanto en el canal Call Center, como en el canal E-Mail, en este segundo caso, el análisis no es directo, dado que la utilización de este como herramienta comercial, se ha disminuido en el último tiempo, convirtiéndose más que nada en un canal informativo y para asuntos de fidelización y blindaje de clientes, por lo que el seguimiento de ventas no es directo a través de ese canal. En este análisis, la unidad de éxito corresponde a altas o transacciones de upselling (mejoras de plan) que se generan a través de la campaña.

| Mes | Llamadas realizadas | Ventas | Efectividad |
|--------------|---------------------|----------------|---------------|
| ene-14 | 4,090,846 | 52,494 | 1.283% |
| feb-14 | 3,827,572 | 46,143 | 1.206% |
| mar-14 | 3,747,160 | 69,244 | 1.848% |
| abr-14 | 4,812,280 | 65,180 | 1.354% |
| may-14 | 7,060,975 | 65,216 | 0.924% |
| jun-14 | 8,381,552 | 62,084 | 0.741% |
| jul-14 | 7,247,920 | 62,813 | 0.867% |
| ago-14 | 8,710,665 | 62,251 | 0.715% |
| sep-14 | 7,538,106 | 49,334 | 0.654% |
| oct-14 | 8,455,405 | 59,460 | 0.703% |
| nov-14 | 7,722,033 | 50,305 | 0.651% |
| dic-14 | 7,968,080 | 51,271 | 0.643% |
| Total | 79,562,594 | 695,795 | 0.875% |

Tabla 4.1: Efectividad campañas Outbound (Fuente: Data provista por la empresa)

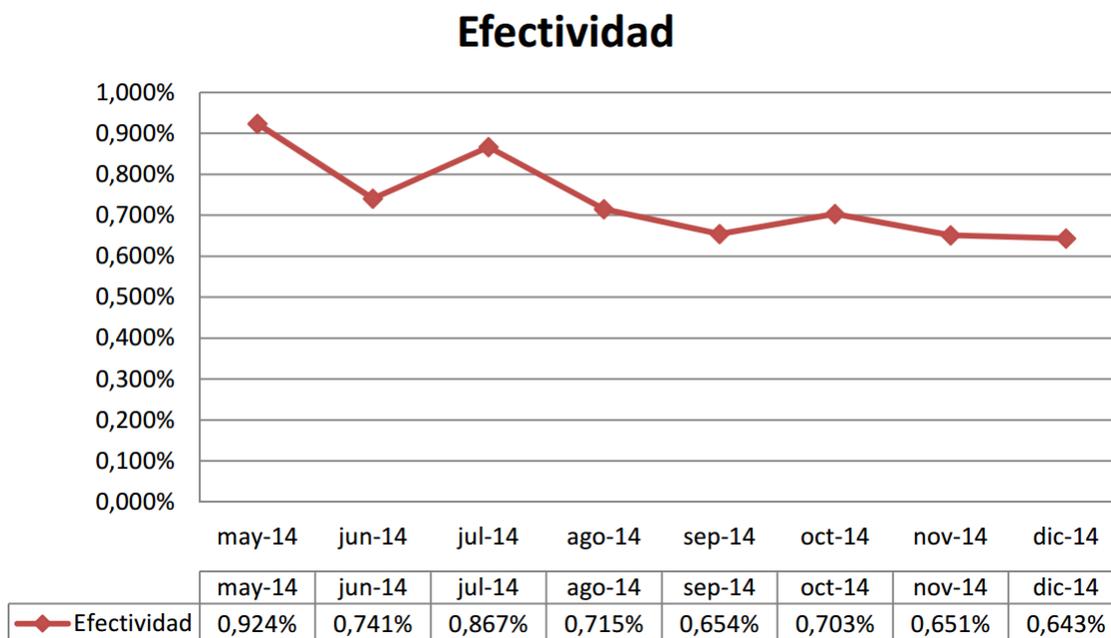


Figura 4.1: Efectividad histórica Call Center

| Mes | Llamadas realizadas NO CQC | Ventas NO CQC | Efectividad NO CQC | Llamadas realizadas CQC | Ventas CQC | Efectividad CQC |
|--------------|----------------------------|----------------|--------------------|-------------------------|----------------|-----------------|
| ene-14 | 2,245,388 | 46,186 | 2.057% | 1,845,458 | 6,308 | 0.342% |
| feb-14 | 2,326,559 | 40,392 | 1.736% | 1,501,013 | 5,751 | 0.383% |
| mar-14 | 2,363,477 | 58,075 | 2.457% | 1,383,683 | 11,169 | 0.807% |
| abr-14 | 2,707,319 | 52,629 | 1.944% | 2,104,961 | 12,551 | 0.596% |
| may-14 | 2,329,343 | 48,174 | 2.068% | 4,731,632 | 17,042 | 0.360% |
| jun-14 | 2,504,801 | 43,340 | 1.730% | 5,876,751 | 18,744 | 0.319% |
| jul-14 | 2,174,459 | 43,086 | 1.981% | 5,073,461 | 19,727 | 0.389% |
| ago-14 | 2,168,419 | 36,225 | 1.671% | 6,542,246 | 26,026 | 0.398% |
| sep-14 | 2,120,972 | 36,646 | 1.728% | 5,417,134 | 12,688 | 0.234% |
| oct-14 | 2,158,621 | 40,398 | 1.871% | 6,296,784 | 19,062 | 0.303% |
| nov-14 | 1,957,406 | 33,766 | 1.725% | 5,764,627 | 16,539 | 0.287% |
| dic-14 | 2,032,263 | 35,958 | 1.769% | 5,935,817 | 15,313 | 0.258% |
| Total | 27,089,027 | 514,875 | 1.901% | 52,473,567 | 180,920 | 0.345% |

Tabla 4.2: Contraste efectividad campañas CQC vs parque (Fuente: Data provista por la empresa)

La separación entre campañas enfocadas en clientes de la compañía y clientes de la competencia se hace evidentemente relevante al observar la tabla anterior. En primer lugar, el tráfico es notoriamente mayor en aquellos clientes que no pertenecen a la compañía, en donde claramente se apuesta por la masividad, al apuntar hacia una captura de clientes desconocidos y sin lograr discriminar en demasía respecto al tipo de persona que se está contactando. Este aumento es muy claro a partir de Mayo, mes en el cual la compañía comenzó una fuerte estrategia tendiente a capturar clientes de la competencia.

Algo negativo de esto es que mes a mes, no es extraño que se contacte al mismo prospecto en forma reiterada - particularmente en las campañas de la competencia, dada la restricción natural de las bases de datos: son limitadas - generándose a la larga un efecto negativo debido a que se puede generar un agobio hacia ellos e impactar sobre la experiencia. Es política reciente de la empresa el evitar esto por medio de ajustes en las políticas más inteligentes de asignación de prospectos.

En las campañas dirigidas hacia clientes pertenecientes a la compañía, las restricciones del negocio son mucho más fuertes y ajustadas a reglas de negocio definidas, que condicionan tanto la cantidad de contactos permitidos como el tipo de ofertas que se pueden ofrecer (definidas en general de acuerdo a objetivos estratégicos y modelos predictivos).

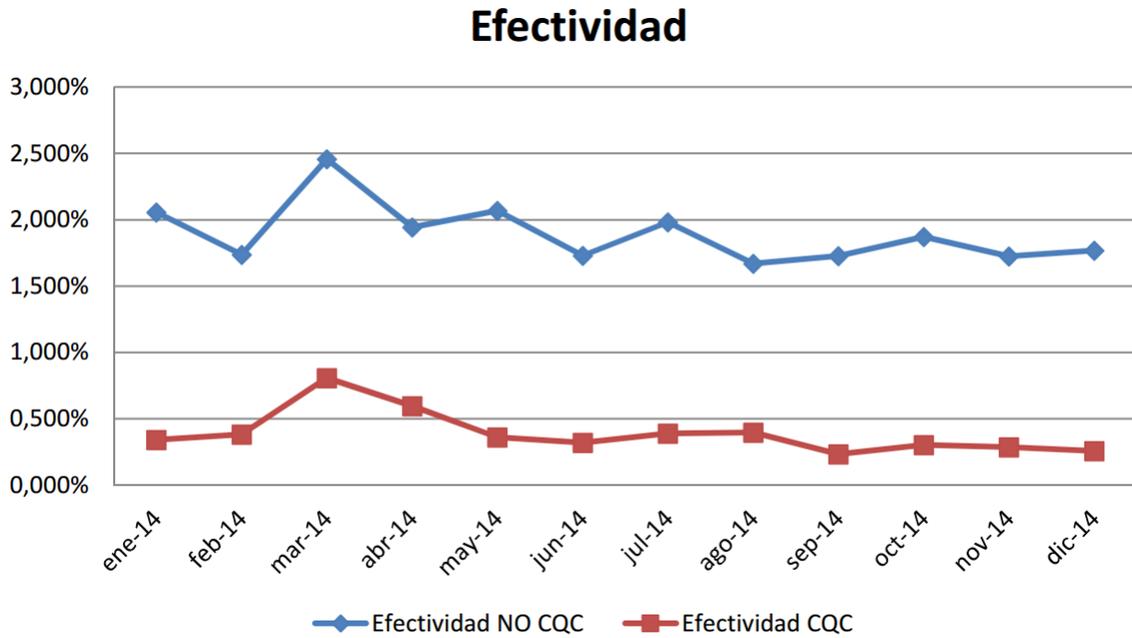


Figura 4.2: Contraste efectividad histórica Call Center CQC vs Parque

4.1.2. Regresión logística

Para evaluar distintos modelos regresivos, se generó un ajuste de las muestras consideradas al considerar una cantidad de respuestas positivas (o aceptaciones de ofertas) similar en términos proporcionales a las respuestas negativas (rechazos de ofertas), desde un 50 % de proporción hasta un 80 %. Con esto, se generaron los siguientes resultados (se ha seleccionado un corte que mantiene un error de predicción equilibrado entre las dos clases):

Modelo agregado

- 50% de porcentaje de respuestas negativas

```
glm(formula = Venta ~ antiguedad_plan + antiguedad_linea + CAMBIOS_PLANES +
gse + edad + equipo_antiguedad + MARCA_PLAN + multimedia +
REGION + aumento_LL_ENTRADA_OM + sd_LTE_3G + aumento_LL_ENTRADA_RF +
disminucion_LL_ENTRADA_CLARO + disminucion_LL_SALIDA_CLARO +
aumento_LL_ENTRADA_ENTEL + aumento_LL_ENTRADA_ON + aumento_LL_SALIDA_ON +
disminucion_LL_SALIDA_ON, family = binomial(), data = data_paq)
```

| | | CORTE | 0.54 | |
|--------|--|--------|-------|-----------|
| | | Obs 0 | Obs 1 | Pred err |
| Pred 0 | | 415229 | 8825 | 0.3497630 |
| Pred 1 | | 223352 | 14660 | 0.3757718 |

| | |
|------------------------|-------------|
| Accuracy | 0.649314419 |
| Error de clasificacion | 0.350685581 |
| Sensitivity (TPR) | 0.624228231 |
| FPR | 0.34976299 |
| Specificity | 0.65023701 |
| Precision | 0.061593533 |
| Prevalence | 0.035472294 |

Tabla 4.3: Modelo agregado (muestras balanceadas 50-50)

- 60% de porcentaje de respuestas negativas

```
glm(formula = Venta ~ antiguedad_plan + gse + antiguedad_linea +
  CAMBIOS_PLANES + equipo_antiguedad + edad + MARCA_PLAN +
  REGION + multimedia + sd_LTE_3G + disminucion_LL_ENTRADA_CLARO +
  aumento_LL_ENTRADA_RF + sd_LL_SALIDA_ENTEL + sd_LL_ENTRADA_CLARO +
  sd_SMS_SAL + disminucion_LL_ENTRADA_ON + sd_SESIONES_PPU +
  aumento_LL_ENTRADA_OM, family = binomial(), data = data_paq)
```

| | CORTE | 0.44 | |
|--------|--------|-------|-----------|
| | Obs 0 | Obs 1 | Pred err |
| Pred 0 | 409280 | 8600 | 0.3590790 |
| Pred 1 | 229301 | 14885 | 0.3661912 |

| | |
|------------------------|-------------|
| Accuracy | 0.640668755 |
| Error de clasificacion | 0.359331245 |
| Sensitivity (TPR) | 0.633808814 |
| FPR | 0.359078958 |
| Specificity | 0.640921042 |
| Precision | 0.060957631 |
| Prevalence | 0.035472294 |

Tabla 4.4: Modelo agregado (muestras balanceadas 60-40)

- 70% de porcentaje de respuestas negativas

```
glm(formula = Venta ~ antiguedad_plan + antiguedad_linea + CAMBIOS_PLANES +
gse + equipo_antiguedad + edad + REGION + MARCA_PLAN + multimedia +
sd_LTE_3G + mean_PAGO + aumento_LL_ENTRADA_RF + disminucion_LL_ENTRADA_CLARO +
puntos_acumulado + disminucion_LL_SALIDA_CLARO + disminucion_LL_SALIDA_RF +
disminucion_GPRS_2G + sd_LL_ENTRADA_CLARO + sd_LL_SALIDA_ENTEL +
sd_LL_ENTRADA_ENTEL + sd_q_dias_naveg, family = binomial(),
data = data_paq)
```

| | CORTE | 0.335 | |
|--------|--------|-------|-----------|
| | Obs 0 | Obs 1 | Pred err |
| Pred 0 | 401369 | 8398 | 0.3714674 |
| Pred 1 | 237212 | 15087 | 0.3575900 |

| | |
|------------------------|-------------|
| Accuracy | 0.629024901 |
| Error de clasificacion | 0.370975099 |
| Sensitivity (TPR) | 0.642410049 |
| FPR | 0.371467363 |
| Specificity | 0.628532637 |
| Precision | 0.059798097 |
| Prevalence | 0.035472294 |

Tabla 4.5: Modelo agregado (muestras balanceadas 70-30)

- 80% de porcentaje de respuestas negativas

```
glm(formula = Venta ~ antiguedad_plan + antiguedad_linea + CAMBIOS_PLANES +
gse + equipo_antiguedad + edad + MARCA_PLAN + REGION + multimedia +
disminucion_LL_ENTRADA_CLARO + sd_IMPORTE_CI + sd_LL_SALIDA_CLARO +
disminucion_LL_ENTRADA_ON + sd_LL_SALIDA_RF + mean_LL_ENTRADA_ON +
sd_SMS_ENT + sd_LTE_3G + disminucion_MM_ENTRADA + disminucion_LL_SALIDA_ON +
disminucion_DEU_INICIAL + aumento_LL_ENTRADA_OM + sd_IMPORTE_VITE +
sd_traf_datos, family = binomial(), data = data_paq)
```

| | | CORTE | 0.235 | | |
|--------|--|--------|-------|-----------|--|
| | | Obs 0 | Obs 1 | Pred err | |
| Pred 0 | | 419061 | 8894 | 0.3437622 | |
| Pred 1 | | 219520 | 14591 | 0.3787098 | |

| | |
|------------------------|-------------|
| Accuracy | 0.654998142 |
| Error de clasificacion | 0.345001858 |
| Sensitivity (TPR) | 0.621290185 |
| FPR | 0.343762185 |
| Specificity | 0.656237815 |
| Precision | 0.062325136 |
| Prevalence | 0.035472294 |

Tabla 4.6: Modelo agregado (muestras balanceadas 80-20)

Como se observa de los resultados anteriores, para los modelos agregados probados para distintos cortes para asignar como aceptador o rechazador de una oferta comercial a los clientes evaluados, no se presentan diferencias significativas en los resultados obtenidos. En todos ellos, el accuracy de la asignación se encuentra en torno al 65%, lo que si bien no es una buena asignación, se valida la identificación de relaciones entre las variables explicativas y variable objetivo.

Se observa además que al cambiar las muestras modelo a modelo, asumiendo que no existe sesgo (al ser aleatoria la toma de muestras), y al cambiar por ende la proporción de individuos de una clase versus otra, las variables que el proceso stepwise identifica como relevantes, no es constante. Esto hace inferir que el comportamiento de los individuos de la clase que no predomina (aquellos que no compran), influye en la definición del modelo.

Modelo desagregado por campaña

Generando una mirada por campaña, sobre la data sin balancear, se obtienen los siguientes resultados:

- Campaña "2DA LN PERS"

| | CORTE | 0.0115 | |
|--------|-------|--------|-----------|
| | Obs 0 | Obs 1 | Pred err |
| Pred 0 | 98006 | 766 | 0.4312756 |
| Pred 1 | 74320 | 1279 | 0.3745721 |

| | |
|------------------------|-------------|
| Accuracy | 0.569389405 |
| Error de clasificacion | 0.430610595 |
| Sensitivity (TPR) | 0.625427873 |
| FPR | 0.431275606 |
| Specificity | 0.568724394 |
| Precision | 0.016918213 |
| Prevalence | 0.011727868 |
| AUC | 0.6392 |

Tabla 4.7: Modelo campaña 2DA LN PERS

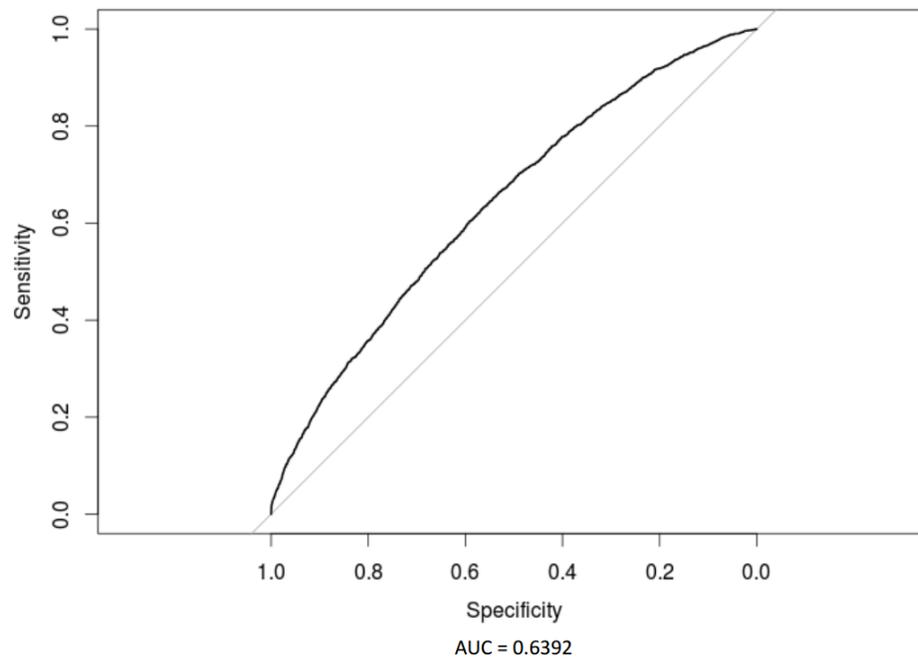


Figura 4.3: Modelo campaña 2DA LN PERS

- Campaña "Bolsas Min/SMS"

| | CORTE | 0.023 | |
|--------|-------|-------|-----------|
| | Obs 0 | Obs 1 | Pred err |
| Pred 0 | 15592 | 330 | 0.3770924 |
| Pred 1 | 9439 | 544 | 0.3775744 |

| | |
|------------------------|-------------|
| Accuracy | 0.622891334 |
| Error de clasificacion | 0.377108666 |
| Sensitivity (TPR) | 0.622425629 |
| FPR | 0.377092405 |
| Specificity | 0.622907595 |
| Precision | 0.054492637 |
| Prevalence | 0.03373866 |
| AUC | 0.67 |

Tabla 4.8: Modelo campaña Bolsas Min/SMS

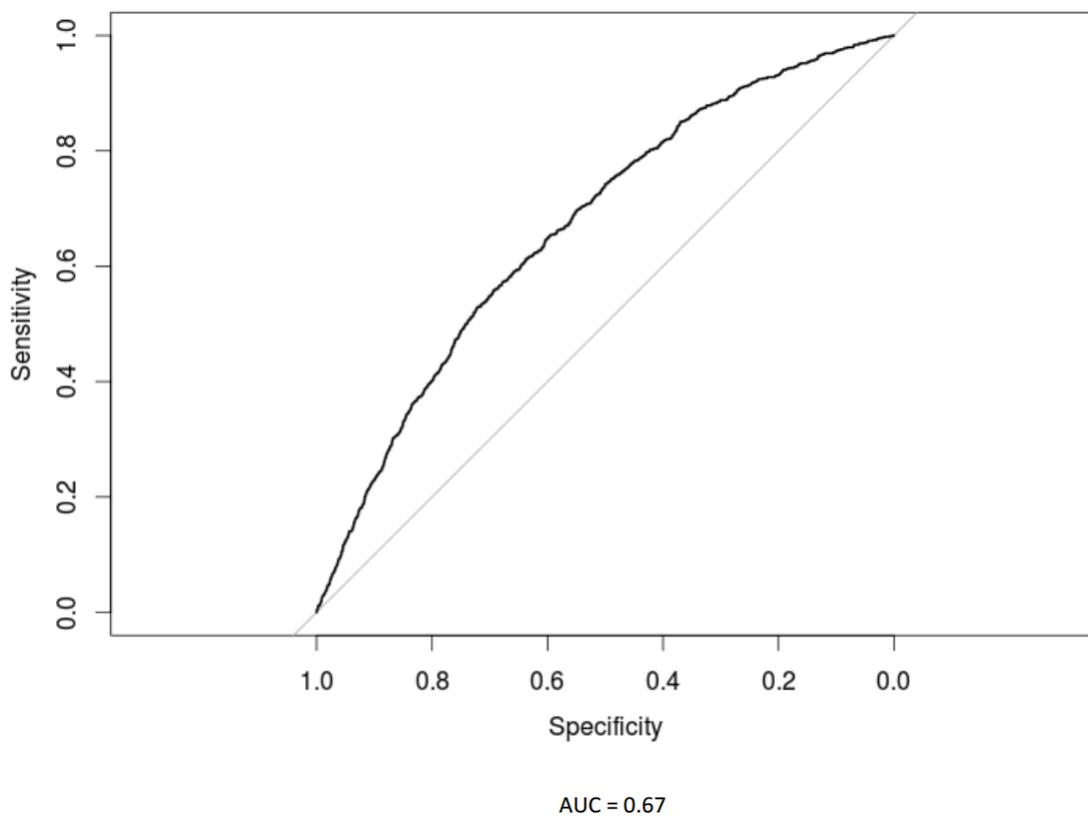


Figura 4.4: Modelo campaña Bolsas Min/SMS

- Campaña "Bolsas Nem Personas"

| | CORTE | 0.028 | |
|--------|-------|-------|-----------|
| | Obs 0 | Obs 1 | Pred err |
| Pred 0 | 24079 | 953 | 0.4400493 |
| Pred 1 | 18923 | 1517 | 0.3858300 |

| | |
|------------------------|-------------|
| Accuracy | 0.562895848 |
| Error de clasificacion | 0.437104152 |
| Sensitivity (TPR) | 0.61417004 |
| FPR | 0.4400493 |
| Specificity | 0.5599507 |
| Precision | 0.074217221 |
| Prevalence | 0.054319141 |

Tabla 4.9: Modelo campaña Bolsas Nem Personas

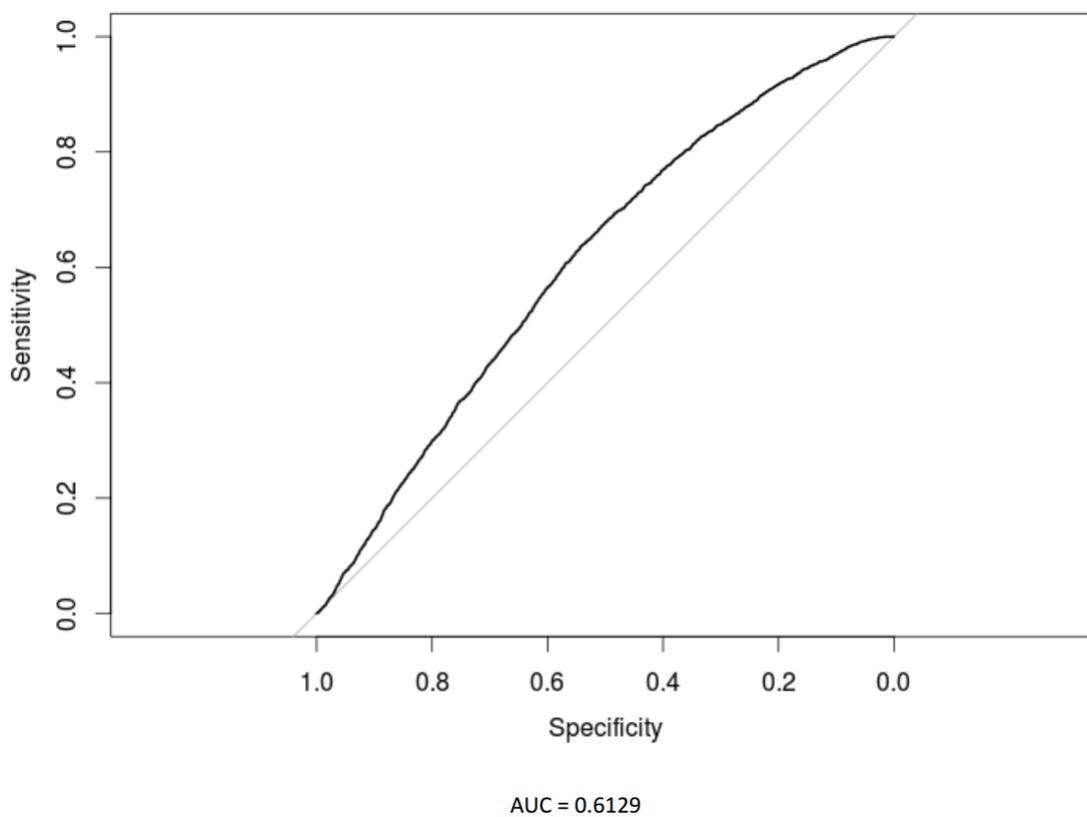


Figura 4.5: Modelo campaña Bolsas Nem Personas

- Campaña “Cross Full Bam Venta 4G”

| | CORTE | 0.028 | |
|--------|-------|-------|-----------|
| | Obs 0 | Obs 1 | Pred err |
| Pred 0 | 32161 | 200 | 0.4408533 |
| Pred 1 | 25357 | 284 | 0.4132231 |

| | |
|------------------------|-------------|
| Accuracy | 0.559377263 |
| Error de clasificacion | 0.440622737 |
| Sensitivity (TPR) | 0.58677686 |
| FPR | 0.440853298 |
| Specificity | 0.559146702 |
| Precision | 0.011076011 |
| Prevalence | 0.00834454 |

Tabla 4.10: Modelo campaña Cross Full Bam Venta 4G

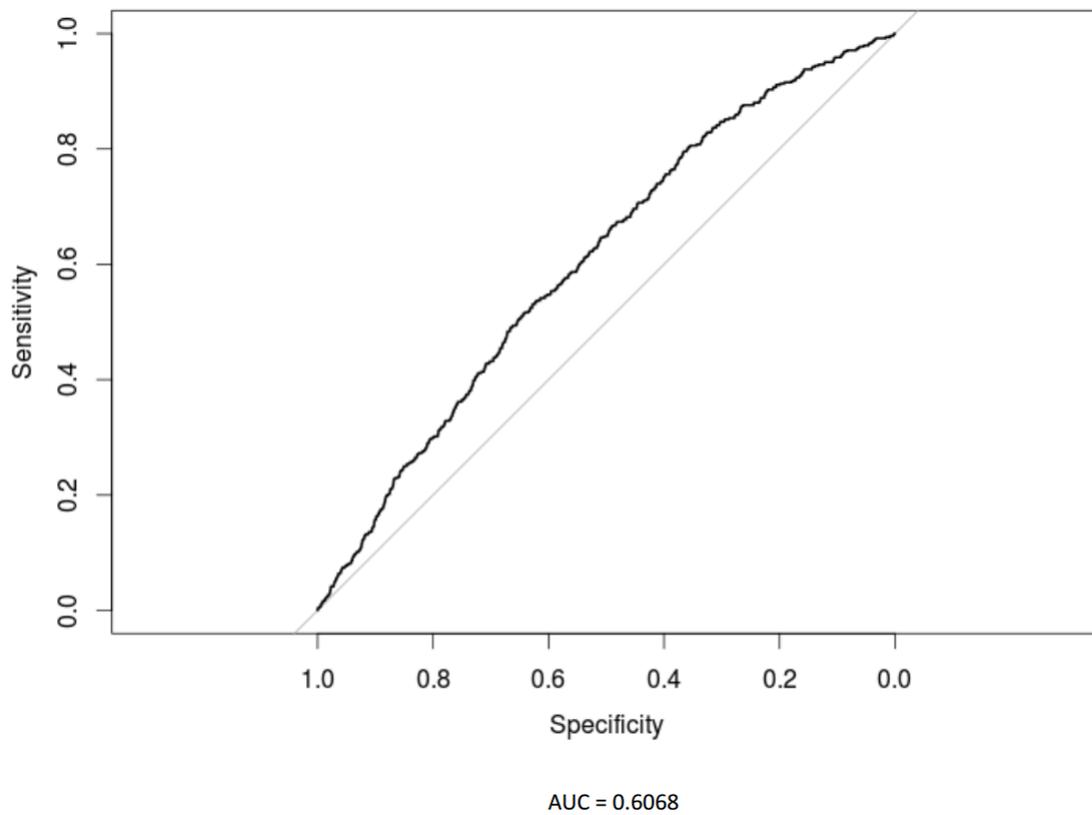


Figura 4.6: Modelo campaña Cross Full Bam Venta 4G

- Campaña “Up Nem Personas”

| | CORTE | 0.038 | |
|--------|-------|-------|-----------|
| | Obs 0 | Obs 1 | Pred err |
| Pred 0 | 6197 | 143 | 0.4661900 |
| Pred 1 | 5412 | 214 | 0.4005602 |

| | |
|------------------------|-------------|
| Accuracy | 0.535768009 |
| Error de clasificacion | 0.464231991 |
| Sensitivity (TPR) | 0.599439776 |
| FPR | 0.466190025 |
| Specificity | 0.533809975 |
| Precision | 0.038037682 |
| Prevalence | 0.029834531 |

Tabla 4.11: Modelo campaña Up Nem Personas

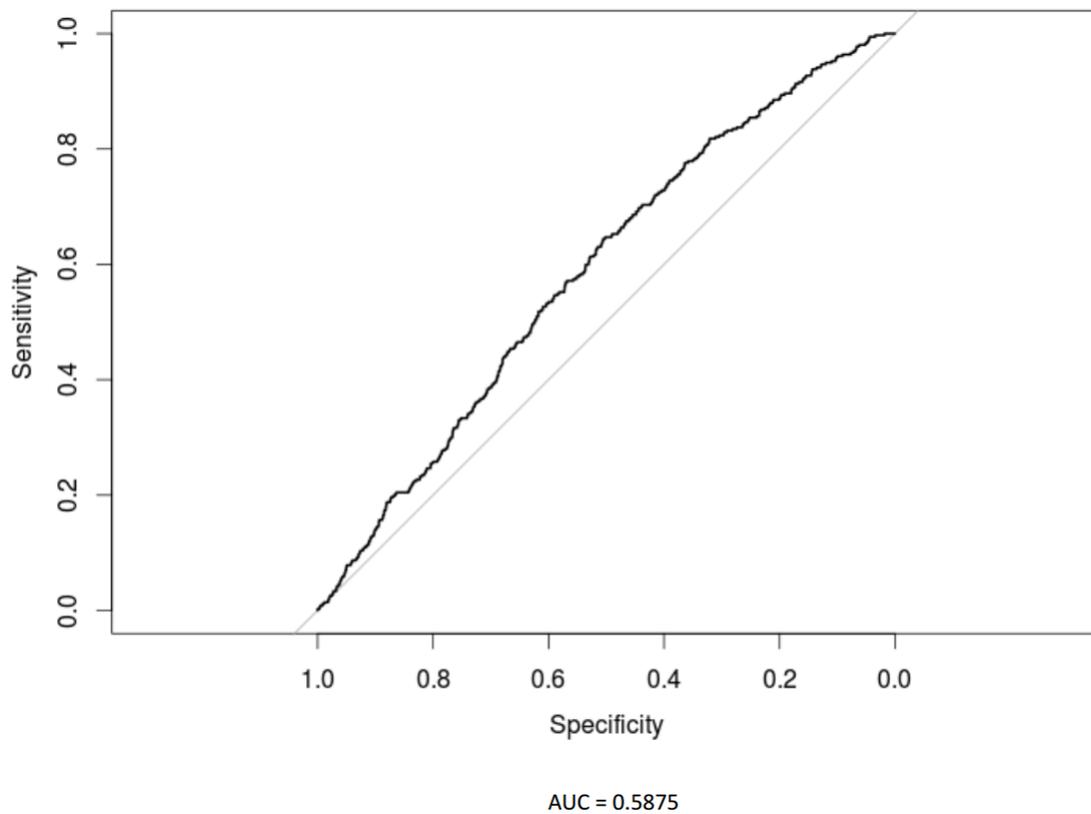


Figura 4.7: Modelo campaña Up Nem Personas

- Campaña “Up Voz Móvil”

| | CORTE | 0.07 | |
|--------|--------|-------|-----------|
| | Obs 0 | Obs 1 | Pred err |
| Pred 0 | 161995 | 5042 | 0.2977988 |
| Pred 1 | 68701 | 12190 | 0.2925952 |

| | |
|------------------------|-------------|
| Accuracy | 0.702562841 |
| Error de clasificacion | 0.297437159 |
| Sensitivity (TPR) | 0.707404828 |
| FPR | 0.297798835 |
| Specificity | 0.702201165 |
| Precision | 0.150696616 |
| Prevalence | 0.06950405 |

Tabla 4.12: Modelo campaña Up Voz Móvil

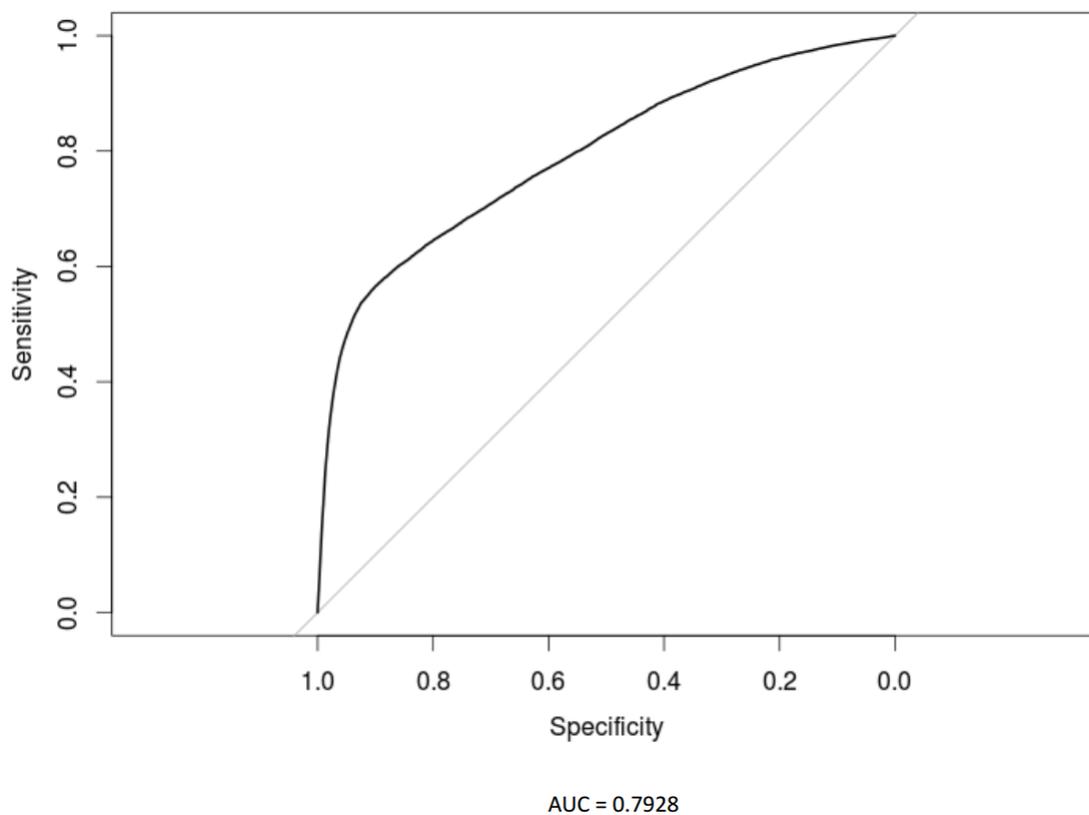


Figura 4.8: Modelo campaña Up Voz Móvil

Las variables que muestran un mayor efecto sobre la variable objetivo que se ha considerado son principalmente variables que tienen que ver con la cantidad de meses que han pasado desde que el cliente cambió su equipo, dio de alta su línea o cambió su plan. Esto tiene sentido en esta clase de mercados, dado que las transacciones se dan con poca frecuencia (los contratos móviles duran inicialmente 18 meses), por lo que es esperable que a medida que pase más tiempo desde la última transacción, la probabilidad de que los clientes adquieran un nuevo producto se vaya incrementando.

Al examinar los resultados de los modelos generados por campaña, en general no se observan cosas muy interesantes, salvo en la campaña de "Up Selling de Voz Móvil", en donde la predicción genera un área bajo la curva ROC de casi 0.8, un valor bastante alto, identificando a dicho modelo como uno con un alto poder predictivo (al menos superior a los demás). Esto denota además que para esta campaña particular, las variables seleccionadas aportan a la capacidad predictora del modelo. Analizando el detalle de este modelo, se observa lo siguiente:

```
glm(formula = Venta ~ antiguedad_plan + equipo_antiguedad + multimedia +
    gse + RECAMBIOS + CAMBIOS_PLANES + REGION + antiguedad_linea +
    puntos_acumulado + sd_SMS_SAL + disminucion_LTE_4G + sd_LL_SALIDA_RF +
    mean_LL_SALIDA_ON + disminucion_LL_ENTRADA_ON + aumento_LL_ENTRADA_OM +
    disminucion_LL_SALIDA_ON + aumento_LL_ENTRADA_ENTEL + aumento_MM_ENTRADA,
    family = binomial(), data = data_sb_cam7)
```

Un detalle importante que salta a la vista es que el modelo favorece variables en las que existe una disminución en variables de tráfico de salida, como la de navegación 4G (*disminucionLTE4G*) y llamadas de salida (*disminucionllsalidaON*), como así también ciertas variables de aumento de tráfico de entrada.

Si se analiza la definición del modelo, es posible detallar de manera evidente el efecto de variables específicas sobre la variable objetivo y cuantificarlo. La ejecución en el software de análisis **R** entrega el siguiente resultado:

```

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.787e+00  9.869e-02 -18.107 < 2e-16 ***
antiguedad_plan  -1.459e-01  1.898e-03 -76.856 < 2e-16 ***
equipo_antiguedad  1.512e-02  6.788e-04  22.273 < 2e-16 ***
multimedia        4.634e-01  2.349e-02  19.724 < 2e-16 ***
gseABC1          -4.885e-01  7.837e-02  -6.233 4.58e-10 ***
gseC2            -2.898e-01  7.421e-02  -3.906 9.38e-05 ***
gseC3            -2.223e-01  7.397e-02  -3.005 0.002652 **
gseD             -1.040e-01  7.372e-02  -1.411 0.158291
gseE             -1.655e-01  1.041e-01  -1.590 0.111757
RECAMBIOS        -8.242e-02  8.189e-03 -10.065 < 2e-16 ***
CAMBIOS_PLANES   7.612e-02  4.856e-03  15.674 < 2e-16 ***
REGION10         9.687e-02  7.127e-02   1.359 0.174081
REGION11        -1.380e-01  1.454e-01  -0.949 0.342730
REGION12        -9.123e-02  8.420e-02  -1.084 0.278559
REGION14         6.376e-02  8.748e-02   0.729 0.466044
REGION15        -7.339e-02  9.102e-02  -0.806 0.420027
REGION2         -7.117e-02  7.065e-02  -1.007 0.313775
REGION3         -5.899e-02  7.709e-02  -0.765 0.444088
REGION4         -8.658e-02  6.878e-02  -1.259 0.208069
REGION5         -2.966e-02  6.374e-02  -0.465 0.641712
REGION6         -2.135e-01  6.780e-02  -3.150 0.001635 **
REGION7         5.428e-02  7.479e-02   0.726 0.468044
REGION8         5.377e-02  6.613e-02   0.813 0.416209
REGION9         1.082e-01  6.991e-02   1.548 0.121518
REGIONRM        -3.133e-02  6.017e-02  -0.521 0.602526
antiguedad_linea -1.581e-03  4.449e-04  -3.554 0.000379 ***
puntos_acumulado -4.895e-06  1.394e-06  -3.511 0.000447 ***
sd_SMS_SAL       7.651e-04  2.543e-04   3.009 0.002619 **
disminucion_LTE_4G -1.331e-01  5.212e-02  -2.553 0.010679 *
sd_LL_SALIDA_RF  3.696e-03  1.290e-03   2.866 0.004154 **
mean_LL_SALIDA_ON -1.911e-04  7.067e-05  -2.704 0.006852 **
disminucion_LL_ENTRADA_ON -6.071e-02  2.827e-02  -2.147 0.031786 *
aumento_LL_ENTRADA_OM -5.004e-02  2.317e-02  -2.160 0.030802 *
disminucion_LL_SALIDA_ON -5.169e-02  2.906e-02  -1.778 0.075330 .
aumento_LL_ENTRADA_ENTEL 3.644e-02  2.193e-02   1.662 0.096539 .
aumento_MM_ENTRADA -3.101e-02  1.865e-02  -1.662 0.096414 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Se observa que variables de tráfico, como la marca de disminución de tráfico 4G o la disminución de llamadas de salida, tienen un parámetro asociado negativo, lo que tiene sentido, dado que ello representaría una tendencia al alza en el consumo de los clientes, lo que motiva la adquisición de un plan de up selling, es decir, de mejoramiento del plan actual del cliente.

Si bien la mayoría de los modelos evaluados no se desempeñan de buena manera en forma clara, el hecho de que la campaña de up selling de voz sí lo hace indica que el análisis construido se puede complementar y mejorar a través de la incorporación de nuevos atributos no considerados, como el tráfico de productos fijos o variables de geolocalización (para intuir

los tipos de requerimientos y consumos más específicos de los prospectos) o bien por medio de la aplicación de otros modelamientos más complejos.

Si bien lo anterior no da cuenta de aplicación de campañas en un ambiente multicanal, el estudio de los resultados observados de las campañas permite identificar correlaciones y causalidades existentes entre las variables consideradas en los distintos modelos y la variable objetivo considerada. Dichos efectos podrían verse amplificados o no al implementar campañas por medio de las distintas combinaciones de canales a testear.

4.2. Diseño experimental

4.2.1. Efectividad Call

Una primera mirada que se puede explorar de los resultados del diseño experimental aplicado, se relaciona con la extracción de las tasas de efectividad generadas por efecto de las llamadas outbound. Estas ventas se registran inmediatamente cuando se genera una llamada con los clientes, por lo que este análisis inicial no captura todas las ventas que cada tratamiento está generando. No obstante, permite darle una primera mirada a como se desempeñan los distintos tratamientos y como los contactos previos apoyan una comunicación futura por medio del Call Center. La tabla a continuación presenta estos resultados, para solo aquellos tratamientos en que el Call se utiliza en alguna instancia de contacto, separando cada una de las instancias de contacto (solo se presenta la efectividad de aquellas instancias de contacto en que se utiliza el Call Center como canal).

| | | Efectividad base recorrida Call (IC1) | Efectividad contactados Call (IC1) | Efectividad contactados 2 Call (IC1) | Efectividad base recorrida Call (IC2) | Efectividad contactados Call (IC2) | Efectividad contactados 2 Call (IC2) | Efectividad base recorrida Call (IC3) | Efectividad contactados Call (IC3) | Efectividad contactados 2 Call (IC3) |
|----------------|--------|---------------------------------------|------------------------------------|--------------------------------------|---------------------------------------|------------------------------------|--------------------------------------|---------------------------------------|------------------------------------|--------------------------------------|
| | Codigo | | | | | | | | | |
| Tratamiento 4 | CC | 0.47% | 0.80% | 1.15% | 0.99% | 3.53% | 3.67% | - | - | - |
| Tratamiento 5 | C | 0.67% | 1.19% | 1.59% | - | - | - | - | - | - |
| Tratamiento 6 | EC | - | - | - | 0.40% | 0.99% | 1.03% | - | - | - |
| Tratamiento 7 | CE | 0.63% | 1.41% | 1.46% | - | - | - | - | - | - |
| Tratamiento 8 | SC | - | - | - | 0.37% | 0.78% | 1.07% | - | - | - |
| Tratamiento 9 | CS | 0.60% | 1.42% | 1.48% | - | - | - | - | - | - |
| Tratamiento 12 | SEC | - | - | - | - | - | - | 0.50% | 2.06% | 2.88% |
| Tratamiento 13 | ESC | - | - | - | - | - | - | 0.33% | 1.37% | 7.30% |
| Tratamiento 14 | SSC | - | - | - | - | - | - | 0.33% | 1.00% | 1.31% |
| Tratamiento 15 | EEC | - | - | - | - | - | - | 0.23% | 0.67% | 0.88% |
| Tratamiento 16 | SCC | - | - | - | 0.20% | 0.73% | 0.95% | 0.26% | 1.20% | 1.25% |
| Tratamiento 17 | ECC | - | - | - | 0.30% | 0.90% | 1.17% | 0.28% | 0.90% | 1.02% |
| Tratamiento 20 | ECE | - | - | - | 0.37% | 1.09% | 1.41% | - | - | - |
| Tratamiento 21 | CES | 0.43% | 1.26% | 1.35% | - | - | - | - | - | - |
| Tratamiento 22 | CSE | 0.57% | 1.40% | 1.49% | - | - | - | - | - | - |
| Tratamiento 23 | CEE | 0.43% | 1.13% | 1.19% | - | - | - | - | - | - |
| Tratamiento 24 | CEC | 0.37% | 1.79% | 2.23% | - | - | - | 0.30% | 1.79% | 2.30% |
| Tratamiento 25 | CCE | 0.33% | 1.81% | 2.02% | 0.16% | 0.98% | 1.21% | - | - | - |

Tabla 4.13: Efectividad tratamientos Call

Se han determinado 3 tipos de efectividades para hacer estas mediciones:

- **Efectividad base recorrida:** Efectividad de las ventas generadas sobre la base total considerada. En este caso, cada tratamiento cuenta con una base de 3.000 individuos, por lo que se mantiene constante ese término en el denominador del cálculo de la efectividad.

- **Efectividad contactados:** Efectividad de las ventas generadas sobre solo los individuos que los calls que gestionaron los envíos, reportaron como contactados. De esta forma, se asegura que la decisión de compra o de rechazo se deben a la oferta misma, y no a que por ejemplo el cliente no quiso responder a la llamada.
- **Efectividad contactados 2:** Igual que el indicador anterior, pero considerando solo aquellas llamadas cuya duración es superior a 15 segundos. Esto busca excluir aquellos casos en que la comunicación con el cliente es tan corta que no es posible entregar la oferta que se busca.

No es extraño que las efectividades de los tratamientos 4, 7, 9, 21, 22, 23, 2, 21, 22, 23, 24 y 25 sean más bien similares, dado que las 3 corresponden a registros en que el cliente solo ha sido contactado por medio del Call Center. Es algo extraño quizás que el tratamiento 4 se desempeñe un poco peor, pero esto se puede atribuir a factores no observables.

En las segundas instancia de contacto, se observan cosas más interesantes. La segunda instancia del tratamiento 4 por ejemplo muestra una alta tasa de efectividad: 3.53%, lo que a priori indica que un contacto previo por medio del mismo Call Center, apoya un segundo contacto por Call Center, apoyando las hipótesis de multicontacto que se postulan en secciones anteriores.

La segunda instancia de los tratamientos 8 y 16 también son idénticos respecto a los contactos recibidos por el cliente: una llamada antecedida de un mensaje de texto. Esto se valida en los resultados, donde se muestra unas tasas de efectividad de contactados en torno al 0.75%. Esto a priori se explica por 2 motivos: el envío de SMS no apoya un segundo contacto por Call, sino que lo empeora; o bien, el envío de SMS "adelanta" transacciones que se generarían por el Call Center, concretándose ventas en el intervalo entre el envío de SMS y la llamada Outbound posterior. Esta última situación se analizará en la sección *4.2.3 Efectividad Total*. En los tratamientos 6, 17 y 20 se observa una situación similar pero en estos casos, la efectividad de contactados se encuentra en torno al 1,00% y la primera instancia de contacto corresponde a un E-Mail y las hipótesis que se pueden plantear son idénticas a lo anterior.

Analizando los tratamientos en que el Call aparece como tercera instancia de contacto, la situación es menos constante. Es importante destacar en primer lugar el tratamiento 12, el cual posee una efectividad de contactados de 2,06%, que si bien no es excesivamente alto, supera a todas las demás tasas de efectividad de esta instancia de contacto. Es importante notar que este tratamiento es particularmente relevante, dado que es el que representa de mejor manera la lógica de crecimiento en la informatividad y accionabilidad del canal de contacto, es decir, se comienza por el menos informativo (SMS) y se va ascendiendo (E-Mail) hasta llegar al más informativo y accionable (Call Center). No obstante, en este caso, para la tercera instancia de contacto, la cantidad de prospectos contactados efectivamente se ha visto reducida: solo 729 contactos para este tratamiento, por lo que si bien se gana efectividad, se pierde contactabilidad de la base.

4.2.2. Efectividad Email

Analizando el desempeño del canal E-Mail, en forma análoga al Call, se pueden cuantificar 2 medidas principales:

- Tasa de apertura de emails.
- Tasa de clics únicos.

El primero de estos ratios tiene relación directamente con el atractivo inicial que puede gatillar en el prospecto el correo electrónico recibido, es decir, en donde la oferta solo se resume por medio del encabezado del mensaje a través de una frase corta. Por su parte, la tasa de clics únicos (sobre correos abiertos) ya da cuenta del interés de los clientes en la oferta misma, dado que en estos casos, el cliente ya ha accedido a un paso más allá y ha solicitado más información de la oferta al acceder a uno de los portales que se redirigen desde el E-Mail. Las tasas observadas en los distintos tratamientos son las siguientes:

| | | Tasa de apertura (IC1) | Tasa de clics únicos (IC1) | Tasa de apertura (IC2) | Tasa de clics únicos (IC2) | Tasa de apertura (IC3) | Tasa de clics únicos (IC3) |
|----------------|--------|------------------------|----------------------------|------------------------|----------------------------|------------------------|----------------------------|
| | Código | | | | | | |
| Tratamiento 2 | EE | 32,37% | 9,35% | 33,86% | 7% | - | - |
| Tratamiento 6 | EC | 38,03% | 8,85% | - | - | - | - |
| Tratamiento 7 | CE | - | - | 40,54% | 7,54% | - | - |
| Tratamiento 10 | SE | - | - | 40,79% | 8,24% | - | - |
| Tratamiento 11 | ES | 38,53% | 8,69% | - | - | - | - |
| Tratamiento 12 | SEC | - | - | 39,85% | 7,43% | - | - |
| Tratamiento 13 | ESC | 40,27% | 7,65% | - | - | - | - |
| Tratamiento 15 | EEC | 40,60% | 7,79% | 34,93% | 6,93% | - | - |
| Tratamiento 17 | ECC | 40,39% | 9,53% | - | - | - | - |
| Tratamiento 18 | SSE | - | - | - | - | 35,84% | 6,21% |
| Tratamiento 19 | EES | 40,91% | 7,60% | 36,31% | 6,88% | - | - |
| Tratamiento 20 | ECE | 38,01% | 7,05% | - | - | 34,41% | 6,91% |
| Tratamiento 21 | CES | - | - | 37,98% | 4,38% | - | - |
| Tratamiento 22 | CSE | - | - | - | - | 41,48% | 6,9% |
| Tratamiento 23 | CEE | - | - | 36% | 2,63% | 39,5% | 7,99% |
| Tratamiento 24 | CEC | - | - | 39,07% | 3,49% | - | - |
| Tratamiento 25 | CCE | - | - | - | - | 42,96% | 7,85% |

Tabla 4.14: Efectividad tratamientos E-Mail

| | | Abiertos (IC1) | Enviados (IC1) | Clics únicos (IC1) | Abiertos (IC2) | Enviados (IC2) | Clics únicos (IC2) | Abiertos (IC3) | Enviados (IC3) | Clics únicos (IC3) |
|----------------|--------|----------------|----------------|--------------------|----------------|----------------|--------------------|----------------|----------------|--------------------|
| | Código | | | | | | | | | |
| Tratamiento 2 | EE | 834 | 2235 | 78 | 771 | 2277 | 54 | - | - | - |
| Tratamiento 6 | EC | 847 | 2227 | 75 | - | - | - | - | - | - |
| Tratamiento 7 | CE | - | - | - | 915 | 2257 | 69 | - | - | - |
| Tratamiento 10 | SE | - | - | - | 959 | 2351 | 79 | - | - | - |
| Tratamiento 11 | ES | 863 | 2240 | 75 | - | - | - | - | - | - |
| Tratamiento 12 | SEC | - | - | - | 942 | 2364 | 70 | - | - | - |
| Tratamiento 13 | ESC | 902 | 2240 | 69 | - | - | - | - | - | - |
| Tratamiento 15 | EEC | 963 | 2372 | 75 | 822 | 2351 | 57 | - | - | - |
| Tratamiento 17 | ECC | 923 | 2285 | 88 | - | - | - | - | - | - |
| Tratamiento 18 | SSE | - | - | - | - | - | - | 982 | 2740 | 61 |
| Tratamiento 19 | EES | 947 | 2315 | 72 | 857 | 2360 | 59 | - | - | - |
| Tratamiento 20 | ECE | 865 | 2276 | 61 | - | - | - | 970 | 2819 | 67 |
| Tratamiento 21 | CES | - | - | - | 1096 | 2886 | 48 | - | - | - |
| Tratamiento 22 | CSE | - | - | - | - | - | - | 1188 | 2864 | 82 |
| Tratamiento 23 | CEE | - | - | - | 1027 | 2853 | 27 | 1127 | 2853 | 90 |
| Tratamiento 24 | CEC | - | - | - | 1117 | 2859 | 39 | - | - | - |
| Tratamiento 25 | CCE | - | - | - | - | - | - | 1235 | 2875 | 97 |

Tabla 4.15: Cantidades reportadas tratamientos E-Mail

Una exploración inicial, muestra que si bien se visualizan diferencias, los resultados a lo largo de los tratamientos y las distintas instancias de contacto no varían en exceso, tanto en la tasa de apertura como en la de clics únicos.

No obstante, algo que salta a la vista es que aquellos tratamientos en los cuales se producen dos envíos de E-Mail (Tratamientos 15, 19 y 20), el segundo envío muestra un desempeño algo peor que en aquellos casos en que el correo electrónico no fue precedido de otro. En particular, las tasas de apertura se reducen en alrededor de 5 puntos porcentuales, lo que demuestra que en este canal, la multicontactabilidad genera efectos negativos y no existe ganancia por el segundo envío de un correo.

Aun así, se puede inferir que esto se debe a que los prospectos, tras el recibimiento del primer correo, ya han logrado capturar la oferta que se les está tratando de comunicar y por tanto, hace sentido que se produzca una disminución en la tasa de apertura. El mismo análisis inicial aplica para las tasas de clics, que muestran un patrón de comportamiento similar.

4.2.3. Efectividad Total

Adicionalmente a lo ya expuesto, la medida más relevante para la validez del diseño experimental en estudio, es sin lugar a dudas las **ventas o altas** generadas por efecto de las distintas combinaciones de estrategias de contacto y los distintos canales utilizados. Tomando la información de ventas, y considerando solo las transacciones en un intervalo acotado de tiempo (como se ha detallado en la sección **3.2.5 Definición de éxito**), los resultados de

las ventas son los siguientes:

| | Código | Ventas totales capturadas | Contactados una instancia | Contactados todas instancias | Efectividad base | Efectividad contactados al menos una instancia | Efectividad contactados todas instancias |
|----------------|--------|---------------------------------|---------------------------------|------------------------------------|---------------------|--|---|
| Tratamiento 1 | CTRL | 12 | 3000 | 3000 | 0.40% | 0.40% | 0.40% |
| Tratamiento 2 | EE | 9 | 1082 | 523 | 0.30% | 0.83% | 1.72% |
| Tratamiento 3 | SS | 2 | 3000 | 3000 | 0.07% | 0.07% | 0.07% |
| Tratamiento 4 | CC | 22 | 2001 | 2001 | 0.73% | 1.10% | 1.10% |
| Tratamiento 5 | C | 28 | 1679 | 1679 | 0.93% | 1.67% | 1.67% |
| Tratamiento 6 | EC | 17 | 1330 | 356 | 0.57% | 1.28% | 4.78% |
| Tratamiento 7 | CE | 25 | 1456 | 221 | 0.83% | 1.72% | 11.31% |
| Tratamiento 8 | SC | 17 | 3000 | 1178 | 0.57% | 0.57% | 1.44% |
| Tratamiento 9 | CS | 22 | 3000 | 1406 | 0.73% | 0.73% | 1.56% |
| Tratamiento 10 | SE | 4 | 3000 | 959 | 0.13% | 0.13% | 0.42% |
| Tratamiento 11 | ES | 12 | 3000 | 863 | 0.40% | 0.40% | 1.39% |
| Tratamiento 12 | SEC | 24 | 1360 | 208 | 0.80% | 1.76% | 11.54% |
| Tratamiento 13 | ESC | 19 | 1342 | 207 | 0.63% | 1.42% | 9.18% |
| Tratamiento 14 | SSC | 17 | 3000 | 1005 | 0.57% | 0.57% | 1.69% |
| Tratamiento 15 | EEC | 16 | 1260 | 287 | 0.53% | 1.27% | 5.57% |
| Tratamiento 16 | SCC | 19 | 3000 | 1236 | 0.63% | 0.63% | 1.54% |
| Tratamiento 17 | ECC | 19 | 1404 | 412 | 0.63% | 1.35% | 4.61% |
| Tratamiento 18 | SSE | 4 | 3000 | 1050 | 0.13% | 0.13% | 0.38% |
| Tratamiento 19 | EES | 8 | 3000 | 947 | 0.27% | 0.27% | 0.84% |
| Tratamiento 20 | ECE | 23 | 1427 | 238 | 0.77% | 1.61% | 9.66% |
| Tratamiento 21 | CES | 13 | 1546 | 514 | 0.43% | 0.84% | 2.53% |
| Tratamiento 22 | CSE | 23 | 1423 | 418 | 0.77% | 1.62% | 5.50% |
| Tratamiento 23 | CEE | 17 | 1501 | 450 | 0.57% | 1.13% | 3.78% |
| Tratamiento 24 | CEC | 25 | 1239 | 312 | 0.83% | 2.02% | 8.01% |
| Tratamiento 25 | CCE | 23 | 1410 | 458 | 0.77% | 1.63% | 5.02% |

Tabla 4.16: Efectividad total tratamientos

Nuevamente, es importante explicar los ratios identificados:

- **Contactados una instancia:** Todos aquellos individuos que fueron efectivamente contactados en alguna de las instancias testeadas. Para el envío de SMS, y dado que no es posible especificar la recepción de mensajes, se asume que todos reciben la comunicación, y por tanto, son contactados.
- **Contactados todas instancias:** Todos aquellos individuos que fueron efectivamente contactados en todas las instancias de contacto.
- **Efectividad base:** Ventas dividido por la base total enviada a gestión (3000 individuos por tratamiento).
- **Efectividad contactados al menos una instancia:** Ventas dividido por número de individuos contactados una instancia.
- **Efectividad contactados todas las instancias:** Ventas dividido por número de individuos contactados todas instancias.

Dado que los órdenes de magnitud son más comparables de acuerdo a tanto efectividad base y la efectividad contactados al menos una instancia, se utilizarán estos ratios en las comparaciones subsiguientes.

Se observa que se dan tratamientos con desempeños peores al del grupo de control (Tratamientos 2 y 3 por ejemplo), generándose incluso menos ventas que en aquel estado en el que no se contacta al cliente. Esto se podría explicar en parte debido al efecto negativo de contactar a los prospectos y "bombardearlos" con ofrecimientos reiterados.

Dejando fuera los tratamientos antes mencionados se observa que en general la aplicación de comunicaciones directas con los clientes generan mejores resultados que el "no hacer nada", confirmando la hipótesis número 2 planteada anteriormente.

Explorando inicialmente la efectividad base de los tratamientos, llama un poco la atención que uno de los tratamientos más exitosos en términos de ventas generadas, resulte ser el Tratamiento 5, el cual corresponde a únicamente un contacto generado por medio de una llamada Outbound. Si bien se requieren más pruebas para concluir respecto al significado de esto, el resultado indica que las estrategias multicontacto son menos efectivas que aquellas en las que solo se contacta mediante una instancia de contacto. Los contactos reiterados generan un efecto inverso al esperado, generando una disminución en las ventas en lugar de aumentarlas.

Esta misma observación anterior desmiente en parte la hipótesis número 3 planteada en secciones anteriores (*H3: La combinación de canales genera mejores tasas de respuesta que las tasas de respuesta individuales de los canales utilizados*), puesto a que el Call por si solo pareciera ser una mejor opción para contactar a los clientes, que comunicaciones en más de una instancia de contacto.

4.2.4. Comparación tratamientos - Analisis post hoc de poder estadístico

Para verificar finalmente el poder estadístico alcanzado por la aplicación del diseño, se puede aplicar un análisis post-hoc de poder, y así concluir respecto a qué tan potentes serán las conclusiones que se pueden hacer de cada uno de los resultados obtenidos tras la implementación del diseño experimental.

| | Poder estadístico |
|---------------------------------|-------------------|
| Tratamiento 1 vs Tratamiento 2 | 0.412 |
| Tratamiento 1 vs Tratamiento 3 | 0.751 |
| Tratamiento 1 vs Tratamiento 4 | 0.816 |
| Tratamiento 1 vs Tratamiento 5 | 0.984 |
| Tratamiento 1 vs Tratamiento 6 | 0.858 |
| Tratamiento 1 vs Tratamiento 7 | 0.981 |
| Tratamiento 1 vs Tratamiento 8 | 0.157 |
| Tratamiento 1 vs Tratamiento 9 | 0.399 |
| Tratamiento 1 vs Tratamiento 10 | 0.529 |
| Tratamiento 1 vs Tratamiento 11 | 0.050 |
| Tratamiento 1 vs Tratamiento 12 | 0.981 |
| Tratamiento 1 vs Tratamiento 13 | 0.918 |
| Tratamiento 1 vs Tratamiento 14 | 0.157 |
| Tratamiento 1 vs Tratamiento 15 | 0.843 |
| Tratamiento 1 vs Tratamiento 16 | 0.237 |
| Tratamiento 1 vs Tratamiento 17 | 0.899 |
| Tratamiento 1 vs Tratamiento 18 | 0.529 |
| Tratamiento 1 vs Tratamiento 19 | 0.140 |
| Tratamiento 1 vs Tratamiento 20 | 0.967 |
| Tratamiento 1 vs Tratamiento 21 | 0.479 |
| Tratamiento 1 vs Tratamiento 22 | 0.969 |
| Tratamiento 1 vs Tratamiento 23 | 0.785 |
| Tratamiento 1 vs Tratamiento 24 | 0.997 |
| Tratamiento 1 vs Tratamiento 25 | 0.969 |

Tabla 4.17: Análisis post-hoc de poder estadístico

Es muy destacable, que en la mayoría de los tratamientos, se alcanza un poder estadístico mayor al 80 % (al menos al contrastar con el grupo de control), por lo que las conclusiones que se pueden hacer respecto a los resultados, tienen un sustento estadístico suficiente.

4.2.5. Comparación tratamientos - Test de hipótesis

El análisis que se puede generar sobre los resultados anteriores, se asocia a la comparación de significatividad estadística entre las proporciones que cada tratamiento entrega. Lo que interesa verificar es si los distintos tratamientos se pueden considerar estadísticamente diferentes a partir de los resultados obtenidos y así concluir respecto a la utilidad o perjuicios que entrega un tratamiento específico. Se usará la metodología del test de comparación de proporciones especificada en la sección de metodología.

En primer lugar, es relevante comparar todos los tratamientos aplicados contra el grupo de control, es decir, aquel conjunto de individuos que no fueron expuestos a ninguna clase de estímulos durante la ejecución del diseño experimental. A priori no es posible saber si un tratamiento es superior o no al grupo de control, por lo que se considerará una hipótesis bidireccional, contrastando sencillamente la diferencia entre dos proporciones, es decir, un test a dos colas:

$$H_0: p_1 = p_2$$

$$H_1: p_1 \neq p_2$$

Con p_1 : efectividad grupo tratado; p_2 : efectividad grupo de control. Es así como se obtienen los siguientes resultados: (Se identifica con una letra 'A' cuando se acepta la hipótesis nula y con una 'R' cuando se rechaza).

| | Comparación efectividad base (Estadístico Z) | Conclusión al 95% | Conclusión al 90% | Comparación efectividad contactados al menos una instancia (Estadístico Z) | Conclusión al 95% | Conclusión al 90% |
|---------------------------------|--|-------------------|-------------------|--|-------------------|-------------------|
| Tratamiento 1 vs Tratamiento 2 | -0.655 | A | A | 1.69 | A | R |
| Tratamiento 1 vs Tratamiento 3 | -2.64 | R | R | -2.64 | R | R |
| Tratamiento 1 vs Tratamiento 4 | 1.70 | A | R | 2.95 | R | R |
| Tratamiento 1 vs Tratamiento 5 | 2.53 | R | R | 4.52 | R | R |
| Tratamiento 1 vs Tratamiento 6 | 0.94 | A | A | 3.27 | R | R |
| Tratamiento 1 vs Tratamiento 7 | 2.13 | R | R | 4.55 | R | R |
| Tratamiento 1 vs Tratamiento 8 | 0.95 | A | A | 0.95 | A | A |
| Tratamiento 1 vs Tratamiento 9 | 1.70 | A | R | 1.70 | A | R |
| Tratamiento 1 vs Tratamiento 10 | -2.03 | R | R | -2.03 | R | R |
| Tratamiento 1 vs Tratamiento 11 | 0 | A | A | 0 | A | A |
| Tratamiento 1 vs Tratamiento 12 | 2.01 | R | R | 4.60 | R | R |
| Tratamiento 1 vs Tratamiento 13 | 1.24 | A | A | 3.69 | R | R |
| Tratamiento 1 vs Tratamiento 14 | 0.95 | A | A | 0.95 | A | A |
| Tratamiento 1 vs Tratamiento 15 | 0.74 | A | A | 3.21 | R | R |
| Tratamiento 1 vs Tratamiento 16 | 1.24 | A | A | 1.24 | A | A |
| Tratamiento 1 vs Tratamiento 17 | 1.24 | A | A | 3.52 | R | R |
| Tratamiento 1 vs Tratamiento 18 | -2.03 | R | R | -2.03 | R | R |
| Tratamiento 1 vs Tratamiento 19 | -0.87 | A | A | -0.87 | A | A |
| Tratamiento 1 vs Tratamiento 20 | 1.87 | A | R | 4.25 | R | R |
| Tratamiento 1 vs Tratamiento 21 | 0.19 | A | A | 1.90 | A | R |
| Tratamiento 1 vs Tratamiento 22 | 1.88 | A | R | 4.27 | R | R |
| Tratamiento 1 vs Tratamiento 23 | 0.95 | A | A | 2.88 | R | R |
| Tratamiento 1 vs Tratamiento 24 | 2.13 | R | R | 5.16 | R | R |
| Tratamiento 1 vs Tratamiento 25 | 1.87 | A | R | 4.29 | R | R |

Tabla 4.18: Test de hipótesis para contraste de tratamientos

En la gran mayoría de los resultados (de la comparación de efectividad de contactados al menos en una instancia), se evidencia una mejora estadísticamente significativa respecto al grupo de control, validándose la hipótesis número 2 propuesta en la sección de metodología respecto a los beneficios que genera la aplicación de campañas de marketing directo para ofrecer un producto específico, debido a que de esa manera, se pueden enfocar de manera mucho más personalizada productos específicos a clientes que prospecten para ellos.

Además, se observan algunas situaciones de impacto negativo en los resultados de los test aplicados. Particularmente, resultados que muestran un desempeño estadísticamente peor o igual al del grupo de control. El tratamiento 10 (SE) demuestra que la combinación de un SMS y un EMAIL no genera una motivación suficiente en los clientes como para incentivar a la venta de un producto particular. Lo mismo ocurre con el tratamiento 18 (SSE), donde

se combinan los mismos canales que en el tratamiento 10. La expectativa natural es que una comunicación de marketing directo al menos genere un mejor resultado que el "no hacer nada" (dejar que la publicidad masiva se encargue de motivar a los clientes).

| | Comparación efectividad base (Estadístico Z) | Conclusión al 95% | Conclusión al 90% | Comparación efectividad contactados al menos una instancia (Estadístico Z) | Conclusión al 95% | Conclusion al 90% |
|----------------------------------|--|-------------------|-------------------|--|-------------------|-------------------|
| Tratamiento 6 vs Tratamiento 7 | -1.21 | A | A | -0.95 | A | A |
| Tratamiento 8 vs Tratamiento 9 | -0.77 | A | A | -0.77 | A | A |
| Tratamiento 10 vs Tratamiento 11 | -2.03 | R | R | -2.03 | R | R |

Tabla 4.19: Test de hipótesis para contraste de tratamientos específicos

Comparando ahora tratamientos en donde los canales utilizados son los mismos, pero en órdenes diferentes, la tabla anterior muestra que solo entre los tratamientos 10 (SE) y 11 (ES) se generan resultados estadísticamente diferentes, es decir, en este caso el orden en que se usan los canales sí tiene una influencia en el resultado obtenido (debido al rechazo en el test evaluado), a diferencia de las otras comparaciones presentadas. Esto se debe posiblemente a la naturaleza de los canales usados (SMS e EMAIL) en contraste a los tratamientos 6, 7, 8 y 9 en donde está presente el Call Center. Este último canal captura la gran mayoría de las transacciones que se generan en cada tratamiento individual, por lo que el resultado que se obtiene es independiente del orden en que se usan los canales en aquellos tratamientos en donde se usa el Call Center.

Tomando en cuenta solo los tratamientos 10 y 11, la hipótesis número 4 (***H4: El orden en que se utilizan los canales tiene influencia en la respuesta de los prospectos, debido al rol particular que cada canal tiene en una interacción con los clientes***) tiene sustento, observándose que el EMAIL tiene un mejor desempeño como iniciador de la comunicación que el SMS, mientras que este último actúa como un buen recordador" de la oferta que ha recibido el cliente (como se muestra en la Tabla 4.16).

Por otro lado, es interesante verificar si la implementación de más contactos trae algún tipo de beneficio adicional, es decir, si generar tres instancias de contacto es mejor que solo generar dos. La tabla a continuación muestra una agregación sencilla en este aspecto sencillamente sumando los valores correspondientes.

| | Ventas totales capturadas | Contactados una instancia | Contactados todas instancias | Efectividad contactados al menos una instancia | Comparación efectividad (Estadístico Z) | Conclusión al 95% | Conclusion al 90% |
|--------------------------|---------------------------|---------------------------|------------------------------|--|---|-------------------|-------------------|
| Contactados 2 instancias | 130 | 20869 | 10507 | 0.62% | -4.082 | R | R |
| Contactados 3 instancias | 250 | 25912 | 7742 | 0.96% | | | |

Tabla 4.20: Test de hipótesis para contraste de tratamientos (agregado por cantidad de contactos)

Como se observa, los resultados muestran que la aplicación de 3 instancias de contacto es estadísticamente mejor que la aplicación de 2 instancias de contacto, tanto al 95% como al 90% de significancia, confirmando a través de este resultado lo dicho en la hipótesis 5 (***H5: La combinación de canales durante 3 instancias de contacto genera mejores tasas de respuesta que aquellas obtenidas mediante solo 2 instancias de contacto***). Esto muestra que si bien tratamientos con 3 instancias de contacto tienen peores resultados que otros con 2 instancias, en términos generales, la aplicación de 3 instancias es una mejor estrategia de comunicación, particularmente los tratamientos 20 (ECE), 22 (CSE), 24 (CEC) y 25 (CCE).

Esto no obstante da espacio para cuestionar hasta qué punto se podría explotar este efecto en campañas futuras: ¿serán cuatro (o incluso más) instancias de contacto mejores que tres instancias? A través de los resultados anteriores no es posible responder esta interrogante, dado que el beneficio marginal al ir aumentando las instancias de contacto es decreciente hasta el punto en el que continuar aumentando la cantidad de instancias de contacto genera costos y no beneficios.

Conclusión

El presente trabajo nace de la necesidad de identificar si la ejecución de campañas por medio de estrategias multicanal en una empresa de telefonía móvil, puede desembocar en mejores resultados de la gestión de estas campañas, principalmente al incrementar la cantidad de ventas generadas y disminuir la cantidad de contactos necesarios para alcanzar dicho nivel de ventas. Los principales factores que juegan un papel en este análisis son los roles de cada canal (y por ende como los clientes pueden responder a cada uno de ellos) y la forma en que cada uno de estos se decide utilizar (orden y relaciones entre canales) para maximizar la capacidad de ellos en la gestión comercial de la empresa.

Dada la inexistencia de campañas aplicadas por medio de estrategias al combinar canales, existe un desconocimiento a priori del efecto que este tipo de campañas puede generar y de cuales combinaciones resultan en mejores niveles de ventas y una mejor experiencia desde la perspectiva del cliente. Por lo anterior, aparece la necesidad de implementar un diseño experimental ad hoc aplicando diferentes combinaciones de los principales canales de marketing directo con que cuenta la compañía: SMS, E-Mail y Call Center de manera de concluir respecto a cómo responden los clientes frente a esta nueva manera de estímulos y así analizar el efecto que estas combinaciones producen.

Esto tiene sentido además como una manera de impulsar la utilización más intensiva de canales diferentes al Call Center (que es el canal principal de campañas en la compañía) debido a los altos costos de utilización de este último y a la invasividad que este tiene sobre los clientes.

El análisis exploratorio de las campañas históricas de la compañía, que permite tener una base sobre la cual evaluar y comparar el desempeño de las combinaciones implementadas, muestran que si bien las campañas apuntadas hacia clientes de la compañía presentan tasas de efectividad superiores que aquellas dirigidas hacia clientes de la competencia, estas son menos masivas (la cantidad de clientes es inferior), por lo que la innovación en este grupo de clientes es muy relevante y la relación con ellos debe llevarse con cautela.

Dado que en general, las campañas generadas por canales que no son Call Center son más bien informativas y no comerciales, es decir, no llaman directamente al consumo de un producto en particular, la evaluación comparativa de canales no es tan directa. No obstante, otros indicadores básicos tales como la tasa de apertura y clicks (*clickthrough rate*) de un correo permiten también tener un punto base sobre el cual evaluar resultados posteriores.

Desde esta perspectiva, los resultados de las regresiones implementadas muestran que los

principales drivers detrás de la aceptación de una oferta comercial se relacionan con variables temporales, principalmente la cantidad de meses pasados desde la última transacción de los clientes mostrando una clara preferencia por adquirir un producto o servicio ofrecido por la compañía a medida que aumentan los meses desde la última transacción generada. Se observa que otras variables demográficas como la edad y el grupo socioeconómico de los clientes también tiene influencia en la decisión, cada uno de forma inversa: a mayor edad menor probabilidad de compra; a mayor gse mayor probabilidad de compra.

Los distintos escenarios testeados muestran que la utilización del Call Center como canal de comunicación es fundamental, dado el alto grado de información que posee y a la alta interactividad que permite a los clientes. Se observa además que tratamientos que van desde canales menos a más invasivos con el cliente desempeñan de mejor manera que aquellos que van en otra dirección, mostrando que el apoyo de una llamada por Call mediante un SMS o un E-Mail previo tiene efectos positivos y refuerza la comunicación con los clientes.

Además, se observa que en términos agregados, la implementación de 3 instancias de contacto es estadísticamente mejor que la implementación de solo 2 instancias de contacto, validando hipótesis de multicontactabilidad. Aun así, es necesario tener un enfoque prudente respecto a esto, dado que lo anterior puede llevar a concluir erróneamente que más contactos es mejor, es decir que el beneficio por cada contacto adicional es siempre positivo, lo que en la práctica no es así debido al efecto de "ahogo" que se produce cuando el cliente es bombardeado con ofertas y comunicaciones, produciéndose a la larga un efecto negativo.

Si bien en el transcurso de este proyecto, se ha detectado que la aplicación de estrategias multicanal y multicontacto pueden tener beneficios en la gestión de las campañas, algunos tratamientos testeados muestran que ciertas combinaciones poseen efectos negativos, lo que puede deberse a factores confounders no identificados ni controlados en el diseño experimental. Uno de ellos puede ser por ejemplo los días en que se contacta a los clientes, puesto que si bien la asignación a tratamientos se generó en forma aleatoria, puede ser que ciertos contactos se generaran en días u horas particulares que bajan el desempeño de un tratamiento particular. Otro elemento que puede influenciar negativamente los resultados es la campaña particular que se está testeando. Campañas CQC o en aquellas en las que no se tiene mucha información de los clientes puede que tengan un mejor desempeño al aplicar estrategias de contacto multicanal, pero dado esto, identificar atributos de los clientes y sus correlaciones con cada tratamiento no es posible a priori.

Cómo extensión o trabajo futuro del estudio, se puede considerar la extrapolación hacia otro tipo de campaña en la cual las transacciones que se produzcan sean de otra especie (upselling o sva) o bien del negocio fijo. Adicionalmente, una medida importante que se puede considerar para medir el efecto de los tratamientos ejecutados es el "churn." la tasa de fuga generada a través de los tratamientos, validando en consecuencia el posible efecto negativo que genera la gestión de campañas sobre los clientes.

Anexos

Anexo A

Descripción del mercado

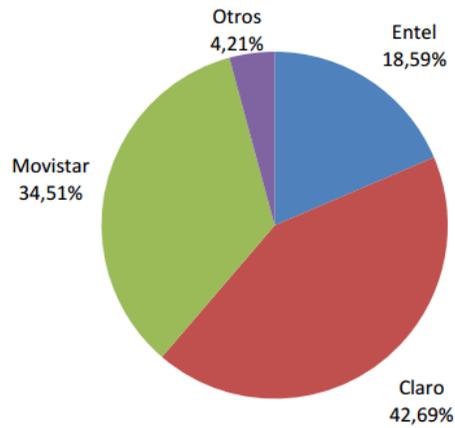


Figura A.1: Abonados contrato 30 días por empresa (Total: 7 374 909) (Fuente: Subtel)

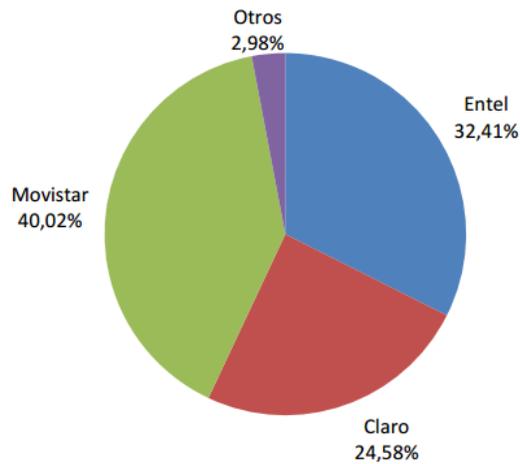


Figura A.2: Abonados prepago 30 días por empresa (Total: 16 308 442) (Fuente: Subtel)

| Año | Número de Líneas en servicio | Crecimiento Año Anterior | Penetración por cada 100 hab. |
|------|------------------------------|--------------------------|-------------------------------|
| 2000 | 3,302,498 | | 21.33 |
| 2001 | 3,478,492 | 5.33% | 22.21 |
| 2002 | 3,467,013 | -0.33% | 21.90 |
| 2003 | 3,252,063 | -6.20% | 20.32 |
| 2004 | 3,345,102 | 2.86% | 20.67 |
| 2005 | 3,460,645 | 3.45% | 21.17 |
| 2006 | 3,383,597 | -2.23% | 20.49 |
| 2007 | 3,459,611 | 2.25% | 20.74 |
| 2008 | 3,524,790 | 1.88% | 20.92 |
| 2009 | 3,555,031 | 0.86% | 20.90 |
| 2010 | 3,458,766 | -2.71% | 20.14 |
| 2011 | 3,370,104 | -2.56% | 19.45 |
| 2012 | 3,284,296 | -2.55% | 18.79 |
| 2013 | 3,338,661 | 1.66% | 18.93 |
| 2014 | 3,416,916 | 2.34% | 19.08 |

Tabla A.1: Líneas fijas en servicio por año (Fuente: Subtel)

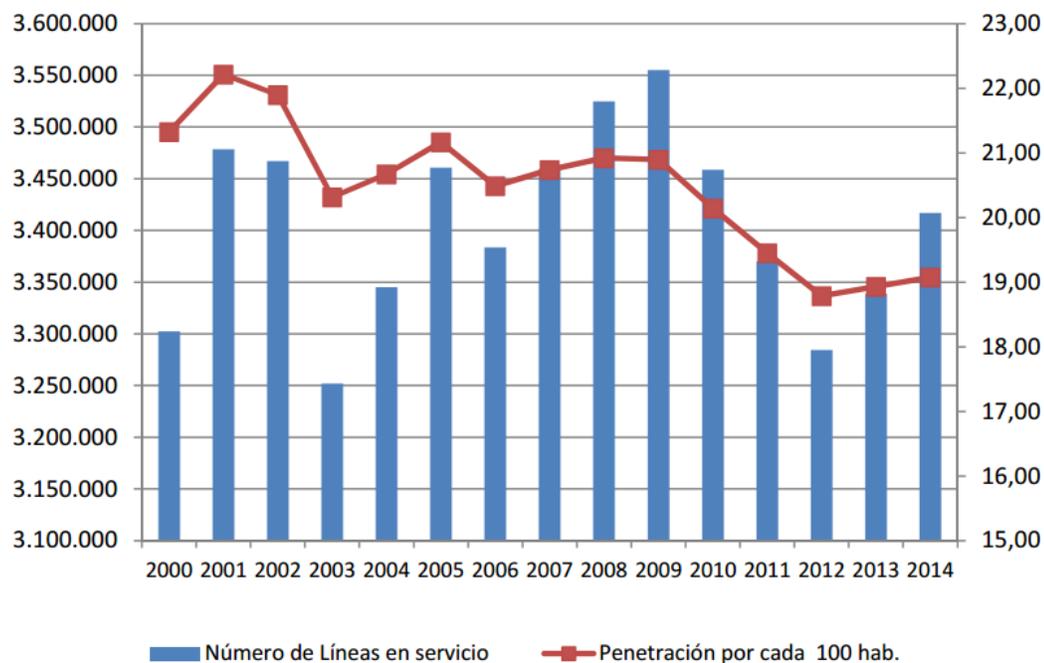


Figura A.3: Líneas fijas en servicio por año (Fuente: Subtel)

Anexo B

Ejemplo utilizacion *G-Power*

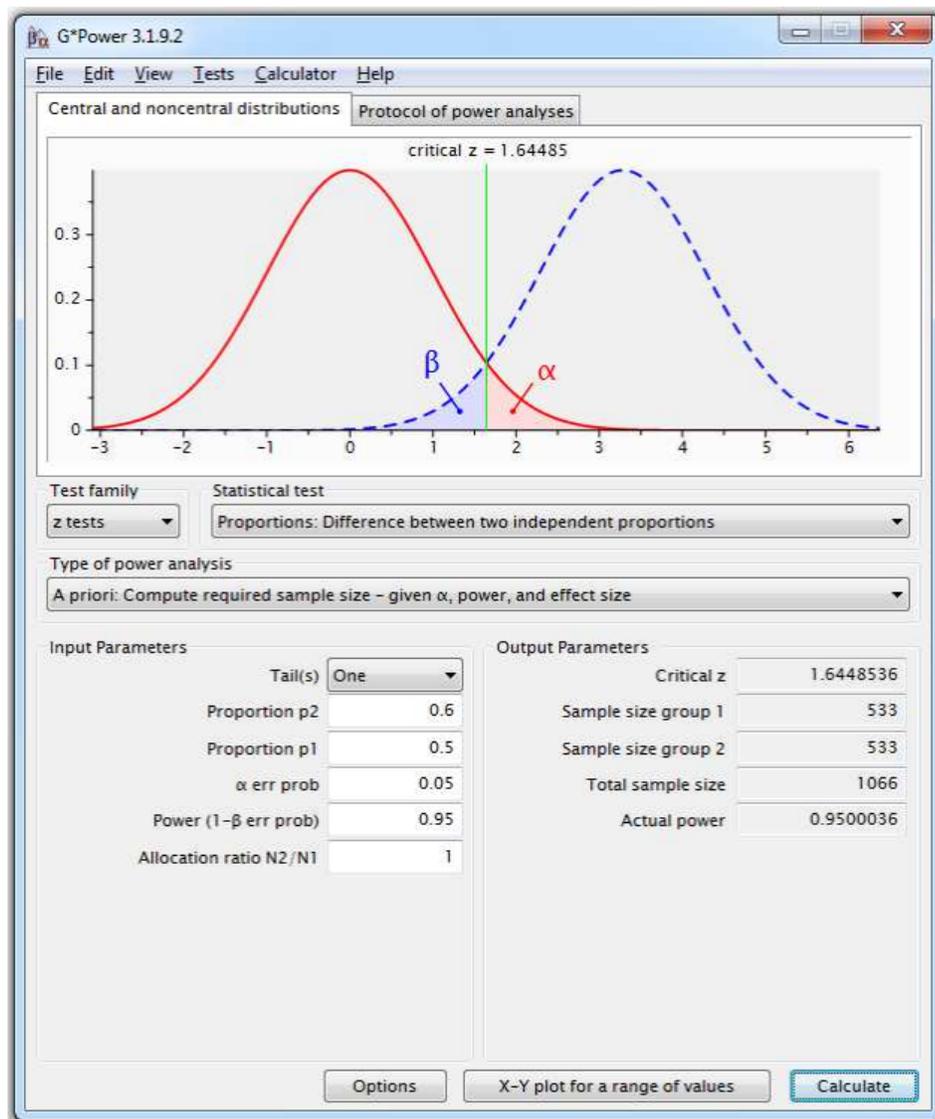


Figura B.1: Interfaz software G-Power

Anexo C

Boxplots

Dada una serie de datos, los boxplots se construyen de acuerdo a un procedimiento de acuerdo a lo siguiente:

1. Ordenar la serie en orden ascendente e identificar la mediana del set de datos, separando este en 2 subconjuntos D_1 y D_2 . La mediana se llamará Q_2
2. Nuevamente, encontrar la mediana de ambos conjuntos, los que se llamarán Q_1 y Q_3 respectivamente. Estos corresponden a los límites en los cuales la data no se considera fuera de rango. Con esto, se puede construir la siguiente figura:

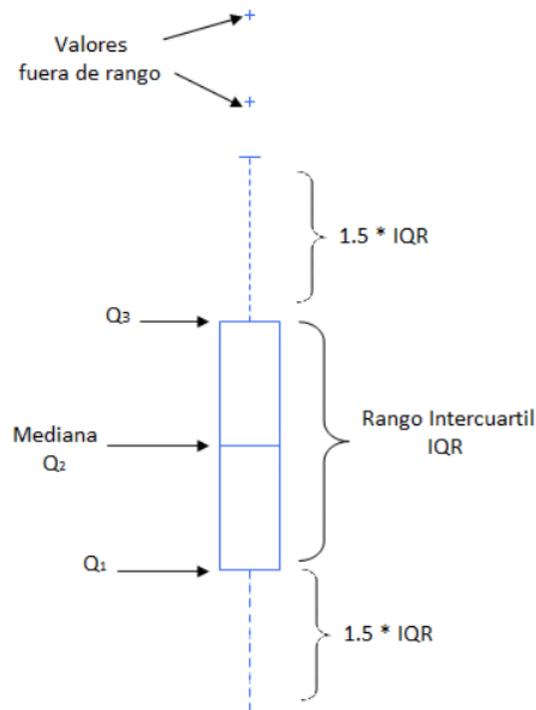


Figura C.1: Definición de cuartiles data

3. De acuerdo a lo anterior, el primer cuartil corresponde a el intervalo entre el valor mínimo y el punto Q_1 , el segundo cuartil corresponde al intervalo entre Q_1 y Q_2 , el tercer cuartil al intervalo entre Q_2 y Q_3 y finalmente, el último cuartil se encuentra en el intervalo entre Q_3 y el valor máximo.
4. Con los valores antes encontrados, se determina el rango intercuartil $IQR = Q_3 - Q_1$, se construye un intervalo superior y otro inferior. Los extremos corresponden a $Q_1 - 1,5 * IQR$ y $Q_3 + 1,5 * IQR$.
5. Los valores que se encuentren dentro de los intervalos $(Q_1 - 1,5 * IQR, Q_1)$ y $(Q_3, Q_3 + 1,5 * IQR)$ se consideran valores que están fuera del rango de mayor concentración de la data, pero no se les considera efectivamente outliers. Los valores fuera de esos rangos se considerarán fuera de rango de acuerdo a este análisis como se ve en la figura anterior.

Anexo D

Construcción binarias (código R)

Se construye inicialmente la variable `cambio_traf_datos`:

$$\begin{aligned} \text{Si } \{ \text{diferencias}_{\text{traf_datos}_{\text{Mes1},\text{Mes2}}} > 0.3 * \text{traf_datos}_{\text{Mes1}} \} &\rightarrow \text{cambio}_{\text{traf_datos}_{\text{Mes1},\text{Mes2}}} = 1 \\ \text{Si } \{ -\text{diferencias}_{\text{traf_datos}_{\text{Mes1},\text{Mes2}}} > 0.3 * \text{traf_datos}_{\text{Mes1}} \} &\rightarrow \text{cambio}_{\text{traf_datos}_{\text{Mes1},\text{Mes2}}} = -1 \end{aligned}$$

El código en el software R que logra lo anterior:

```
traf_datos06_05 = traf_datos[,3]-traf_datos[,2]
traf_datos07_06 = traf_datos[,4]-traf_datos[,3]
traf_datos08_07 = traf_datos[,5]-traf_datos[,4]
traf_datos09_08 = traf_datos[,6]-traf_datos[,5]
traf_datos10_09 = traf_datos[,7]-traf_datos[,6]

diferencias_traf_datos =
cbind(traf_datos06_05,traf_datos07_06,traf_datos08_07,traf_datos09_08,traf_datos10_09)

cambio_traf_datos =
cbind(rep(0,nrow(diferencias_traf_datos)),rep(0,nrow(diferencias_traf_datos)),
rep(0,nrow(diferencias_traf_datos)),rep(0,nrow(diferencias_traf_datos)),
      rep(0,nrow(diferencias_traf_datos)))

for(i in 1:nrow(diferencias_traf_datos)){
  for(j in 1:ncol(diferencias_traf_datos)){
    if(diferencias_traf_datos[i,j]>0.3*traf_datos[i,j+1]){cambio_traf_datos[i,j]=1}
    else if(-diferencias_traf_datos[i,j]>0.3*traf_datos[i,j+1]){cambio_traf_datos[i,j]=-1}
  }
}
```

Con lo anterior, se construye la variable `tendencia_traf_datos`

```

for(i in 1:nrow(diferencias_traf_datos)){
if((cambio_traf_datos[i,1]+cambio_traf_datos[i,2]+cambio_traf_datos[i,3]+cambio_traf_datos[i,4]+cambio_traf_datos[i,5])>0){tendencia_traf_datos[i]=1}
else
if((cambio_traf_datos[i,1]+cambio_traf_datos[i,2]+cambio_traf_datos[i,3]+cambio_traf_datos[i,4]+cambio_traf_datos[i,5])<0){tendencia_traf_datos[i]=-1}
}

```

Y con ello, los inputs que interesan de aumento y disminución:

```

aumento_traf_datos = rep(0,length(tendencia_traf_datos))
aumento_traf_datos[which(tendencia_traf_datos==1)]=1

disminucion_traf_datos = rep(0,length(tendencia_traf_datos))
disminucion_traf_datos[which(tendencia_traf_datos==-1)]=1

```

Bibliografía

- [1] Diseño de estrategias integradas de comunicación entre múltiples canales. Ceret - Universidad de Chile, 2014.
- [2] Eric Almquist and Gordon Wyner. Boost your marketing roi with experimental design. *Harvard business review*, 79(9):135–141, 2001.
- [3] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rudiger Wirth. Crisp-dm 1.0 step-by-step data mining guide. 2000.
- [4] Scott A Czepiel. Maximum likelihood estimation of logistic regression models: theory and implementation. *Available at czep.net/stat/mlelr.pdf*, 2002.
- [5] Esther Duflo, Rachel Glennerster, and Michael Kremer. Using randomization in development economics research: A toolkit. *Handbook of development economics*, 4:3895–3962, 2007.
- [6] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191, 2007.
- [7] EP George, William Gordon Hunter, and J Stuart Hunter. *Statistics for experimenters: Design, innovation, and discovery*. Wiley, 2005.
- [8] Tim Hughes. New channels/old channels: Customer management and multi-channels. *European Journal of Marketing*, 40(1/2):113–129, 2006.
- [9] Gary King and Langche Zeng. Logistic regression in rare events data. *Political analysis*, 9(2):137–163, 2001.
- [10] John G Lynch Jr and Dan Ariely. Wine online: Search costs affect competition on price, quality, and distribution. *Marketing Science*, 19(1):83–103, 2000.
- [11] David P MacKinnon, Jennifer L Krull, and Chondra M Lockwood. Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, 1(4):173–181, 2000.
- [12] Scott A Neslin, Dhruv Grewal, Robert Leghorn, Venkatesh Shankar, Marije L Teerling, Jacquelyn S Thomas, and Peter C Verhoef. Challenges and opportunities in multichannel customer management. *Journal of Service Research*, 9(2):95–112, 2006.

- [13] John O Rawlings, Sastry G Pantula, and David A Dickey. *Applied regression analysis: a research tool*. Springer Science & Business Media, 1998.
- [14] William R. Shadish, Thomas D Cook, and Donald Thomas Campbell. *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage learning, 2002.
- [15] Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.
- [16] Rajkumar Venkatesan and Vita Kumar. A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of marketing*, 68(4):106–125, 2004.
- [17] Vijay Vishwanath and Gerry Mulvin. Multi-channels: the real winners in the b2c internet wars. *Business Strategy Review*, 12(1):25–33, 2001.