



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

DESARROLLO DE UN MODELO PREDICTIVO PARA EL AUMENTO DE LA TASA DE
ASISTENCIA A CITAS MÉDICAS AMBULATORIAS EN EL HOSPITAL CLÍNICO DE LA
UNIVERSIDAD DE CHILE

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

JUAN CARLOS RUBILAR ORELLANA

PROFESOR GUÍA:
ROBERTO SEGOVIA SEGOVIA

MIEMBROS DE LA COMISIÓN:
DANIEL VARELA LÓPEZ
RICARDO LOYOLA MORAGA

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: JUAN CARLOS RUBILAR ORELLANA
FECHA: 2019
PROF. GUÍA: ROBERTO SEGOVIA SEGOVIA

DESARROLLO DE UN MODELO PREDICTIVO PARA EL AUMENTO DE LA TASA DE ASISTENCIA A CITAS MÉDICAS AMBULATORIAS EN EL HOSPITAL CLÍNICO DE LA UNIVERSIDAD DE CHILE

Uno de los problemas más habituales en los centros de atención ambulatoria corresponde a las inasistencias de los pacientes a las citas médicas agendadas. Existen variadas metodologías y factores de causalidad en la literatura, sin embargo, la variación del comportamiento de los pacientes es sensible al medio en el cual se encuentran inmersos, de esta forma, se realiza un análisis previo de variables involucradas para establecer los factores que afectan en mayor medida el comportamiento de los pacientes.

El presente estudio tiene por objetivo generar un modelo estructural bajo dos focos de investigación. El primero lo constituye la transformación de data en información útil para la toma de decisiones. Bajo esta base se establecen hipótesis asociadas a variables que eventualmente se relacionan con las inasistencias de los pacientes, de este modo, derribar asunciones erróneas y dimensionar de forma clara y fehaciente el impacto que generan los factores de investigación sobre el comportamiento de asistencia del individuo.

El estudio muestra la estructuración del modelo con foco explicativo, del cual se establecen probabilidades promedios asociadas a diversos factores testeados. Dichas probabilidades son de suma utilidad para establecer políticas de gestión adecuadas para lograr un mayor nivel de asistencia a citas médicas ambulatorias y que dichas políticas sean sostenibles en el tiempo.

Una vez se han evaluado las distintas especificaciones de utilidad y se ha obtenido el nivel de significancia adecuado, se obtiene el ajuste de un modelo estructural con foco predictivo. Dicho modelo mantiene un ajuste robusto logrando un 84.44 % de predicciones correctamente clasificadas. Si se extrapolan los resultados históricos a gestiones futuras utilizando el modelo predictivo, se puede establecer el impacto anual que generaría la aplicación de dicho modelo, llegando a un aumento de 9.5 % de los ingresos asociados a citas médicas ambulatorias de la institución (en un caso promedio), lo cual es equivalente a una reducción de la tasa de inasistencias de 21.7 % a 14.25 %, aumentando en 38.000 el número de citas médicas asistidas anualmente. Es importante mencionar que dicha aplicación debe estar acompañada de políticas de gestión adecuadas para evitar fuga de pacientes ocasionadas por problemas de gestión interna (tiempos de espera, aglomeraciones, tiempos de atención, etc.).

De esta forma, la dualidad en la aplicación del modelo, evaluado bajo distintos focos de investigación, genera un impacto positivo de forma directa al predecir las inasistencias de los pacientes e indirecta al establecer numéricamente la probabilidad de asistencia de un paciente bajo distintos factores del medio, mejorando la toma de decisiones y políticas de acción asociadas.

En memoria de mi familia por todo el apoyo incondicional que me han brindado

Tabla de Contenido

1. Antecedentes Generales	1
1.1. Introducción	1
1.1.1. Sistema de salud en Chile	1
1.1.2. Hospital Clínico de la Universidad de Chile	1
1.2. Descripción del problema	6
1.3. Justificación del proyecto	9
2. Marco Teórico	10
2.1. Metodología KDD (Knowledge Discovery in Databases)	10
2.2. Modelos estructurales	11
2.2.1. Modelos de elección discreta	11
2.2.2. Modelo Logit	12
2.2.3. Modelo Probit	13
2.2.4. Modelo Logit Mixto	13
2.2.5. Interpretación de modelos	14
2.2.6. Evaluación de bondad de ajuste	15
3. Estructura de la investigación	16
3.1. Objetivos	16
3.1.1. Objetivo General	16
3.1.2. Objetivos Específicos	16
3.2. Hipótesis de investigación	17
3.3. Metodología	18
3.4. Resultados esperados	19
3.5. Alcances	20
4. Tratamiento y Análisis de datos	21
4.1. Bases de datos	21
4.1.1. Fuentes internas de información	21
4.1.2. Fuentes externas de información	22
4.2. Tratamiento de data	23
4.2.1. Selección de pacientes	23
4.2.2. Selección de registros y transformación de data	26
4.3. Análisis descriptivo	27
4.4. Selección de variables	36

5. Modelamiento	40
5.1. Validación de la muestra	41
5.2. Estimación y evaluación del modelo empírico: Foco explicativo	44
5.2.1. Modelo estructural con efectos fijos	46
5.2.2. Modelo estructural con efectos fijos y aleatorios	51
5.3. Análisis del modelo	54
5.3.1. Análisis del modelo con foco explicativo y validación de hipótesis	54
5.3.2. Análisis del modelo con foco predictivo	64
5.3.3. Discusión	67
6. Aplicación del modelo estructural sobre políticas de gestión	69
6.1. Evaluación de política de extensión de agenda	69
6.2. Evaluación de aplicación del modelo con foco predictivo sobre la gestión de citas médicas en Call Center	74
6.2.1. Aspectos a tener en cuenta en la implementación	74
Conclusión	76
A. Apéndice A: Anexo metodológico	80
A.1. Modelos de elección discreta y maximización de utilidad aleatoria	80
B. Apéndice B: Ajuste del modelo estructural: Foco explicativo	82
B.1. Ajuste de coeficientes	82
B.2. Odds ratios para modelos con ajuste Logit y Mixed Logit (efectos fijos y aleatorios)	93
B.3. Ajuste del Modelo III con función de enlace LOGIT con extensión de data.	104

Índice de Tablas

1.1. Especialidades HCUCH	3
4.1. Porcentaje de asistencia por sucursal (incluye sobrecupos)	28
4.2. Porcentaje de asistencia por género del paciente	29
4.3. Definición de variables utilizadas en el modelo predictivo	39
5.1. Resumen de ajuste Modelo I con función de enlace LOGIT	47
5.2. Resumen de ajuste Modelo I y II con función de enlace LOGIT	49
5.3. Resumen de ajuste Modelo I, II y III con función de enlace LOGIT	50
5.4. Resumen de ajuste del Modelo III con función de enlace LOGIT y PROBIT	52
5.5. Resumen de ajuste del Modelo III con función de enlace LOGIT para la muestra y extensión de data	65
5.6. Resumen de clasificación de predicciones: sensibilidad y especificidad	66
6.1. Variación de la probabilidad de la probabilidad de asistencia frente a variación mar- ginal de la latencia	73

Índice de Ilustraciones

1.1.	Variación de consultas médicas consolidadas 2016 - 2018	4
1.2.	Agendamiento web anual de la institución	5
1.3.	Consultas médicas totales por previsión de salud	6
1.4.	Distribución etaria por género del paciente para citas médicas solicitadas	6
1.5.	Solicitudes de cita médica con registro de email periodo 2017 - 2018	7
2.1.	Pasos que compone el proceso de KDD	10
3.1.	Fases principales en la obtención de un modelo de ecuaciones estructurales	19
4.1.	Proceso de requerimiento de información	25
4.2.	Evolución de frecuencia de solicitudes por edad periodo 2013 - 2018	28
4.3.	Evolución de frecuencia de solicitudes por edad periodo 2013 - 2018	30
4.4.	Evolución mensual de asistencia a citas médicas periodo 2013 - 2018	30
4.5.	Variación diaria promedio de asistencia a citas médicas año 2018	31
4.6.	Variación diaria promedio anual (por día de la semana) de asistencia a citas médicas año 2018	31
4.7.	Asistencia a citas médicas por canal de agendamiento (incluye sobrecupos)	32
4.8.	Nivel de asistencia promedio a citas médicas por latencia periodo 2013 - 2018	33
4.9.	Nivel de asistencia promedio a citas médicas por género y latencia periodo 2013 - 2018	33
4.10.	Nivel de asistencia promedio a citas médicas por canal y latencia periodo 2013 - 2018	34
4.11.	Nivel de asistencia promedio a citas médicas por edad y género periodo 2013 - 2018	35
4.12.	Nivel de asistencia promedio a citas médicas por tipo de cita y edad periodo 2013 - 2018	35
5.1.	Frecuencia promedio de citas médicas y antigüedad promedio de pacientes por cantidad de solicitudes del panel	42
5.2.	Gráficos de ajuste de Modelo I mediante función de enlace LOGIT	48
5.3.	Gráficos de ajuste de Modelo II mediante función de enlace LOGIT	49
5.4.	Gráficos de ajuste de Modelo III mediante función de enlace LOGIT	51
5.5.	Gráficos de ajuste de Modelo III mediante la función de enlace PROBIT y del Modelo IV mediante la función de enlace MIXED LOGIT	53
5.6.	Curva ROC del Modelo III y IV con funciones de enlace PROBIT y MIXED LOGIT respectivamente	54
5.7.	Variación de probabilidad promedio asociada a la cita médica por edad y género	56
5.8.	Probabilidad esperada asociada a la cita médica por especialidad	57

5.9. Probabilidad esperada bajo variaciones en el nivel de latencia de la cita médica . . .	59
5.10. Probabilidad de asistencia promedio por horarios de atención	59
5.11. Probabilidad de asistencia promedio nivel de fiabilidad de declaraciones del pa- ciente y nivel de asistencia histórico	61
5.12. Probabilidad de asistencia esperada por nivel de precipitación	63
6.1. Descripción del proceso de extensión de citas	70

Capítulo 1

Antecedentes Generales

1.1. Introducción

1.1.1. Sistema de salud en Chile

El sistema de salud Chileno es un sistema mixto a nivel de financiamiento, aseguramiento y provisión de las prestaciones de salud. El ente regulador del sistema en conjunto es el Ministerio de Salud. El financiamiento de salud es compartido de acuerdo a los diferentes sistemas de aseguramiento, estos son: FF.AA., FONASA (Fondo Nacional de Salud), ISAPRE (Instituciones de Salud Provisional) y Mutuales de seguridad.

El Estado financia completamente a FF.AA. y de forma parcial al sistema FONASA, éste último financia su complemento con cotizaciones previsionales legales provenientes del trabajador (7 % del sueldo bruto) y asigna a cada usuario, de acuerdo a sus ingresos, a una categoría determinada, donde el tramo A es el beneficiario con mayor vulnerabilidad (carente de recursos) y accede a atención gratuita en instituciones públicas, los tramos B al D corresponden a beneficiarios con ingresos correspondiente a rentas provenientes de trabajo formal o pensionados. Los beneficiarios que poseen ingresos tiene la posibilidad de atención en instituciones de salud pública con un porcentaje de bonificación de acuerdo al tramo o utilizar el "sistema de libre elección", el cual corresponde a la posibilidad de atención médica en instituciones de salud privadas adscritas a través de compra de bonos, estos últimos son co-pagos de los asegurados, donde su porcentaje de bonificación será dependiente del tramo al que pertenezca. El área aseguradora ISAPRE es financiada en su totalidad por el cotizante suscrito, el cual puede escoger entre las distintas coberturas de salud propuestas por las instituciones. Finalmente existen las Mutuales de seguridad, área aseguradora que administra fondos proporcionados por el empleador (cotización variable obligatoria y dependiente del tipo de contrato de trabajo), otorga cobertura total a los siniestros por accidentes laborales.

1.1.2. Hospital Clínico de la Universidad de Chile

El Hospital de la Universidad de Chile fue fundado en 1874 bajo el nombre de San Vicente de Paul durante el gobierno de Federico Errázuriz Zañartu, luego de una serie de modificaciones en 1952 nace el Hospital Clínico Universidad de Chile con el nombre del más ferviente propulsor, impulsor del concepto de un Hospital Universitario, Dr. José Joaquín Aguirre. A fines de 1994

concluyó un convenio que existía desde la década del 50 entre el Hospital y el Ministerio de Salud, en el cual este último pagaba por las prestaciones médicas realizadas a pacientes derivados de los servicios médicos estatales. El monto estatal financiaba de manera insuficiente al Hospital, aumentando su deuda año a año. El término del convenio obligó al Hospital Clínico asumir el desafío de un cambio en su modelo de gestión financiera y entrar al mercado de la salud, lo que le ha permitido en la actualidad generar prácticamente el 90 % de su presupuesto y realizar una serie de inversiones en tecnología de punta, planta física y recursos humanos, que lo ubican entre los mejores establecimientos de salud del país.

Misión

Ser el principal hospital universitario del país que, junto con la formación de profesionales de la salud de excelencia, la investigación y la mejor práctica clínica, otorgue a la comunidad soluciones integrales de salud.

Sucursales

La institución consta de una filial de menor tamaño llamada Clínica Universidad de Chile Quilín, la cual está capacitada para brindar soluciones integrales en salud: atención primaria y secundaria. Cuenta con una amplia gama de servicios en cirugía mayor ambulatoria y un centro de diagnóstico. La atención en la filial permite derivar a los pacientes al Hospital Clínico Universidad de Chile cuando es necesario brindar soluciones más complejas. Estructuralmente se encuentra también la clínica psiquiátrica, sin embargo, las gestiones ambulatorias son realizadas por medio del Hospital.

Servicios

El Hospital Clínico de la Universidad de Chile (HCUCH) es un "Hospital Escuela", es decir, una institución en donde sus médicos deben realizar horas de docencia en la Facultad de Medicina de la Universidad, al mismo tiempo de servir como lugar de práctica para los futuros profesionales de la salud que estudian en dicha facultad.

Dada la dualidad en objetivos mencionada previamente, existe una categorización en las consultas médicas: consultas institucionales, consultas realizadas en horario de trabajo (docencia) y por ende no reciben vinculación (renta variable asociada a servicios médicos) y consultas no institucionales, consultas médicas realizadas fuera del horario de trabajo (de acuerdo al contrato de trabajo vigente) y por ende recibe una remuneración variable asociada a los servicios prestados. Las consultas no institucionales son realizadas generalmente en las tardes y no existe variación en el costo de la consulta médica, además el paciente no tiene conocimiento del tipo de consulta a la hora de su atención, ya que esta no es de conocimiento explícito.

Los servicios de asistencia se dividen en: Atención ambulatoria, atención cerrada y atención de urgencias. Los servicios de atención ambulatoria ofrecidos por la institución se categorizan en: cirugía, obstetricia y ginecología, medicina y especialidades (ver tabla 1.1).

De acuerdo al informe estadístico 2018, las consultas médicas totales para los centros HCUCH y clínica Quilín para el año 2016 corresponden a 446.405 consultas consolidadas. Para el año 2017 la cantidad sufre una disminución de 3,35 % llegando a 431.454 consultas, finalmente para el año 2018 sufre una disminución de 5,1 % con respecto al año anterior y un 8,3 % en comparación al año 2016, llegando a 409.539 consultas médicas consolidadas (ver figura 1.1). Para el año 2018, 74,8 % de las consultas médicas realizadas por la institución corresponden al Hospital y un 25,2 %

SERVICIOS DE MEDICINA	Medicina Interna	Gastroenterología
	Nefrología	Neumología
	Hematología	Reumatología
	Endocrinología	Oncología
	Genética	Inmunología
	Geriatría	Infectología
SERVICIO DE CIRUGÍA		
SERVICIO DE OBST. Y GINECOLOGÍA		
SERVICIOS DE ESPECIALIDADES	Pediatría	Cardiología
	Urología	Oftalmología
	Otorrinolaringología	Traumatología
	Neurología	Neurocirugía
	Dermatología	Psiquiatría
	Medicina Física y Rehabilitación	Laboratorio Clínico
	Unidad del Dolor	Anestesiología
	Neonatología	Radiología Intervencional
	Medicina Nuclear	Centro de Consultas y Procedimientos
	Cuidado Continuo y Paliativo	

Tabla 1.1: Especialidades HCUCH
(Fuente: Elaboración propia)

fueron realizadas en Clínica Quilín.

En el año 2018, el hospital presenta un 69,4 % de las citas médicas consolidadas como consultas institucionales, observándose una reducción de 4,5 % con respecto a la misma categoría del año anterior y 30,6 % como no Institucionales, observándose una reducción de 9,3 % con respecto a la misma categoría del año anterior. En cuanto a la distribución de los servicios de asistencia ambulatoria, se observa que el 18,6 % corresponde a servicios de medicina, donde la mayor contribución es realizada por Reumatología y Gastroenterología con un 18 % y 17,8 % respectivamente, 12 % correspondiente a servicios de cirugía, 16,1 % corresponden a servicios de Obstetricia y ginecología y finalmente el 53,3 % es contribuido por servicios de especialidades.

Dimensionamiento de la actividad

De acuerdo a la memoria de la Universidad de Chile 2014-2018 [7], para el año 2016 el Hospital Clínico de la Universidad de Chile consta de:

- 1.155 profesionales clínicos
- sobre 400.000 consultas realizadas anualmente
- 20.000 cirugías efectivas
- 630.000 procedimientos clínicos
- 1.330.000 exámenes de laboratorio e imagenología

HCUCH es reconocido como una de las 50 instituciones más innovadoras del país (ranking UAI), una de las mejores instituciones de salud (ranking América Economía) y una de las instituciones más eficientes en gestión en Chile (ranking Ministerio de Salud).

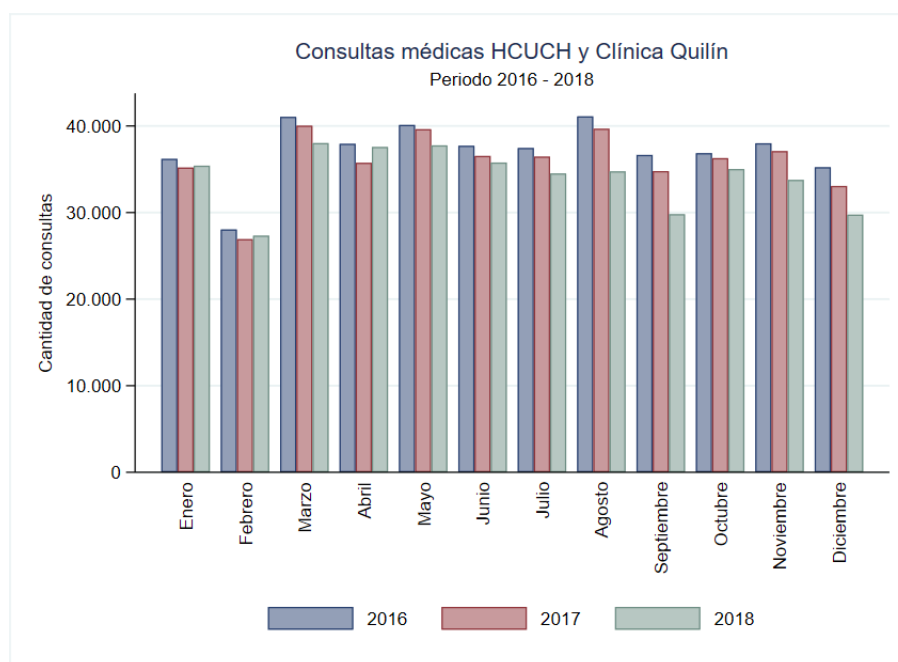


Figura 1.1: Variación de consultas médicas consolidadas 2016 - 2018
(Fuente: Informe estadístico HCUCH Diciembre 2018)

En cuanto a la inversión y financiamiento, el año 2013 el hospital presentaba pérdidas¹ por 10.000 millones de dólares, sin embargo, al año 2017 se registró un superávit de 2.400 millones de dólares. De igual forma, se genera una reducción de la deuda con proveedores de \$34.000 millones en 2013 a \$13.700 millones en 2017. Para el año 2018 se genera una inversión de 17.000 millones de dólares en programas de actualización del equipamiento y formación de especialistas.

Canales de atención

Los canales de atención del hospital son presenciales, Call Center y web. Hoy en día todas las consultas médicas agendadas se concentran en dos bases de datos, AppContact y TICARES, la primera utilizada para la gestión y flujo de información con el paciente (gestiones de confirmación, información de las citas, observaciones en la comunicación, etc.) y la segunda corresponde a la ficha médica electrónica, donde se almacena toda la información personal del paciente (RUT, género, dirección, morbilidad, etc.) (ver sección 3.1).

El sistema de agendamiento web es uno de los canales emergentes de la institución. De acuerdo a la base de datos Ticares, desde el año 2013 al año 2018 el porcentaje de agendamientos web en comparación a otro medio prácticamente se duplicó en base al total de agendamientos anuales, en el año 2018 se ha observado una leve disminución de su utilización en comparación al anterior con respecto a sus citas totales, sin embargo, mantiene un porcentaje sólido en la captación de citas con un 17,7 % (ver figura 1.2). Call Center es el canal de comunicación más activo del hospital en cuanto a transferencias de información debido a que gestiona las llamadas entrantes (inbound)

¹Es importante destacar que las pérdidas referidas en este apartado provienen de un origen multifactorial y corresponden a estados de resultados financieros públicos, de esta forma, la información entregada tiene como propósito informar acerca del dimensionamiento de la actividad de la institución, y no hace alusión directa a las pérdidas relacionadas al problema tratado en el presente estudio.

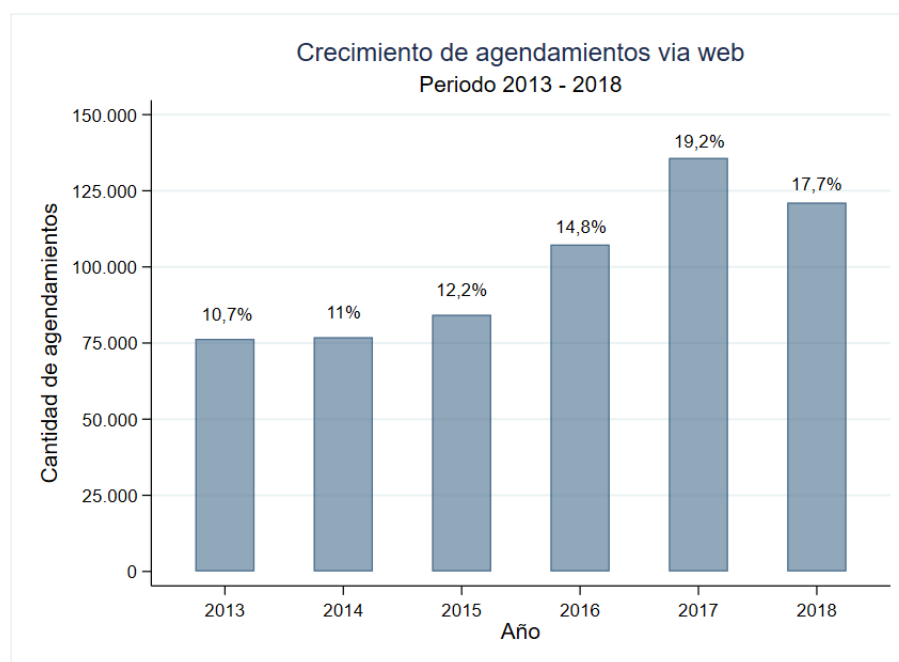


Figura 1.2: Agendamiento web anual de la institución
(Fuente: Elaboración propia)

como canal independiente y las llamadas salientes (outbound) mediante la confirmación o bloqueos de agendamientos realizados por algún canal de la institución siempre y cuando dicho agendamiento se encuentre dentro de la cartera de especialidades a gestionar por el área, dicha cartera es fija y su complemento es gestionado por otros departamentos.

Descripción del segmento

Para el año 2018, se observa que, del universo de consultas consolidadas de la institución, el 54,2 % de las citas médicas consolidadas corresponden a pacientes suscritos a FONASA, el 40,3 % a pacientes suscritos a ISAPRE y 5,5 % a otros medios. Se observa una homogeneidad en dichos ponderadores de forma mensual (ver figura 1.3).

Se observa que para el año 2018, aproximadamente el 63 % de las citas solicitadas corresponden a mujeres y el 37 % a hombres. Al evaluar la distribución etaria de los pacientes, se observa una concentración con tendencia homogénea en hombres con concentraciones entre 45 y 70 años, se muestran tendencias acentuadas en varones menores de 10 años. En el caso de las mujeres, se observa una concentración etaria de 25 a 65 años (ver figura 1.4).

Otro aspecto interesante a observar, es si el paciente posee email como medio de contacto, ya que de ser efectivo abriría nuevas líneas de investigación o márgenes de acción frente a gestiones de citas médicas. Se puede observar que, en el año 2013, en promedio el 50,7 % de los pacientes registraron un email al agendar una cita médica, el número de registros aumenta paulatinamente hasta llegar a un promedio de 66,3 % para el año 2018. Si se estudia el comportamiento de registro con la edad del paciente, se observa que 72 años es el punto de corte donde los pacientes tienden a no registrar su email en las citas médicas agendadas, lo que indica un nivel de comunicación a través de este medio para una amplia mayoría (ver figura 1.5).

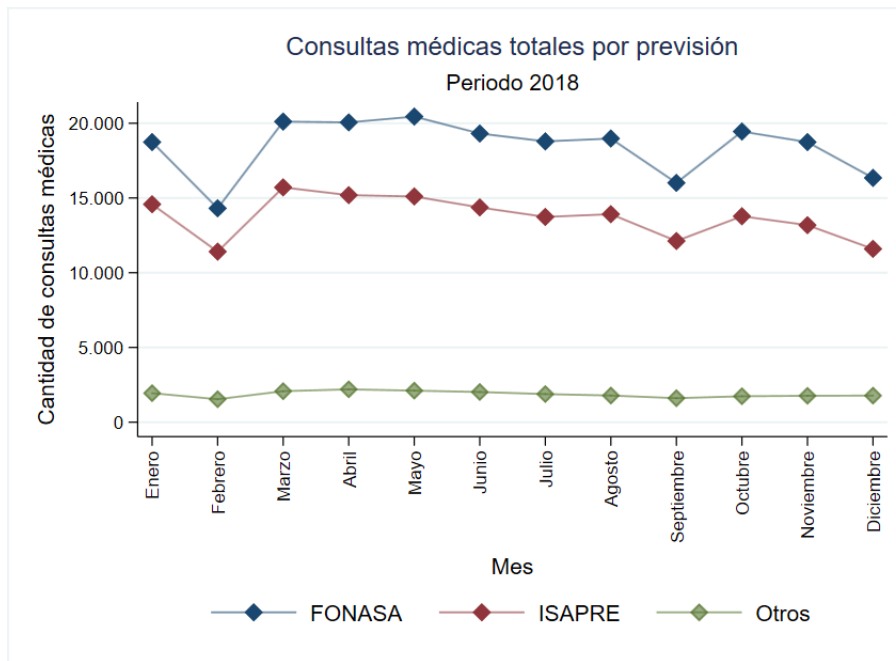


Figura 1.3: Consultas médicas totales por previsión de salud (Fuente: Informe estadístico HCUCH Diciembre 2018)

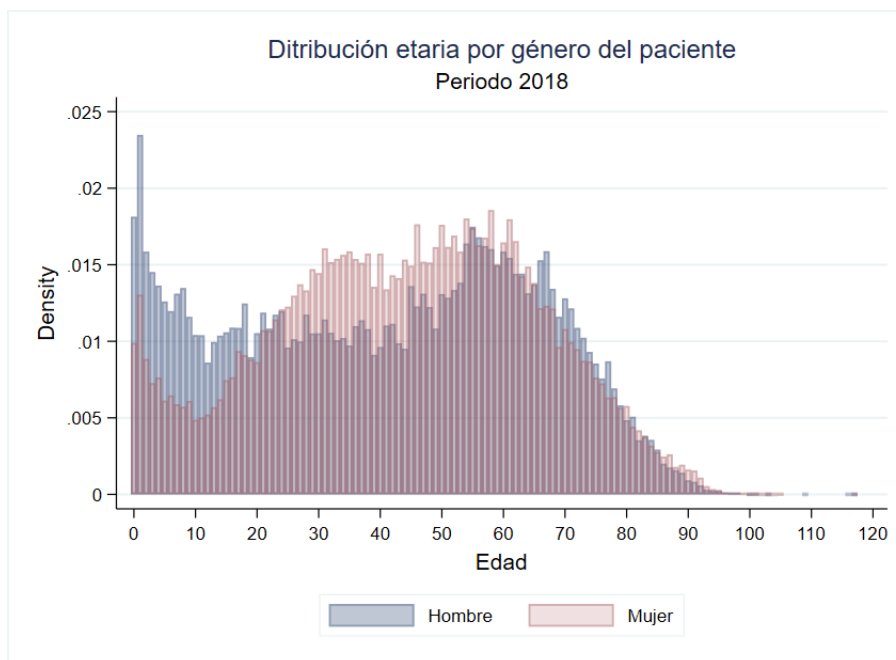


Figura 1.4: Distribución etaria por género del paciente para citas médicas solicitadas (Fuente: Elaboración propia)

1.2. Descripción del problema

Los sistemas de reserva poseen un riesgo latente si no son controlados de forma adecuada. En la industria de transporte se mantienen políticas establecidas acerca de la cancelación de la reserva, donde existe una motivación monetaria por parte del pasajero a informar su inasistencia y devolu-

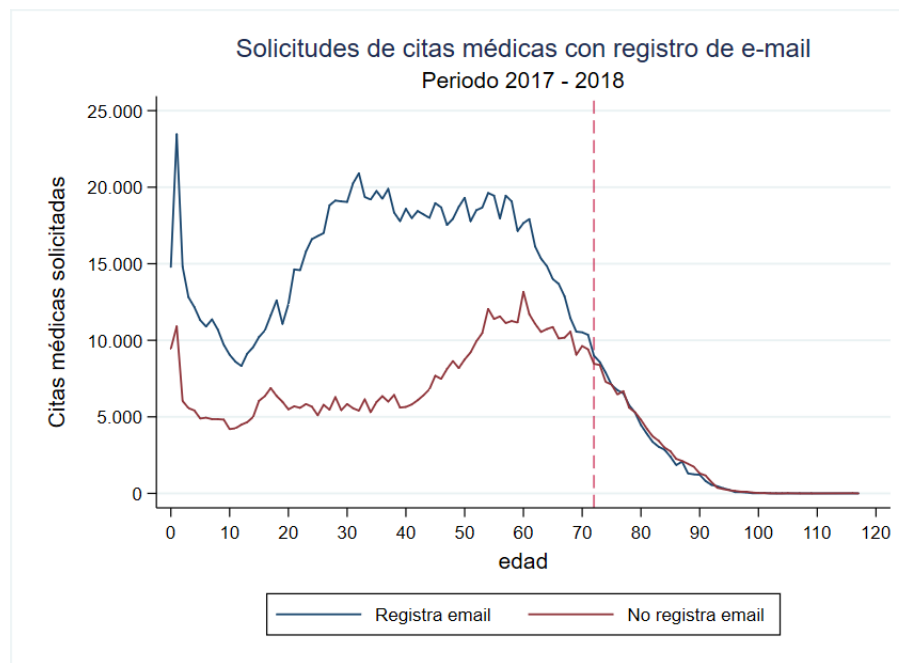


Figura 1.5: Solicitudes de cita médica con registro de email periodo 2017 - 2018
(Fuente: Elaboración propia)

ción de su boleto, de esta forma la empresa involucrada puede cubrir la gestión realizada por medio del porcentaje de retención del costo del pasaje. Si bien es cierto dicha retención cubre los costos de gestión, existen costos de oportunidad muy altos asociados a la compra del servicio disponible y su contribución a la cobertura de costos fijos del proceso. En la industria de aerolíneas les fue imprescindible evaluar otras alternativas para convertir dicho costo en aumento de los ingresos, políticas como la aplicación de overbooking en líneas aéreas han sido fundamentales para el crecimiento, un claro referente data de 1990, donde se expone la aplicación de overbooking en American Airline, donde el 40 % de los beneficios totales es producto de su aplicación [2], hoy en día la estrategia de overbooking es el pilar fundamental para la presencia de las aerolíneas en el mercado mundial.

Contrario a lo anterior es el caso de las atenciones médicas, donde el proveedor no tiene ningún seguro en caso de inasistencia de pacientes a sus citas agendadas (no-shows), no existe cobertura porcentual a causa de gestiones realizadas y las estrategias de sobrecupo están limitadas a la disposición del médico a establecer un número determinado de sobrecupos de acuerdo al momento de la solicitud. A menudo, se establecen sobrecupos fijos por iniciativas del médico, sin embargo, no existe una medida establecida dada la variabilidad intrínseca que conlleva dicha acción. La literatura demuestra que los sistemas de agendamiento de citas médicas presentan niveles de inasistencia con rangos de 5 % a 60 % en la práctica [12].

De acuerdo a la base de datos TICARES, para el año 2018, el 7,44 % de las citas médicas agendadas son anuladas y de las citas activas el 21,7 % de las consultas médicas ambulatorias agendadas en la institución no fueron concretadas debido a ausentismo del paciente (no-shows), excluyendo a los sobrecupos.

Para efectos del desarrollo del presente estudio, los problemas asociados al sistema de atención

ambulatoria se agrupan en dos tipos: En primer lugar, se encuentran los problemas asociados al ausentismo del paciente a la cita médica agendada y, en segundo lugar, se encuentran problemas relacionados al conocimiento del comportamiento del paciente frente a variaciones marginales de estados (análisis de contrafactuales), de esta forma generar políticas de acción beneficiosas para la institución y sin necesidad de testear las políticas para conocer su efecto en el nivel de asistencia. Lo anterior otorga análisis teórico previo acerca del impacto causado por variaciones de estados.

En primera instancia y de forma más explícita se encuentra el *costo de oportunidad de la cita no asistida*. Con el fin de cuantificar este costo, se estima lo siguiente: Si se consideran 400.000 consultas anuales asistidas y que el 21,7 % de las citas médicas agendadas corresponden a inasistencias, entonces, se estiman 110.856 consultas médicas no asistidas². Si el costo de una consulta ambulatoria de medicina general es de \$33.520 pesos [8], independientemente la forma de pago o seguro del paciente (FONASA, ISAPRE, etc.), el costo de oportunidad anual corresponde a 3.715.893.120. Es decir, al reducir un 1 % el porcentaje de inasistencias (de 21.7 % a 20.7 %), se aumentarían aproximadamente 171 millones de pesos anualmente a los ingresos del hospital, tan sólo considerando el perjuicio del costo de oportunidad. Otro impacto no menor y en la misma línea del anterior se encuentra en el propósito del hospital, el cual es entregar soluciones de salud a la comunidad, es así como el comportamiento de inasistencia genera perjuicio a personas que eventualmente requieren solucionar su problema de salud y no tuvieron la oportunidad debido a la utilización del cupo disponible.

Otro efecto adverso ocasionado por el ausentismo, corresponde al impacto en la oferta médica:

- Para las citas médicas institucionales, el impacto de inasistencias se traduce en tiempos muertos (Idles Times) de especialistas, dado que este tipo de citas médicas se encuentran asociadas a horarios de trabajo establecidos de forma contractual.
- Para citas médicas no institucionales, el impacto está asociado a la oferta médica, ya que este tipo de citas tienen una distribución de ingreso variable (Dado que es horario fuera del establecido en contrato, el beneficio es mediante un porcentaje de la atención) y sujeto a la cantidad de atenciones efectivas (pagadas). Si los pacientes no asisten a las citas, las motivaciones por generar oferta por parte de los médicos (agendas que involucren horarios fuera de contrato) en el hospital se reducen, generando necesidad de profesionales (ya que los actuales eventualmente prefieren cambiarse de institución (al menos en el horario no institucional o fuera de contrato) a una institución con mejor capacidad de captación de pacientes o a sus propias consultas médicas), reduciendo de esta forma los ingresos de la institución.

Como parte del impacto asociado al manejo de información teórica, se encuentra la creación de políticas de gestión adecuadas que requieren de un análisis profundo de la data. Hoy en día, se están evaluando las siguientes políticas en el departamento:

- Extensión de la agenda médica con el fin de capturar mayor cantidad de pacientes y así evitar desechar posibles candidatos por no tener mayor cantidad de horas disponibles en el tiempo (juicio que requiere ser fundado mediante análisis numérico). Un testeado de este cambio de política requeriría demasiado tiempo, lo que se traduce en un riesgo si corresponde a una política errada.
- Dado que la literatura señala que existe un aumento en la probabilidad de asistencia del paciente al ser confirmado, se estableció la gestión de confirmación para ser realizada un día

²Detalle de la estimación: $400000(0,217/(1 - 0,217)) = 110856$

antes de la cita médica y sin límites de intentos (a mayor probabilidad de confirmación, mayor probabilidad de asistencia). Lo anterior genera una brecha importante para comportamientos deshonestos tales como llamar una gran cantidad de veces a un número mal ingresado para lograr una aparente confirmación. De esta forma el agente aumenta su sueldo variable (relacionado a la cantidad de intentos de confirmación) sin generar beneficios a partir de su gestión.

1.3. Justificación del proyecto

La importancia de comprender la información asociada a la cita médica y en particular la información del paciente radica en la segmentación y selección de alternativas para proveer el mejor modelo de gestión de información y de servicio. En primer lugar, conocer al segmento objetivo, en específico, conocer el impacto de sus decisiones y acciones, por ejemplo, saber el grado de fiabilidad de las declaraciones del paciente, evaluaciones teóricas de políticas de gestión, evaluar de forma transversal los efectos causados por variables exógenas tales como el clima o la inflación y cómo su comportamiento afecta a la institución compone un conjunto de información crucial para la toma de decisiones, información que hoy en día es desconocida para la institución.

Por otro lado, la aplicación práctica del conocimiento atañe directamente a predecir el comportamiento del paciente, evaluar nuevos sistemas de gestión para la optimizar los recursos disponibles y como dicho comportamiento va cambiando a través del tiempo. Crear políticas y medidas de eficiencia que sean sostenibles en el tiempo es un desafío, y para ello es necesario establecer parámetros y medidas de acción estratégicas basados en datos y predicciones fiables, de esta forma el actuar de la institución será controlado y no reactivo.

Desde el punto de vista práctico, el análisis de datos y creación de un modelo factible incentiva la creación de políticas de acción. Desde el punto de vista tecnológico, el estudio ayudará a identificar y almacenar variables fundamentales para la predicción del comportamiento, tales como fiabilidad de respuesta del paciente junto con evaluar la posibilidad de captura de nueva información asociada al paciente, que eventualmente podría agregar mayor nivel de predicción al modelo o entregar información adicional para la toma de decisiones de distintas áreas.

Finalmente, las acciones mencionadas alinearán los objetivos de los stakeholders con los de la institución, generando disminución de malas prácticas e incentivando la oferta de citas médicas y retención de pacientes/especialistas.

Capítulo 2

Marco Teórico

En el siguiente capítulo se abordarán las principales bases teóricas acerca de las herramientas utilizadas en el presente estudio. Este capítulo tiene por objetivo establecer información contextual para presentar al lector los procedimientos utilizados en cada análisis y su interpretación. Se abordarán además temáticas asociadas a la metodología utilizada, al ajuste del modelo y herramientas complementarias para lograr su aplicación en los procesos requeridos.

2.1. Metodología KDD (Knowledge Discovery in Databases)

KDD es una metodología presentada por Fayyad (1996), el proceso es interactivo e iterativo (ver figura 2.1) y se compone principalmente por siete pasos fundamentales [17]:

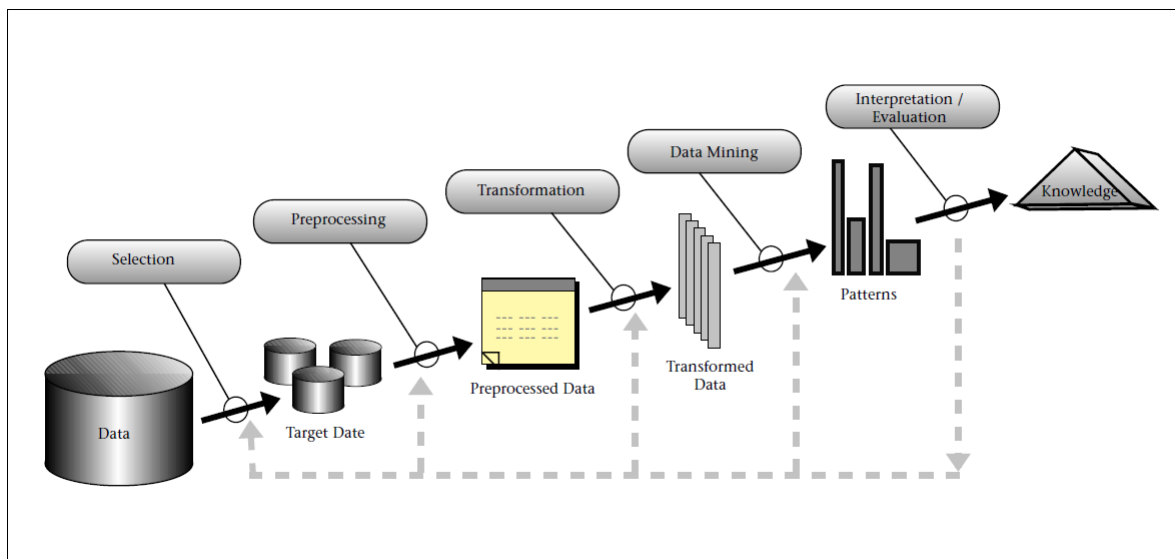


Figura 2.1: Pasos que compone el proceso de KDD [13]

1. Entender el dominio de aplicación, determinar el problema a resolver y los objetivos del proyecto, todo desde el punto de vista del cliente.

2. Selección e integración de un conjunto de datos objetivos provenientes de fuentes heterogéneas, teniendo en cuenta el objetivo del proyecto. Homogeneizar formatos y descartar variables irrelevantes.
3. En la etapa de limpieza y pre procesamiento se deben tomar decisiones acerca del proceder con datos faltantes, atípicos, erróneos, etc. Evaluar la forma de eliminar inconsistencias y valores duplicados. La etapa de preparación y limpieza es a veces una etapa descuidada, pero de suma importancia en este proceso, dado que grandes cantidades de datos son recolectados por medio de métodos automáticos. A veces el método por el cual los datos fueron obtenidos no fue cuidadosamente controlado, y así los datos podrían contener valores fuera de rango (por ejemplo, edades negativas), combinaciones incorrectas de datos, y otros.
4. Encontrar características útiles para representar a los datos dependiendo de los objetivos. Reducción de dimensiones para llevar adelante el trabajo con un número reducido de variables. Eliminar columnas dependientes, tales como fecha de nacimiento y edad, reducir la cantidad e información en memoria activa, optimizando los procesos de almacenamiento dinámico de resultados.
5. Corresponde al análisis exploratorio, escoger las herramientas de data mining adecuadas al problema a resolver, teniendo en cuenta el objetivo (predecir, explicar, clasificar, agrupar, etc.) y comprobar requisitos o supuestos base para la utilización de la herramienta y finalmente utilizarla. Una vez realizada la tarea, se procede con el descubrimiento de patrones y relaciones en los datos, para presentárselos al usuario de una manera adecuada (gráficos, árboles, reglas, etc.).
6. Interpretación de los datos, llevada a cabo por el analista. Desde este punto del proceso es posible regresar a cualquiera de los pasos anteriores.
7. En esta etapa se debería consolidar el conocimiento ganado, probando los modelos creados contra los resultados obtenidos de la aplicación de estos modelos en el mundo real.

2.2. Modelos estructurales

Se diferencian dos focos claros para la creación de modelos: El primero se le denomina *Modelo estructural*, el cual se basa en teoría para derivar una especificación econométrica que puede ser estimada sobre datos, por ejemplo, consideración de agentes racionales, tendencias del comportamiento del paciente, entre otros. El segundo foco se denomina *Modelos de forma reducida*, los cuales tienen un foco puramente estadístico, sin base teórica y se centra netamente en el ajuste de la data. Es claro que, para lograr el doble objetivo de predecir y explicar, es necesario mantener una base teórica de los factores incidentes en el comportamiento del paciente, de esta forma, para efectos del presente estudio el foco adecuado a abordar corresponde al modelo estructural [15].

2.2.1. Modelos de elección discreta¹

De acuerdo a [25], los modelos de elección discreta describen las elecciones que los decisores hacen entre diferentes alternativas. Para encajar en un marco de elección discreta, el conjunto de

¹Para mayor información, ver *Apéndice A: Anexo metodológico*

alternativas, llamado conjunto de elección, tiene que presentar tres características. En primer lugar, las alternativas deben ser *mutuamente excluyentes* desde el punto de vista del decisor. Escoger una alternativa necesariamente implica no escoger ninguna de las alternativas restantes. En segundo lugar, el *conjunto de elección debe ser exhaustivo*, en el sentido de que todas las posibles alternativas deben estar contempladas. En tercer lugar, el *número de alternativas debe ser finito*.

2.2.2. Modelo Logit

El modelo Logit resulta de asumir que cada variable de error ε_{nit} es independientemente distribuido de acuerdo a una distribución *Gumbel* o de valor extremo tipo I [15].

$$F(\varepsilon_{nit}) = e^{-e^{-\varepsilon_{nit}}} \quad f(\varepsilon_{nit}) = e^{-\varepsilon_{nit}} \cdot e^{-e^{-\varepsilon_{nit}}} \quad (2.1)$$

2.1: Función de distribución tipo Gumbel sobre ε_{nit}

Aplicando la ecuación 2.1, se puede demostrar que la probabilidad de elección en un modelo Logit corresponde a la formula cerrada presentada en la ecuación 2.2.

$$\begin{aligned} P_{nit} &= \int_{\varepsilon} I(\varepsilon_{njt} - \varepsilon_{nit} < V_{nit} - V_{njt} \quad \forall i \neq j | \varepsilon_{nit}) f(\varepsilon_{nit}) d\varepsilon_{nit} \\ &= \int_{\varepsilon} \left(\prod_{i \neq j} e^{-e^{-(V_{nit} - V_{njt} + \varepsilon_{nit})}} \right) e^{-\varepsilon_{nit}} e^{-e^{-\varepsilon_{nit}}} d\varepsilon_{nit} \\ &= \frac{e^{V_{nit}}}{\sum_j e^{V_{njt}}} \end{aligned} \quad (2.2)$$

2.2: Probabilidad de elección de un modelo Logit

Dado que para el caso de estudio la elección es dicotómica, además se cumple que la diferencia de dos variables de tipo valor extremo distribuye de forma logística, entonces la probabilidad de elección se simplifica como muestra la ecuación 2.3.

Si $\varepsilon_{nit}, \varepsilon_{njt} \sim Gumbel$ iid., entonces $\varepsilon_{njt}^* = \varepsilon_{njt} - \varepsilon_{nit} \sim Logistica$.

$$P_{nit} = \frac{e^{V_{nit}}}{1 + e^{V_{nit}}} \quad (2.3)$$

2.3: Probabilidad de elección asociada a una transformación lineal para normalizar la utilidad

El modelo anterior es altamente utilizado dada su simplicidad y estructura cerrada (explicita), de esta forma, se pueden obtener probabilidades de ocurrencia tan solo evaluando las características del evento en la función de utilidad ajustada V_{nit} . Esta última es obtenida a partir del método de Máxima verosimilitud [25].

2.2.3. Modelo Probit

El modelo Probit resulta de asumir que el vector de errores ε_{ni} se distribuye por una normal multivariada centrada en cero ($\varepsilon_{ni} \sim N(0, \psi)$). La normalidad de los errores provee bastante flexibilidad para acomodar una amplia variedad de estructuras de las preferencias. Un modelo con errores normales permite acomodar factores sistemáticos no observables en la utilidad. Formalmente, el supuesto de normalidad de la componente aleatoria de la utilidad implica que su función de densidad viene dada por la ecuación 2.4, donde ψ corresponde a la varianza-covarianza de tamaño $I \times I$, donde I es el número de alternativas disponibles para el tomador de decisión [15].

$$\phi(\varepsilon_n) = \frac{1}{(2\pi)^{I/2} |\psi|^{1/2}} e^{-\frac{1}{2} \varepsilon_n' \psi^{-1} \varepsilon_n} \quad (2.4)$$

2.4: Probabilidad de elección asociada a una transformación lineal para normalizar la utilidad

De esta forma, la probabilidad que el individuo n elija la alternativa i viene dada por la ecuación 2.5.

$$\begin{aligned} P_{ni} &= Prob(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \quad \forall i \neq j) \\ &= \int_{\varepsilon_n} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \quad \forall i \neq j) \phi(\varepsilon_n) d\varepsilon_n \end{aligned} \quad (2.5)$$

2.5: Probabilidad que el individuo n escoja la alternativa i mediante la función de enlace Probit

A diferencia del modelo Logit, la integral sobre la densidad $\phi(\cdot)$ no tiene primitiva analítica y por lo tanto, no se dispone de una fórmula cerrada para P_{ni} , por lo que su ajuste debe ser a través de simulación por métodos numéricos.

2.2.4. Modelo Logit Mixto

Si se asume que los parámetros del modelo siguen una distribución con densidad $f(\beta)$ a lo largo de la población, para obtener la probabilidad incondicional P_{ni} , se deberá integrar la probabilidad de elección del modelo Logit estándar a lo largo de todos los posibles valores de

β (ver ecuación 2.6) [15].

$$P_{ni} = \int \left(\frac{e^{V_{ni}(\beta)}}{\sum_j e^{V_{nj}(\beta)}} \right) f(\beta) d\beta \quad (2.6)$$

2.6: Probabilidad que el individuo n escoja la alternativa i mediante Mixed Logit

Donde $V_{ni}(\cdot)$ es la porción observable de la utilidad del individuo n al elegir la alternativa i . Si se asume linealidad en la función de utilidad, la formulación toma la estructura de la ecuación 2.7.

Al igual que el modelo anterior, el modelo Logit Mixto no posee una fórmula de probabilidad cerrada. La principal ventaja de utilizar este modelo subyace en su capacidad para ajustar el modelo utilizando variables aleatorias asociadas a comportamientos no observables, es decir, establece

$$P_{ni} = \int \left(\frac{e^{\beta' x^{ni}}}{\sum_j e^{\beta' x^{nj}}} \right) f(\beta) d\beta \quad (2.7)$$

2.7: Probabilidad que el individuo n escoja la alternativa i mediante Mixed Logit asumiendo linealidad

medidas de heterogeneidad (comportamiento no observable) en cada caso, lo cual resulta extremadamente beneficioso para el modelo estructural con foco explicativo (debido a que establece medidas de probabilidad de ocurrencia más precisas).

2.2.5. Interpretación de modelos

Los modelos Logit y Mixed Logit se pueden interpretar mediante variadas técnicas:

Odds Ratios [21] Un Odds Ratio (OR) es una medida de asociación entre una exposición y un resultado. Se utiliza principalmente para comparar las probabilidades relativas de ocurrencia de un resultado de interés, dada la exposición a la variable de interés. También se puede usar para determinar si una exposición en particular es un factor de riesgo para un resultado en particular, y para comparar la magnitud de varios factores de riesgo para ese resultado. De esta forma:

- Para $OR = 1$ La exposición no afecta las probabilidades de resultados
- Para $OR > 1$ Exposición asociada con mayores probabilidades de resultados
- Para $OR < 1$ Exposición asociada con menores probabilidades de resultados

Efectos marginales [15] [18] Los efectos marginales corresponden a tasas de cambio instantáneo; se calcula un predictor para el efecto, mientras que los otros predictores en el modelo se mantienen en un valor dado, por lo general en su media. En la ecuación 2.8, se puede observar el efecto sobre la probabilidad de escoger la alternativa i al variar alguna componente de la utilidad de la misma alternativa. De igual forma, en la ecuación 2.9 se muestra el efecto producido sobre la probabilidad de escoger i al variar alguna componente de la alternativa j .

$$\frac{dP_{nit}}{dx_{nit}} = \frac{\partial V_{nit}}{\partial x_{nit}} \cdot P_{nit}(1 - P_{nit}) \quad (2.8)$$

2.8: Variación de efectos marginales producidos por la variación de alguna componente de la misma alternativa

$$\frac{dP_{nit}}{dx_{njt}} = \frac{\partial V_{njt}}{\partial x_{njt}} \cdot P_{nit} \cdot P_{njt} \quad (2.9)$$

2.9: Variación de efectos marginales producido por la variación de alguna componente de otra alternativa

Otras aplicaciones Existen otras medidas enfocadas en la variación discreta de ciertos parámetros, evaluando el impacto en la probabilidad del evento, algunas técnicas son: Adjusted Predictions at the Means (APM), evalúa predicciones sobre las medias de las covariables bajo la variación de la

covariable de estudio. Average Adjusted Predictions (AAP's) evalúa el impacto promedio generado bajo variaciones de la covariable de estudio dejando el resto de los factores intactos.

2.2.6. Evaluación de bondad de ajuste

Pseudo- R^2 (de McFadden) Está basado en el logaritmo de la función de verosimilitud (ver ecuación), donde $l(\hat{\beta})$ es el logaritmo de la función de verosimilitud para el modelo estimado y $l(\hat{\beta}_0)$ el de un modelo sólo con termino constante. Como $|l(\hat{\beta})| < l(\hat{\beta}_0)$, el $Pseudo - R^2$ se encuentra entre 0 y 1, tomando el valor de 1 para el mejor ajuste teórico. Es importante destacar que McFadden mediante la expresión *Rule of Thumb* señala que si se cumple: $0,20 \leq McFaddenPseudo - R^2 \leq 0,40$, el modelo presenta un excelente ajuste [19].

Criterio de información de Akaike (AIC) y Bayesiano (BIC) [15] Los dos indicadores más usados para comparar modelos con distintos números de parámetros son AIC y BIC en que se penaliza la verosimilitud por el número de parámetros para capturar el hecho que al incluir nuevos parámetros la verosimilitud necesariamente crecerá. La diferencia entre AIC y BIC es que el primero tiene una penalización constante por número de parámetros mientras que la penalización del segundo depende de la cantidad de data disponible. Si la log-verosimilitud de un modelo con n observaciones y k parámetros es LL entonces AIC y BIC vienen dados por la ecuación 2.10

$$AIC = -1LL(\hat{\theta}) + 2k \quad BIC = -2LL(\hat{\theta}) + k \ln(n) \quad (2.10)$$

2.10: Criterios de información AIC y BIC

Capítulo 3

Estructura de la investigación

3.1. Objetivos

3.1.1. Objetivo General

Desarrollar un modelo predictivo con el fin de reducir la tasa de inasistencia a citas médicas al menos en un 1 %¹ en el Hospital Clínico de la Universidad de Chile.

3.1.2. Objetivos Específicos

Objetivo específico 1

Evaluar variables endógenas y levantar base de datos de variables exógenas para un ajuste adecuado del modelo.

Objetivo específico 2

Caracterizar las variables explicativas de la asistencia a una cita médica agendada.

Objetivo específico 3

Ajustar distintos modelos basados en variables endógenas y exógenas, y determinar el modelo más adecuado, evaluar ajuste de modelos con heterogeneidad no observable para una correcta especificación.

Objetivo específico 4

Evaluar el comportamiento de las variables de acuerdo a sus sensibilidades y obtener probabilidades promedios y marginales.

¹La reducción de la tasa de inasistencia de citas médicas en un 1 % equivale a reducir la tasa de 21.7 % a 20.7 %, logrando un aumento de aproximadamente 5109 asistencias anuales (ver sección 1.2. *Descripción del problema*).

Objetivo específico 5

Aplicar el estudio a la evaluación de políticas actuales. Evaluar la eventual inserción del modelo predictivo ajustado en la gestión de citas.

3.2. Hipótesis de investigación

Las hipótesis de investigación, dada su naturaleza, se encuentran anidadas; en primer lugar, se validarán hipótesis asociadas al estudio y creación del modelo, luego se validarán hipótesis asociadas a su aplicación:

Hipótesis asociadas al modelo

1. Existen diferencias entre las probabilidades de asistencia entre pacientes de distinto género
2. Los pacientes de edad avanzada tienen mayores probabilidades de asistencia que los pacientes de menor edad
3. Las citas médicas asociadas a especialidades determinadas tienen mayores probabilidades de asistencia que medicina general
4. Los pacientes que agendan vía web tienen menos probabilidades de asistencia que otra vía de agendamiento
5. Los sobrecupos afectan positivamente a la probabilidad de asistencia del paciente
6. La gestión de confirmación genera un impacto positivo en la asistencia del paciente
7. A mayor diferencia entre la fecha de solicitud de la cita y la fecha de atención mayor es la probabilidad de asistencia, ya que el paciente eventualmente organizaría mejor sus tiempos y cumpliría con su compromiso.
8. Los pacientes que agendan una cita médica para los días sábado tienen menor probabilidad de asistencia, debido a que es fin de semana.
9. Los pacientes que agendan una cita médica para atenciones después de las 18:00, tienen menos probabilidades de asistir.
10. La confirmación de la cita médica por parte del paciente y su posterior inasistencia, tiene que ver con eventualidades del momento y no es producto de su comportamiento habitual
11. Una cita médica agendada con el mismo médico de la última cita asistida, posee mayores probabilidades de asistencia que una cita médica cuya última vez fue asistida con un médico diferente (intenta demostrar la fidelidad del paciente con su médico de tratamiento).
12. Mientras más lejos se encuentre el hogar del paciente, menor es la probabilidad de asistir a su cita agendada
13. Mientras más demore el paciente en llegar al centro médico, menor es la probabilidad de

asistir a la cita agendada

14. La temperatura y precipitaciones afectan el comportamiento de asistencia del paciente a su cita médica agendada
15. Las manifestaciones en Santiago afectan a la asistencia del paciente a su cita médica agendada
16. Los eventos deportivos masivos, como juegos de fútbol nacional afectan el comportamiento de asistencia del paciente a su cita médica
17. Las consultas post operatorias tienen una fuerte relación con la asistencia dada su importancia intrínseca
18. Es posible predecir el comportamiento de un paciente a partir de un modelo estructural

Hipótesis asociadas a la aplicación del modelo

1. Aperturas de agendas más extensas (agendar a mayor periodo de anticipación) mejora la captura de pacientes y por ende involucra mayores beneficios como política de gestión
2. El sistema de gestión actual posee brechas de acción asociada a malas prácticas producto de políticas mal especificadas
3. La aplicación de un modelo predictivo otorga una solución a los problemas de inasistencia a citas médicas agendadas

3.3. Metodología

Para el propósito del presente estudio, existen una serie de metodologías adecuadas para el desarrollo del modelo, tales como KDD (Knowledge Discovery in Databases), SEMMA (acrónimo de las fases: Sample, Explore, Modify, Model, Assess), CRISP-DM (Cross-Industry Standard Process for Data Mining), entre otras. Sin embargo, se decide utilizar la metodología KDD² dado que, a diferencia de las metodologías anteriores, posee una fase explícita de interpretación de data junto a la evaluación del modelo, lo cual es una ventaja al evaluar el foco explicativo y predictivo del modelo estructural. De esta forma, el presente estudio se desarrollará sobre la base de una variación de KDD.

La variación estructurada a esta metodología se basa principalmente en la adaptación de una estrategia de modelización, a los pasos 5 y 6 de la metodología, generalmente utilizada para la creación de modelos de ecuaciones estructurales llamada *Estrategia de desarrollo del modelo*[27].

La estrategia de desarrollo del modelo tiene como propósito mejorarlo de forma continua. El investigador que sigue esta estrategia intenta re especificar un modelo básico de teoría para encontrar uno nuevo.

Comienza con la definición de un modelo teórico, el cual se intentará contrastar estadísticamente, y finaliza con una verificación parcial o global del modelo (ver figura 3.1).

²Ver Capítulo 2: Marco Teórico, sección 2.1 Metodología KDD

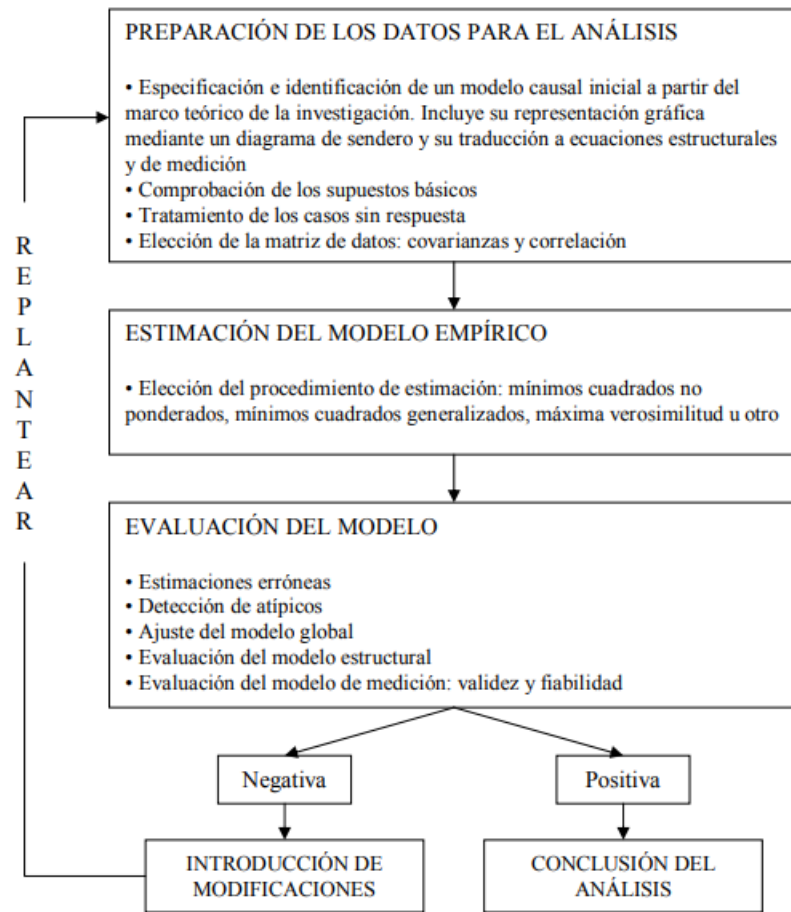


Figura 3.1: Fases principales en la obtención de un modelo de ecuaciones estructurales [27]

La falta de confirmación de una o varias hipótesis establecidas en la elaboración del modelo de ecuación estructural conlleva a la modificación de este, manteniendo un significado lógico-sustantivo, además de una significatividad estadística.

Toda modificación del modelo inicial lleva forzosamente a la repetición de todo el análisis estadístico.

El proceso de análisis acaba en el momento en el que se consigue un modelo con un nivel de ajuste igual o superior al propuesto por el investigador.

3.4. Resultados esperados

Por medio del presente estudio se espera obtener los siguientes resultados:

- Establecer probabilidades de asistencia sujetas a factores de investigación asociados (factores endógenos y exógenos)
- Entender el impacto producido por cada factor testado en el nivel de asistencia de las citas médicas por medio de probabilidades promedio, marginales y conjuntas.

- Comprender y dimensionar los efectos producidos por las variaciones en los factores estudiados sobre la tasa de asistencia de pacientes.
- Aplicar las probabilidades obtenidas a políticas de gestión.
- Entender y cuantificar el beneficio futuro si se integrara el modelo predictivo en el proceso de gestión de citas médicas.

3.5. Alcances

Los alcances del presente estudio corresponden a la obtención de un modelo estructural con foco predictivo, que demuestre ser robusto a la hora de estimar el comportamiento de asistencia de los pacientes (testear distintos modelos y establecer resultados a partir del modelo que presente mejor ajuste). Se espera además obtener probabilidades promedios y marginales asociadas a las distintas variables del modelo, por ejemplo, probabilidad promedio de asistencia de un sobrecupo, probabilidad marginal de asistencia de una cita médica confirmada, etc. A partir de lo anterior, validar distintas hipótesis asociadas a la toma de decisiones.

Es importante destacar que el presente estudio sólo sugiere la inclusión del modelo e interpretación de resultados, no establece una metodología de adaptación tecnológica al software actual ni comprueba su factibilidad técnica, dado que no existen conocimientos suficientes acerca del modelo, estructuración y programación de la base de datos utilizadas actualmente para la gestión de citas médicas.

Capítulo 4

Tratamiento y Análisis de datos

4.1. Bases de datos

La extracción de datos fue realizada desde distintas fuentes de información categorizadas por la proveniencia como Fuentes Internas y externas de información.

4.1.1. Fuentes internas de información

HCUCH

HCUCH es un software de gestión de citas médicas creado por la institución, es utilizado de forma paralela a TICARES para estos objetivos, sin embargo, con el fin de conseguir una mejora tecnológica y práctica en el proceso, se decide dejar de utilizarlo en el año 2017, luego comienza la utilización de AppContact. La información extraída del software comprende información histórica a partir del año 2007.

TICARES

TICARES es un software que gestiona la ficha electrónica del paciente, se utiliza de forma paralela a un software de gestión de citas médicas dado que almacena la información personal del paciente tal como su dirección, información de contacto, etc. la cual debe estar constantemente actualizada.

AppContact

AppContact es un nuevo software de gestión de citas médicas implementado con el fin de reemplazar a HCUCH y mejorar el sistema de gestión e información del departamento, dicho software comienza su implementación en enero de 2017 y ha sido utilizado con éxito hasta ahora. Las principales funcionalidades que clasifican a este sistema corresponde a la forma de apertura de agenda, las cuales se categorizan por campañas de apertura, dichas campañas son gestionadas, extrayendo estadísticas asociadas a las gestiones destinadas al departamento, es importante mencionar que el departamento posee un portafolio de especialidades a gestionar, las que serán especificadas en la sección 5 con la descripción del proceso.

OMEGA

Crossnet es una empresa que otorgaba a la institución servicios de almacenamiento y gestión de información por medio de Hyperion de Oracle en la nube, sin embargo, dicha relación se detuvo a inicios del año 2018, donde se reemplaza el sistema outsourcing con un sistema propio de la institución, es aquí donde nace OMEGA. OMEGA es un sistema de almacenamiento y gestión de información asociado a las llamadas, esto es, almacenamiento de grabaciones, registro de tiempos de llamadas inbounds, outbounds, tiempos de espera, estados de llamadas, agentes, departamentos, etc. De esta forma, el sistema es fundamental para realizar gestiones de calidad de atención, niveles de eficiencia de operadores y toma de decisiones para el área.

4.1.2. Fuentes externas de información

Para la obtención de variables exógenas del modelo, se recurre a fuentes de información externa:

Meteorología

Para obtener información acerca de la temperatura y precipitación, se recurre a fuentes de información meteorológica de Chile, específicamente se recurre a información registrada por la estación meteorológica [5] 330021 ubicada en Pudahuel - Santiago. Es importante destacar que la información disponible corresponde a promedios diarios de temperatura en Santiago y niveles de precipitación diarios.

Manifestaciones sociales

Para el caso de manifestaciones sociales, dado que no fue posible obtener bases de datos asociadas, fue necesario realizar una investigación independiente en información histórica presentada en periódicos de Chile, específicamente se investiga la información asociada a manifestaciones realizadas en Santiago correspondiente al periodo 2013 al 2018 en el portal web Emol [1], donde es posible filtrar la búsqueda por conceptos asociados a sucesos, de esta forma, se fue virtualizando manualmente una base de datos asociada a los días de manifestaciones en Santiago.

Costo de vida

Dado que no es posible obtener características socioeconómicas de cada paciente, se decide recurrir a la variación macro del costo de vida de las personas en Chile, es así como se recurre a la obtención de la variación empalmada del IPC mensual por medio de la base de datos del Banco Central [4], de esta forma obtener una aproximación de sensibilidad asociada a la respuesta del paciente frente a alzas en el costo de vida.

Eventos deportivos

Dado que existe un cambio en el comportamiento de las personas asociado a eventos deportivos masivos de competencias entre países, el cual es exacerbado con la publicidad que este tipo de eventos frecuentemente lleva consigo, fue imprescindible analizar dichos eventos con el fin de capturar el efecto que genera sobre el comportamiento de un paciente, tales como partidos de la selección Chilena de fútbol. Para ello se recurre a una base de datos de la web solofutbol.cl [24],

donde se encuentra una base de datos de los partidos de fútbol de la selección chilena a partir del año 1910 hasta la fecha.

Días feriados

Para capturar el comportamiento de un paciente en días previos a un feriado, feriado y post feriado, se recurre a la base de datos de feriados.cl [14], la cual señala los feriados históricos de Chile.

Distancia y tiempo de viaje

A pesar de que las bases de datos internas poseen las direcciones de los usuarios, dicha información no es útil si no se tiene la distancia entre la dirección y el centro asociado a la cita médica o el tiempo de viaje. Para ello fue necesario recurrir a la extracción de aproximadamente 290.000 pacientes de la base de datos TICARES, de los cuales se extrajo su dirección particular para generar una conexión con GOOGLE CLOUD [16] y levantar una base de datos de distancias reales¹ entre el centro asociado a la cita médica y la dirección particular del paciente. En cuanto al tiempo, se decide levantar información asociada al tiempo de viaje utilizando transporte público.

4.2. Tratamiento de data

4.2.1. Selección de pacientes

El presente estudio demanda analizar el comportamiento del paciente, para ello una de las fases más importantes tiene que ver con el número de pacientes a estudiar y si mantiene información de su comportamiento histórico, es así como una de las medidas que requiere dicha selección y que acota la información de acuerdo a su transformación son las variables distancia y tiempo de viaje. Para ello se recurre a Web Scraping.

El Web Scraping es un proceso que implica la recuperación de información semi estructurada de Internet, generalmente una página web se encuentra en un lenguaje demarcado como HTML o XHTML, y el análisis de esa información es utilizado para extraer datos específicos para su uso en otro contexto [26].

Para el presente estudio se recurrió a la programación de web Scraping mediante CURL en PHP. CURL es una librería que permite realizar peticiones HTTP para transferir información. En la práctica, permite simular a un usuario navegando por internet y realizando peticiones a otros servidores.

A partir de mediados de 2018, Google comienza a controlar las peticiones de GoogleMap de forma gratuita, todo esto debido a que lanza Google Cloud al mercado, es así como se puede acceder a la información de forma programada sólo si quien programa dicha simulación posee una llave de acceso, a menudo los intentos sin llave de acceso quedan bloqueados, debido a que Google capta un comportamiento no habitual de navegación y lista a la IP utilizada en su "*black list*" del servidor,

¹Se define distancia real como la distancia entre dos puntos que incluye obstáculos, calles y sentidos del tránsito. El beneficio de utilizar dicha información es que a diferencia de la distancia obtenida a menudo con la fórmula de Haversine utilizando dos coordenadas geográficas, la distancia real posee un grado superior de exactitud, lo que ayuda significativamente la predicción del modelo en caso de ser una variable significativa.

esto queda reflejado al solicitar en cada petición de búsqueda un *reCAPTCHA* para demostrar que quien navega no es una simulación, sino un usuario real. Es así como al adquirir la llave de acceso desde Google Cloud, se programa un script en PHP con el objeto de leer la dirección del paciente en una base de datos local, leer la dirección del centro correspondiente y enviar la consulta por medio del script. La consulta consiste en obtener dos variables que podrían ser incidentes en el comportamiento del paciente:

1. Dirección: Corresponde a la distancia real entre la dirección personal del paciente y la dirección del centro correspondiente.
2. Tiempo: Corresponde al tiempo que demora el paciente en asistir a la cita médica en caso que utilice transporte público, este último es calculado para el horario de mayor congestión vehicular (pero caso).

Es importante notar que la correlación de ambas variables podría ser un problema, sin embargo, la diferencia fundamental es que la distancia es fija y óptima y el tiempo es variable dado que se enfoca sólo en transporte público (podría eventualmente utilizar otra dirección de viaje, dado que este tipo de transporte posee un recorrido determinado, sin mencionar que en ocasiones se deben realizar trasbordo entre medios de transporte público como bus, metro o simplemente caminar. Si se toma esta variabilidad en cuenta, las variables podrían no poseer fuerte correlación.

De esta forma se crea un script iterativo que realiza aproximadamente 290.000 consultas, correspondiente a la base de datos de pacientes de TICARES para el periodo 2017-2018, de esta forma, se selecciona a pacientes que han tenido contacto reciente con el servicio, obteniendo información histórica del comportamiento de dicho paciente (dado que la información recolectada data del año 2007).

El proceso de requerimientos de información se especifica a continuación:

1. Programar una función de solicitud y ejecutar programa
2. El código consultará la dirección de un ID de usuario en una base de datos local
3. Se formatea la información y se envía la petición junto con la key de acceso
4. Google Cloud procesa los datos de petición
5. Se recibe la respuesta y se formatea la información
6. Se almacena la información de recibida en una base de datos local
7. El proceso anterior entra en un bucle para completar las 290.000 solicitudes (pacientes), luego se extrae la información final de la base de datos local, construyendo de esta forma la base de datos de dirección y tiempo de viaje.

A continuación, se detalla el script de la función utilizada, la cual se encuentra alojada en Controller de la estructura de programación MVC (Model View Controller) `solicitud.php`:

```
1 | public function solicitud(){  
2 |
```

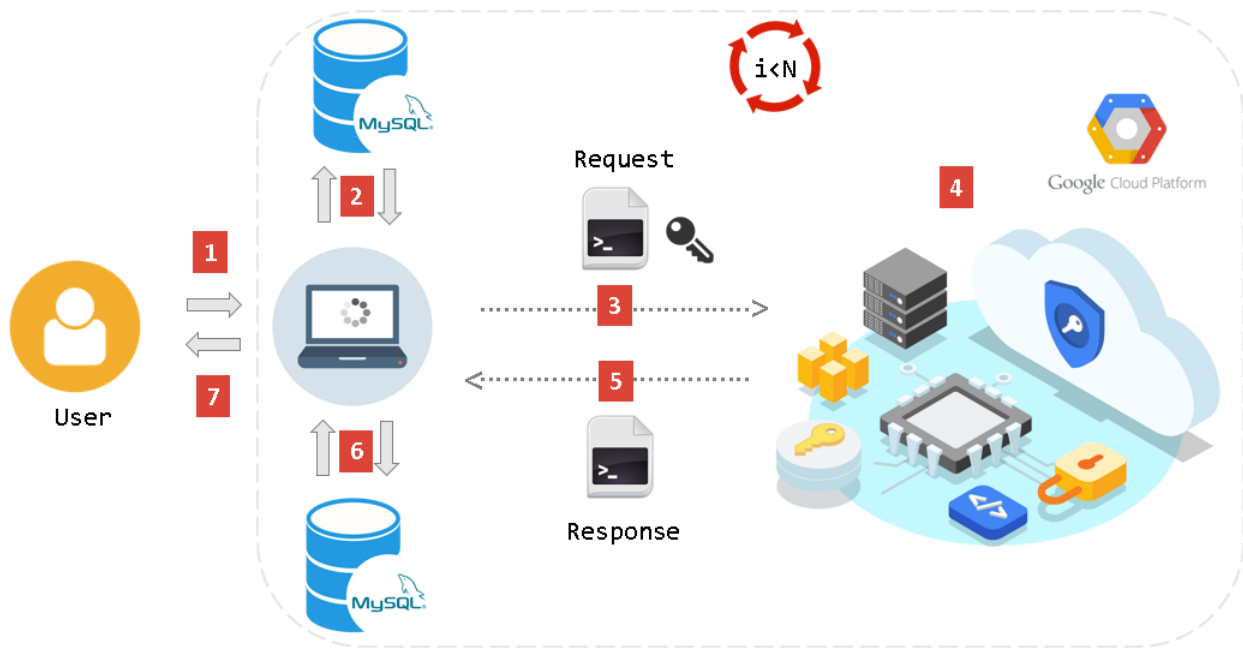


Figura 4.1: Proceso de requerimiento de información
(Fuente: Elaboración propia)

```

3  $limit=290000;
4  $r['timei']=new DateTime('h:i:s');
5  $r['error']='';
6  $apiKey;
7
8  for($id=1;$id<=$limit;$id++){
9      $con=$this->distancia_model->lectura_direccion($id);
10     //configurar direccion fija
11     if($con->centro==3){
12         $addressTo='Santos Dumont 999, Independencia, Region Metropolitana'; //direccion del hospital
13     } else {
14         if($con->centro==5){
15             $addressTo='Mar Tirreno 3349, Penalolen, Region Metropolitana'; //direccion quilin
16         } else {
17             false;
18         }
19     }
20     // se formatea la direccion en caso de error de ingreso
21     $formattedAddrTo= str_replace(' ', '+', $addressTo);
22     $sex0 = array('*','CASA', 'DEPTO', '.');
23     $sin0= trim(mb_strtolower(str_replace($sex0,'',$con->direccion),'UTF-8'));
24     $addressFrom = $sin0.",santiago, region Metropolitana google map";
25     // se formatea la informacion para el request
26     $formattedAddrFrom = str_replace(' ', '+', mb_strtolower($addressFrom,'UTF-8'));
27     $url ="https://maps.googleapis.com/maps/api/distancematrix/json?origins=".$formattedAddrFrom."&destinations=".$formattedAddrTo."&mode=transit&trafficModel=pessimistic&transit_mode=bus|train|tram|subway&key=".$apiKey;
28     // se utiliza la libreria CURL
29     $ch = curl_init();
30     curl_setopt($ch, CURLOPT_URL, $url);
31     curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);
32     curl_setopt($ch, CURLOPT_PROXYPORT, 3128);
33     curl_setopt($ch, CURLOPT_SSL_VERIFYHOST, 0);
34     curl_setopt($ch, CURLOPT_SSL_VERIFYPEER, 0);
35     $response = curl_exec($ch);
36     curl_close($ch);
37     $response_a = json_decode($response, true);
38     // en caso de error por informacion no encontrada

```

```

39     if (!empty($response_a->error_message)){
40         $this->distancia_model->actualizar_distancia2($id,0,0);
41         array_push($r['error'],'('.$id.')'. $response_a->error_message);
42     } else {
43         $dista = $response_a['rows'][0]['elements'][0]['distance']['text'];
44         $time = $response_a['rows'][0]['elements'][0]['duration']['value'];
45         //se actualiza la informacion en la BBDD local
46         $this->distancia_model->actualizar_distancia2($id,$dista,$time/60);
47     }
48 }
49 $r['timef']=new DateTime(date('h:i:s'));
50 // se visualiza el nivel de avance del proceso y el tiempo restante (opcional)
51 $this->load->view('vistas/distancia',$r);
52 }

```

4.2.2. Selección de registros y transformación de data

El siguiente paso corresponde al paso número 2 de la metodología. A partir de la información de distancias y tiempos, se cruzan las bases de datos especificadas en el capítulo anterior. Es importante mencionar que la base de datos de pacientes se cruza con la base de datos de citas médicas, asignando la distancia y tiempo de viaje a cada cita médica a partir del año 2013, esto se realiza al asumir que cada paciente ha permanecido en su dirección actual al menos 5 años.

Para el caso de variables asociada a las citas médicas y los pacientes, se utiliza el folio de la cita como enlace. Para el caso de asignación de data asociada al paciente con los datos de la cita médica, se utiliza el RUT junto con el nombre del paciente como punto de enlace, de esta forma el cruce de información es exitoso. Para datos exógenos, se utiliza el día como punto de enlace a la base de datos principal.

Una vez agregado los datos y notando que la cota inferior de información la posee el registro de pacientes con información de distancias transformada, se consolida como muestra principal la base de datos de pacientes con direcciones efectivas.

Finalmente, se ejecuta el paso número 3 de la metodología correspondiente a limpieza y pre procesamiento de información:

- La información original de HCUCH se encuentra ordenada por número de gestión, es decir, se encuentra explícita la información de cada gestión asociada a cada cita médica de cada paciente, es así como se eliminan los datos duplicados manteniendo sólo el número de gestiones generadas en cada cita médica, su estado final, tiempos promedios, etc.
- Una vez se obtienen los registros únicos por cita, se ordenan de forma temporal tomando como cluster el ID del paciente
- Se diferencian las ocasiones y frecuencias de citas médicas por paciente mediante variables auxiliares. Las gestiones realizadas por cita son tratadas de igual forma.
- Se crean variables asociadas al comportamiento del cluster (pacientes) a través del tiempo (trazabilidad), tales como porcentaje de asistencia relativa con un retardo (porcentaje obtenido hasta la cita anterior), cantidad de citas, porcentaje de fiabilidad con un retardo (paciente que afirma que asistiría a la cita y que finalmente asiste), retardos directos de asistencia denominados *Lags* (1 si asiste a la cita médica anterior y 0 en caso contrario), niveles de lealtad con retardo (asociados a pacientes que asisten con el mismo profesional) y otras variables

asociadas al comportamiento histórico. Es importante notar que hasta el momento la información constituida corresponde a citas médicas con fecha de atención entre los años 2007 y 2018.

- Una vez se obtiene la información de comportamiento histórico especificado en cada cita registrada de forma relativa, y dado que la cota inferior de data corresponde a la distancia y tiempo de viaje, donde se utiliza el supuesto de asumir que los pacientes han vivido en su última dirección declarada durante los últimos 5 años, se decide eliminar la información² previa a 2013.
- Las especialidades en la base de datos, se encuentran categorizadas por *servicios de especialidad*³, por lo que sería complejo entender y comparar el comportamiento para especialidades categorizadas de forma distinta, por ejemplo, geriatría o gastroenterología (categorizada como servicio de medicina) y sólo se encuentra especificada como información en la subespecialidad. De esta forma, se transforma la variable especialidad, explicitando la especialidad médica de la cita agendada, es así como la nueva categoría especialidad resulta de una combinación entre las categorías *servicios de medicina* y *servicios de especialidades*.
- Finalmente se obtiene un set de 1.586.000 datos limpios correspondientes a citas médicas del periodo 2013 a 2018, e incluye variables de comportamiento histórico desde el año 2007.

4.3. Análisis descriptivo

El siguiente análisis es realizado sobre la base de datos TICARES y tiene por objetivo encontrar tendencias de comportamiento a partir de un análisis descriptivo de las características de las citas médicas, para ello se analizan variaciones intertemporales, de interacciones y de segmento.

Comportamiento por tipo de cita médica

Se puede observar que la cantidad de citas médicas agendadas con cupos definidos anualmente ha variado menos de 6 %⁴ (ver gráfico 4.2). Para el caso de citas médicas agendadas por sobrecupo, se puede observar que se han reducido cerca de 4 puntos porcentuales en el periodo 2013 al 2018. Para el año 2018, sólo el 7 % de las citas médicas corresponden a citas agendadas por sobrecupo. En cuanto al nivel de asistencia para cada tipo de cita médica, se observa que el comportamiento de asistencia es contrario a su cantidad de agendamiento, se observa que el porcentaje de asistencia para los cupos agendados se encuentran entre 71,9 % y 77,5 % para los años 2013 y 2018 respectivamente, mientras que el nivel de asistencia de citas médicas agendadas como sobrecupo se encuentra sobre el 79,5 % con 81,1 % para el año 2013 y 90,9 % para el año 2018, con una mínima de 79,5 % para el año 2015. Se observa además una diferencia notoria en el nivel de asistencia para cada tipo de cita en el año 2018, donde la asistencia a citas médicas agendadas como sobrecupo tuvo un alza de 9 puntos porcentuales. Dados los resultados anteriores, es probable que el comportamiento de asistencia del paciente sea altamente influenciado por el tipo de cita.

²La acción de traer de forma relativa la información temporal por medios de retardos corresponde a una técnica para tratar datos de panel como datos pertenecientes a data de corte transversal, sin perder información en el proceso.

³Ver sección 1.1.2. *Hospital Clínico de la Universidad de Chile.*, apartado *Servicios*

⁴La capacidad de las instituciones no presenta variación en el periodo 2013 - 2018. Además los cupos se encuentran definidos a partir de la utilización máxima de la estructura del centro y/o la oferta médica presente en el periodo.

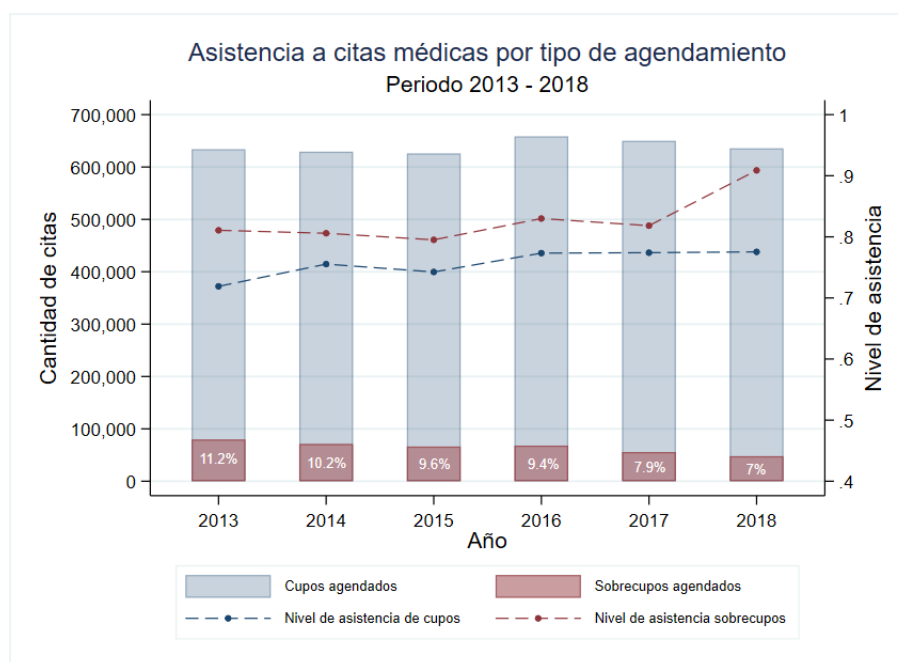


Figura 4.2: Evolución de frecuencia de solicitudes por edad periodo 2013 - 2018
(Fuente: Elaboración propia)

Comportamiento por sucursal

Se puede observar que clínica Quilín posee menor nivel de asistencia de pacientes a sus citas médicas en comparación a HCUCH con 72,4 % y 80,3 % respectivamente, para la institución en conjunto se observa un nivel de asistencia de 78,5 % incluyendo sobrecupos. Si se compara la evolución que ha sufrido la asistencia de pacientes, se puede observar un aumento progresivo para ambas sucursales (ver tabla 4.1).

Año	HCUCH	Clínica Quilín	Institución
2013	75,5 %	66,1 %	72,9 %
2014	78,8 %	68,2 %	76,1 %
2015	77,9 %	65,2 %	74,8 %
2016	81,3 %	66,7 %	77,9 %
2017	81,0 %	67,7 %	77,8 %
2018	80,3 %	72,4 %	78,5 %

Tabla 4.1: Porcentaje de asistencia por sucursal (incluye sobrecupos)
(Fuente: Elaboración propia)

Comportamiento por género

Se observa que en promedio el 63 % de las citas solicitadas corresponden a mujeres, siendo sólo un 37 % de las citas solicitadas por hombres. En la tabla 4.2 se puede observar que el comportamiento de asistencia entre géneros⁵ en la institución se mantiene con promedios de 71,4 % para

⁵Calculados sobre las citas solicitadas por el género respectivo.

mujeres y 74,8 % para hombres, no se evidencia una mejora significativa por género desde el año 2013 al año 2018. Se observa además que los hombres tienen una leve tendencia a asistir con mayor frecuencia a sus citas médicas.

Año	<i>HCUCH</i>		<i>Clínica Quilín</i>		<i>Institución</i>	
	<i>Femenino</i>	<i>Masculino</i>	<i>Femenino</i>	<i>Masculino</i>	<i>Femenino</i>	<i>Masculino</i>
2013	72,8 %	75,3 %	61,5 %	64,9 %	69,7 %	72,4 %
2014	75,5 %	77,9 %	63,3 %	64,9 %	72,2 %	74,4 %
2015	74,2 %	76,7 %	59,9 %	61,8 %	70,6 %	72,8 %
2016	76,5 %	78,9 %	59,9 %	61,4 %	72,5 %	74,6 %
2017	75,4 %	77,8 %	59,0 %	61,5 %	71,2 %	73,7 %
2018	74,9 %	77,5 %	63,2 %	66,3 %	72,0 %	74,8 %

Tabla 4.2: Porcentaje de asistencia por género del paciente
(Fuente: Elaboración propia)

Comportamiento por edad

Se puede observar que el segmento etario presenta variaciones entre los años 2013 y 2018 (ver gráfico 4.3), para edades menores a 54 años se observan reducciones en sus frecuencias de solicitudes de hasta 0,9 % a los 2 años de edad, lo que se traduce como 2690 citas agendadas de diferencia, para edades mayores existe un aumento en sus frecuencias, con máxima de 0,4 % a los 58 años de edad, lo que se traduce en 2498 solicitudes de diferencia. Es así como se reconoce un cambio en el segmento de la institución con una leve tendencia de atención a adultos mayores. En cuanto a las citas asistidas, para el segmento perteneciente a los 18 y 30 años, pese a su reducción de frecuencia de solicitudes, se observa un aumento en el nivel de asistencia del paciente.

Comportamiento por mes

El comportamiento de asistencia en la institución presenta variaciones mensuales (ver gráfico 4.4), se observa que las variaciones de asistencia por mes se acentúan en mayor medida el año 2013, a medida que transcurre el tiempo, las asistencias muestran menores variaciones, sin embargo se observa un comportamiento persistente a través de los años. Para el año 2018, se observan principalmente variaciones en los meses de Junio, Julio y Agosto.

Comportamiento por día del mes

Si se analiza la variación de asistencia diaria promedio para el año 2018 (ver gráfico 4.5) se puede observar que no existe una variación significativa por día, sin embargo, existen variaciones significativas por intervalos, para este caso se clasifican 3 subconjuntos; menores o iguales a 10, entre 10 y 20 y mayores de 20, de esta forma se podría postular en el modelo explicativo que existen variaciones en la asistencia a citas médicas para las clasificaciones anteriores.

Comportamiento por día de la semana

El comportamiento por día de la semana varía notoriamente (ver gráfico 4.6), se puede observar el promedio anual por día de la semana, donde existen variaciones diarias en la asistencia de pa-

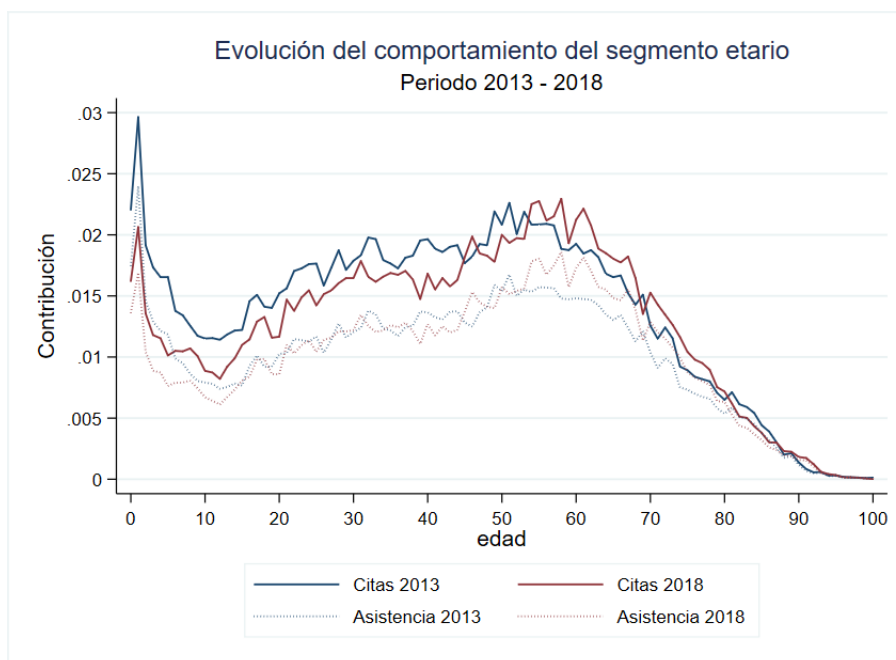


Figura 4.3: Evolución de frecuencia de solicitudes por edad periodo 2013 - 2018
(Fuente: Elaboración propia)

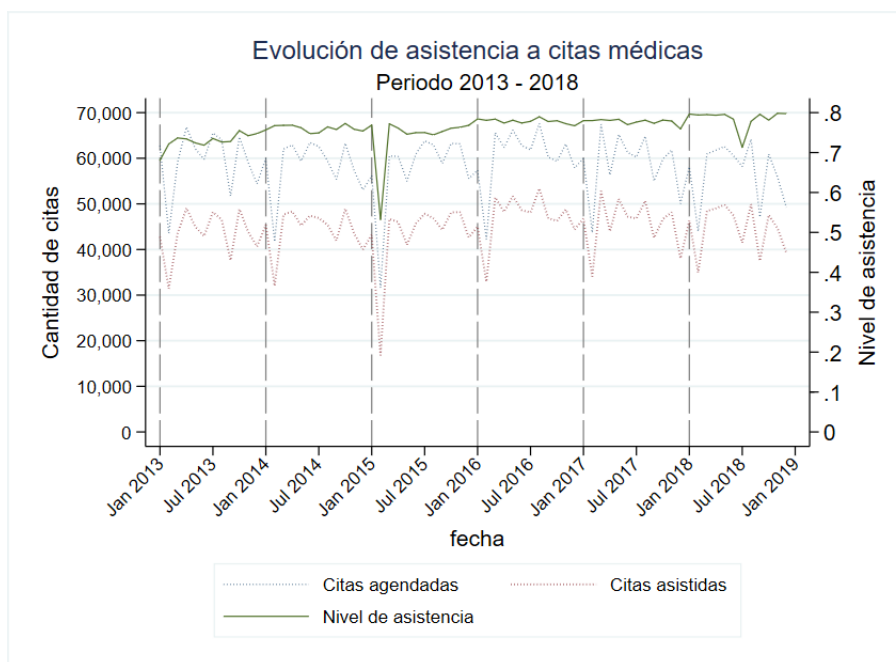


Figura 4.4: Evolución mensual de asistencia a citas médicas periodo 2013 - 2018
Fuente: Elaboración propia

cientes con el nivel más bajo para el día Sábado, con asistencias cercanas al 70 %. Se observa un comportamiento similar para las citas promedio en semanas de cada año. Bajo esta evidencia, es probable que el día de la semana al cual el paciente deba asistir a su cita médica sea una variable explicativa para el modelo.

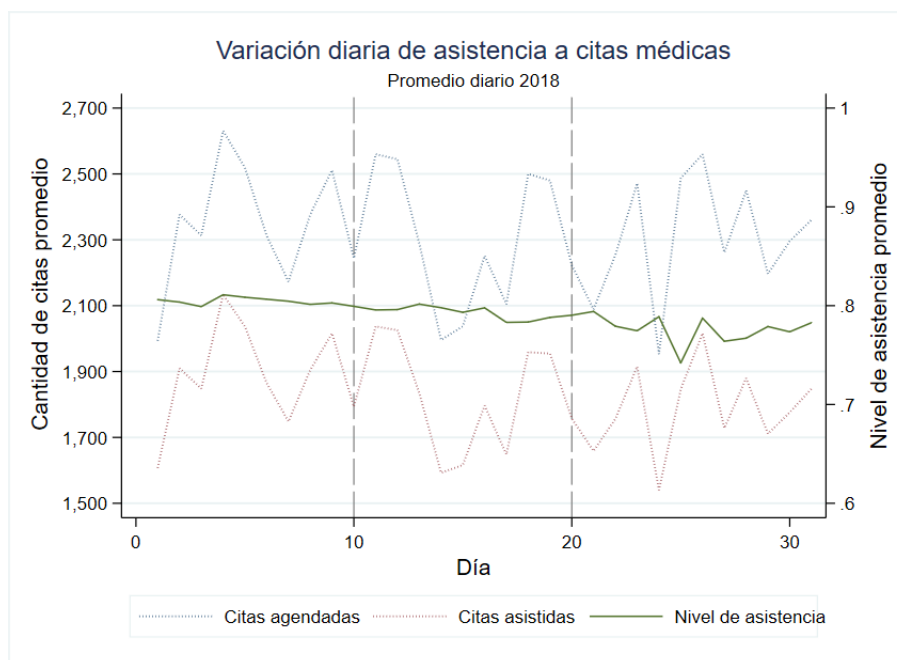


Figura 4.5: Variación diaria promedio de asistencia a citas médicas año 2018
(Fuente: Elaboración propia)

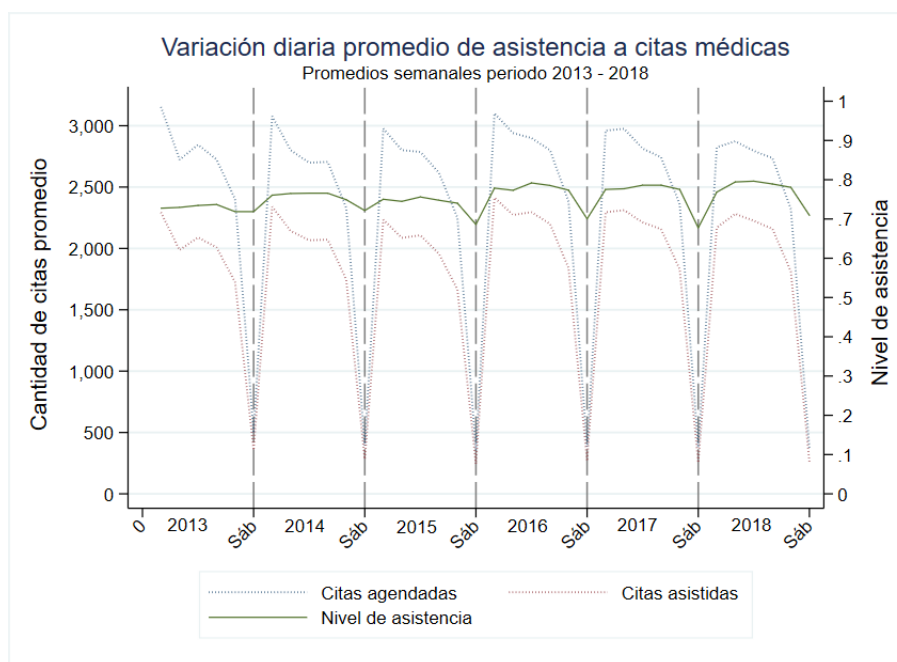


Figura 4.6: Variación diaria promedio anual (por día de la semana) de asistencia a citas médicas año 2018
(Fuente: Elaboración propia)

Comportamiento por canal

Al comparar las citas médicas asociadas al canal web (Ver gráfico 4.7), se observa que ha habido un aumento en el periodo 2013 a 2018, llegando a 17,7% para el 2018. En cuanto al nivel de asistencia, se observa que las citas agendadas por el canal web poseen menor nivel de asistencia que citas agendadas por otro medio. Para el año 2018, el agendamiento web presentó un nivel de

asistencia de 74 % mientras que para citas agendadas por otro canal poseen un nivel de asistencia de 79,4 %.

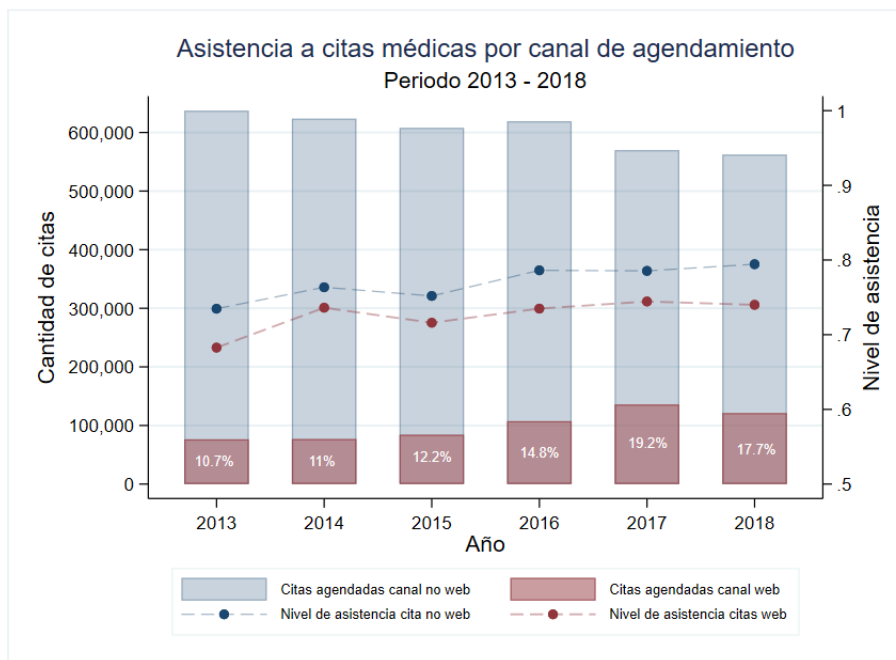


Figura 4.7: Asistencia a citas médicas por canal de agendamiento (incluye sobrecupos)
(Fuente: Elaboración propia)

Efecto de la latencia

Definición 4.1 Se define *latencia* como la diferencia temporal en días entre la fecha de atención de la cita médica y la fecha de su solicitud.

En el gráfico 4.8 se observa que en latencia 0, el nivel de asistencia promedio es 91,2 %, a partir de latencia igual a 1, el nivel de asistencia promedio disminuye notoriamente a 77 %, donde disminuye paulatinamente hasta latencia igual a 11 con un nivel de asistencia promedio igual a 70,2 %, a partir de este último punto, comienza un leve aumento en el nivel de asistencia promedio menor a 4 puntos porcentuales, comportamiento de continua hasta llegar a una latencia igual a 40 días. Para latencias mayores se observa variabilidad en el nivel de asistencia, la cual podría ser explicada en mayor medida por la cantidad de citas médicas asociadas a dicho nivel.

Interacción género y latencia

Si se analiza el comportamiento de los pacientes utilizando las dimensiones de género y latencia (ver gráfico 4.9), donde se observa que el nivel de asistencia en hombres es mayor al de mujeres, sin embargo, su comportamiento en cuanto a la latencia es similar. A partir de 40 días de latencia existe mayor variabilidad debido a una reducción de frecuencias para ambos sexos, de esta forma no se observan eventuales injerencias de la interacción latencia y género para el comportamiento del paciente.

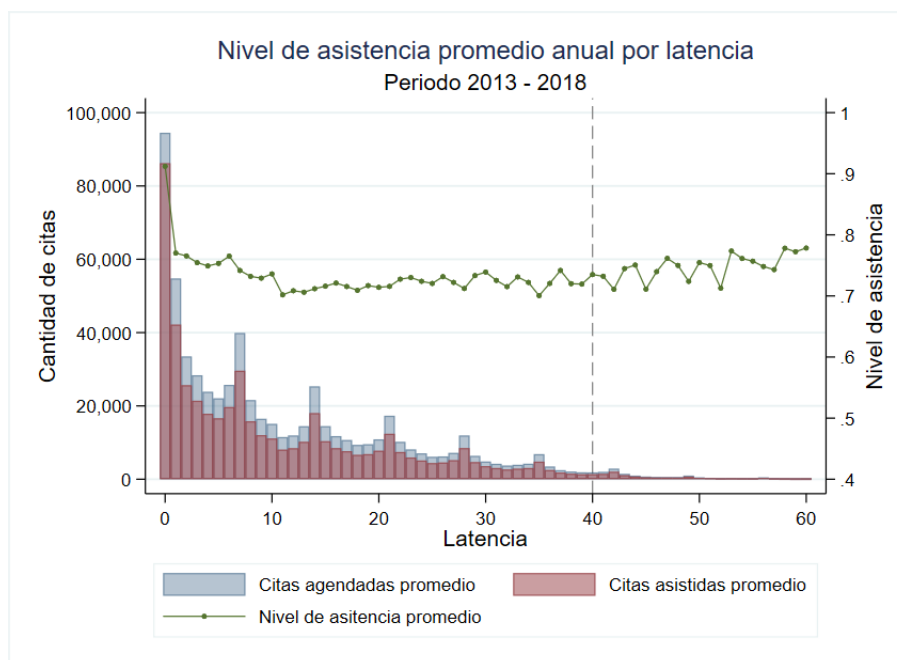


Figura 4.8: Nivel de asistencia promedio a citas médicas por latencia periodo 2013 - 2018
(Fuente: Elaboración propia)

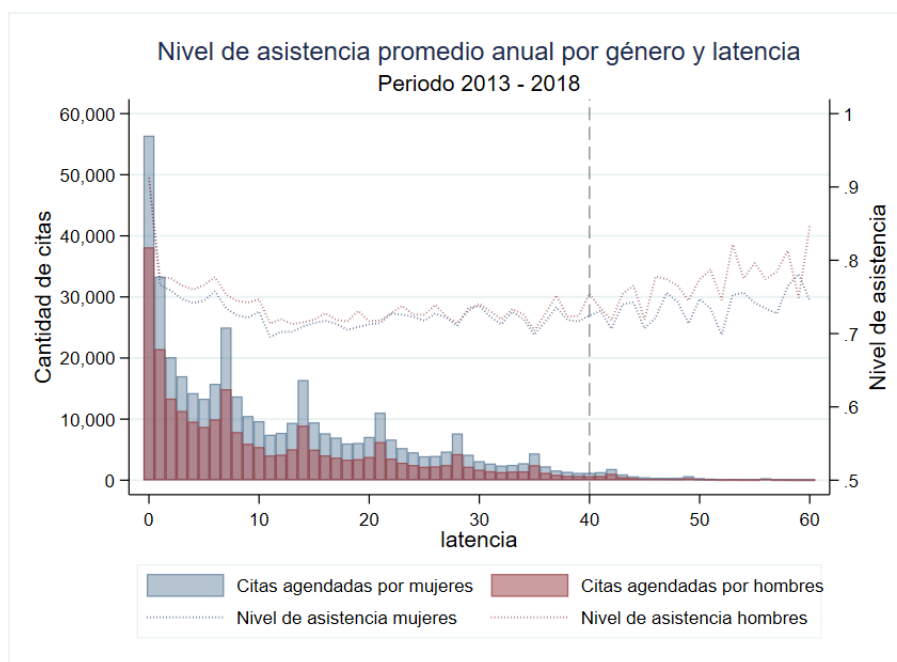


Figura 4.9: Nivel de asistencia promedio a citas médicas por género y latencia periodo 2013 - 2018
(Fuente: Elaboración propia)

Interacción canal web y latencia

Al analizar la variación entre la latencia y el canal de agendamiento web (ver gráfico 4.10), se puede observar que para latencia igual a cero, es decir, citas médicas que fueron solicitadas con asistencia para el mismo día de solicitud, existe una diferencia de 12 puntos porcentuales en el nivel de asistencia promedio con 92 % para citas médicas agendadas por un canal no web y 80 %

de asistencia promedio para citas médicas agendadas por el canal web respectivamente. Para citas médicas con latencia entre 1 día y 21 días, el nivel de asistencia promedio es similar para ambos canales, con leves diferencias de asistencia por periodo. Para latencias entre 21 y 40 días, el nivel de asistencia promedio decae notoriamente en citas agendadas vía web llegando a una diferencia de 18 puntos porcentuales, con 67 % de asistencia promedio para citas agendadas vía web y 74,9 % de asistencia promedio para citas agendadas por otro canal. A partir de 40 días de latencia, existe mayor variabilidad debido a la frecuencia de las citas para este periodo, por lo que no es posible concluir dada la dimensión de la muestra en este intervalo.

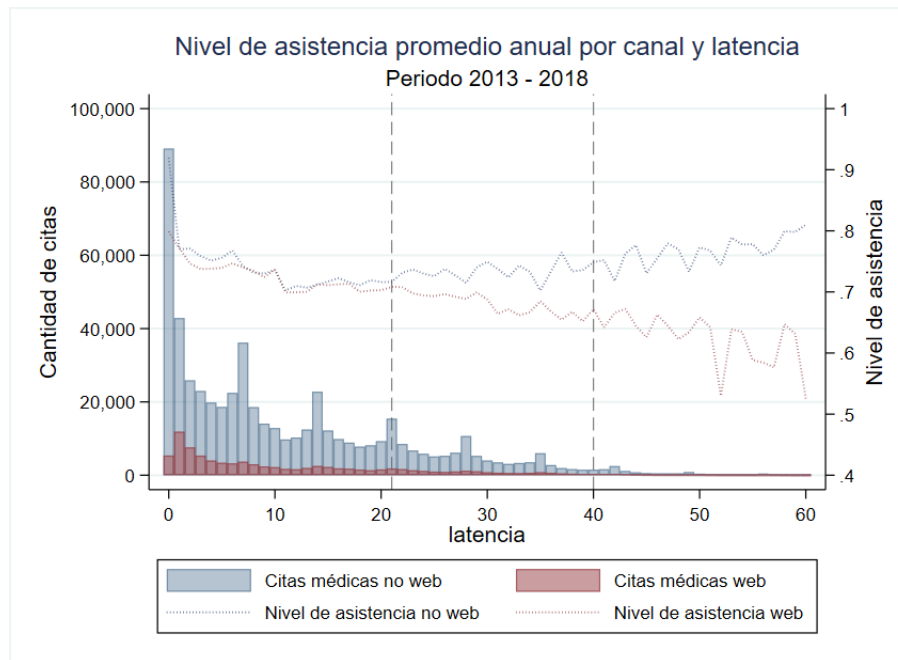


Figura 4.10: Nivel de asistencia promedio a citas médicas por canal y latencia periodo 2013 - 2018
(Fuente: Elaboración propia)

Interacción género y edad

Al analizar las variaciones entre edad y género (ver gráfico 4.11), se puede observar que la cantidad promedio de citas médicas agendadas por hombres es mayor a la de mujeres hasta los 13 años de edad, para edades mayores la cantidad de citas médicas agendadas por mujeres es mayor al de hombres con hasta un 73 % del total de citas médicas promedios agendadas para la edad de 31 años. En cuanto al nivel de asistencia, se puede observar que, entre los 0 y 48 años, no se observan diferencias significativas en el nivel de asistencia, para edades entre 48 y 90 años, se observa que el nivel de asistencia para el género masculino es hasta 3.5 puntos porcentuales mayor a asistencias para el género femenino. A partir de los 94 años, la cantidad de citas médicas promedio para ambos géneros se reduce de forma significativa, generando una notoria variabilidad en el nivel de asistencia.

Interacción edad y sobrecupo

De acuerdo al gráfico 4.12, el nivel de asistencia promedio para citas médicas agendadas a pacientes de 0 a 62 años de edad es 22 puntos porcentuales mayor para citas medidas agendadas como

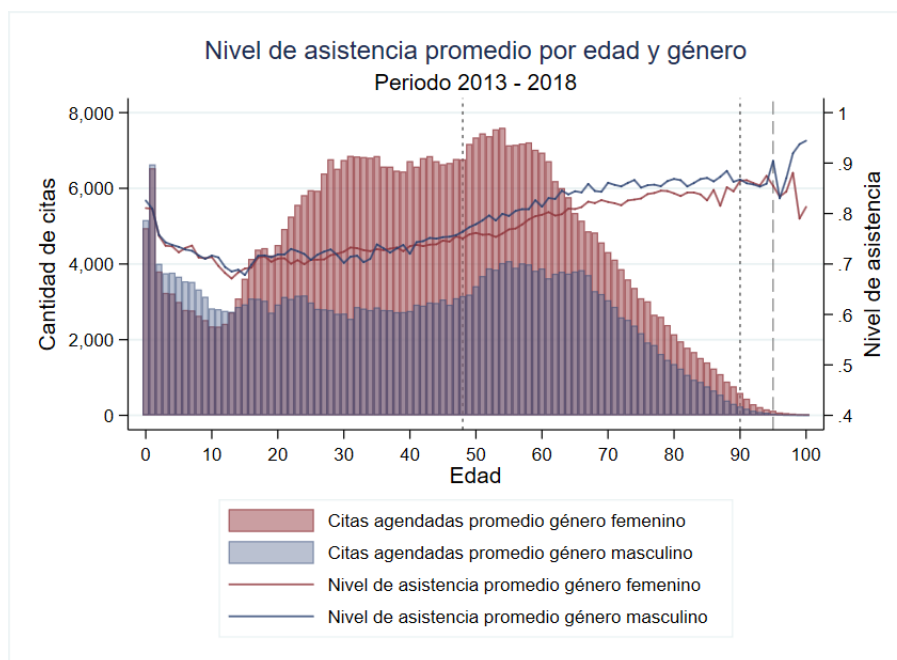


Figura 4.11: Nivel de asistencia promedio a citas médicas por edad y género periodo 2013 - 2018
(Fuente: Elaboración propia)

sobrecupo que citas médicas agendadas por cupo determinado con 95,6 % y 73,9 % respectivamente. A partir de los 62 a los 90 años, el comportamiento de asistencia para ambos tipos de citas es similar. Se observa que, a partir de los 90 años, la cantidad de citas se reduce significativamente, generando variabilidad en los niveles de asistencia respectivos.

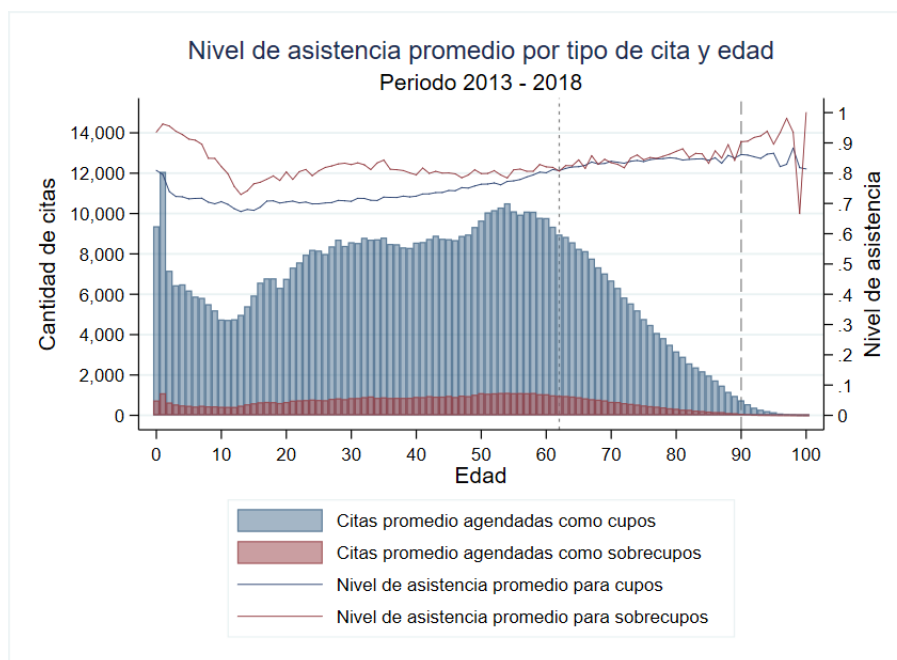


Figura 4.12: Nivel de asistencia promedio a citas médicas por tipo de cita y edad periodo 2013 - 2018
(Fuente: Elaboración propia)

4.4. Selección de variables

A partir de lo anterior, se continua con el paso número 4 de la metodología, correspondiente a reducción de data, selección de variables, eliminación de columnas dependientes como fechas de nacimiento y edad, proceso que tiene como fin optimizar el almacenamiento de información.

A continuación, se enuncian las variables a utilizar en el modelo estructural:

<i>Variable</i>	<i>Tipo</i>	<i>Descripción</i>
Especi	Categoría	Variable dependiente correspondiente a la especialidad de la cita agendada, éstas se encuentran categorizadas en 40 especialidades. Para resolver el de forma parcial el problema de información multinivel ⁶ , por ejemplo, citas médicas agendadas para una misma especialidad en ambos centros denotada con distintos nombres o etiquetas, se estandarizan las categorías por áreas de estudio generales asociadas a la especialidad ⁷ . Se utilizará <i>cod.17 medicina general (m.g.)</i> como especialidad pivote (para comparar probabilidades de asistencias. El conjunto de especialidades, para efectos de expresiones en ecuación, será <i>ESP</i> .
Centro	Dicotómica	Variable dependiente correspondiente al centro o sucursal de la institución, HCUCH o Clínica Quilín, donde esta última corresponde a la categoría base (0).
Scupo	Dicotómica	Variable dependiente correspondiente al tipo de cita médica: 1 si la cita fue agendada como sobrecupo, 0 en otro caso.
Web	Dicotómica	Variable dependiente correspondiente al canal de agendamento: 1 si es web, 0 en otro caso.
Conf	Dicotómica	Variable dependiente correspondiente al estado de confirmación de la cita médica: 1 para citas confirmadas, 0 si no fue confirmada.
Gescaec	Dicotómica	Variable dependiente correspondiente al tipo de cobertura: 1 si pertenece al sistema GES ⁸ /CAEC ⁹ , 0 si no.
Post	Dicotómica	Variable dependiente correspondiente a citas médicas clasificadas como post operatorias, de esta forma 1 corresponde a citas post operatorias y 0 en otro caso.
Laten	Numérica	Variable dependiente correspondiente a la latencia ¹⁰ de la cita médica.
Diasem	Categoría	Variable dependiente correspondiente al día de la semana de la cita médica. El conjunto para efectos de la ecuación corresponde a <i>DS</i> , con categoría base igual a 1 (día lunes)

(Continúa en la página siguiente)

⁶Se define como información multinivel a información subcategorizada o anidada.

⁷Las áreas de estudio utilizadas corresponden a especialidades generales que posee la institución [9].

⁸Las Garantías Explícitas en Salud (GES), constituyen un conjunto de beneficios garantizados por ley para las personas afiliadas a Fonasa y a las Isapres [11].

⁹La Cobertura Adicional para Enfermedades Catastróficas (CAEC) es un beneficio adicional al plan de salud otorgado por algunas Isapres que permite financiar, cumplidos ciertos requisitos, hasta el 100 % de los gastos derivados de atenciones de alto costo, tanto hospitalarias como algunas ambulatorias, realizadas en el país y cubiertas por el plan de salud [10].

¹⁰Ver definición en sección 3.3 *Análisis descriptivo* apartado *Efecto de la latencia*, definición 3.1.

(Continúa de la página anterior)

<i>Variable</i>	<i>Tipo</i>	<i>Descripción</i>
Mes	Categórica	Mes del año de la fecha de atención de la cita médica. El conjunto para efectos de la ecuación corresponde a M , con categoría base igual a 1 (Enero).
Diag	Categórica	Los días del mes (1 al 31) correspondiente a la fecha de atención de la cita médica, fueron agrupados en 3 estratos: 1 para los días 1 al 10, 2 para los días 11 al 20 y 3 para los días 21 al 31. El conjunto para efectos de la ecuación corresponde a DG , con categoría base igual a 1.
Horag	Categórica	Las horas del día fueron segmentadas en cuatro grupos, el principal criterio para su segmentación se basó en la frecuencia de atención para dicho horario, con el fin de dejar estratos con similar nivel de frecuencia. De esta forma se clasifica como 1 a horarios de atención entre 08:00 hrs. y 9:00 hrs., 2 a horarios entre 10:00 hrs. y 12:00 hrs., 3 a horarios entre 13:00 hrs. y las 17:00 hrs. y 4 a horarios entre 18:00 hrs. y 20:00 hrs. El conjunto para efectos de la ecuación corresponde a HG , con categoría base igual a 1.
Gender	Dicotómica	Variable dependiente correspondiente al género del paciente asociado a la cita médica: 1 si es femenino, 0 si es masculino.
Edad	Numérica	Variable dependiente correspondiente a la edad en años del paciente asociado a la cita médica.
Edad ²	Numérica	Variable dependiente correspondiente al cuadrado de la edad del paciente.
Ffijo	Dicotómica	Variable dependiente correspondiente a citas médicas cuyos pacientes poseen un número de teléfono fijo registrado para la gestión de citas: 1 para quienes lo han registrado y 0 para quienes no.
Fmovil	Dicotómica	Variable dependiente correspondiente a citas médicas cuyos pacientes poseen un número de teléfono móvil registrado para la gestión de citas: 1 para quienes lo han registrado y 0 para quienes no.
Ftrab	Dicotómica	Variable dependiente correspondiente a citas médicas cuyos pacientes poseen un número de teléfono del trabajo registrado para la gestión de citas: 1 para quienes lo han registrado y 0 para quienes no.
Mail	Dicotómica	Variable dependiente correspondiente a citas médicas cuyos pacientes poseen email registrado para la gestión de citas: 1 para quienes lo han registrado y 0 para quienes no.
Dist	Numérica	Variable dependiente correspondiente a la distancia real ¹¹ en kilómetros entre la dirección de la vivienda del paciente y el centro de atención respectivo.

(Continúa en la página siguiente)

¹¹Distancia que incluye vías óptimas considerando caminos permitidos en sitios urbanos.

(Continúa de la página anterior)

<i>Variable</i>	<i>Tipo</i>	<i>Descripción</i>
Time	Numérica	Variable dependiente correspondiente al tiempo de viaje por transporte público, medido en minutos, recurriendo al evento pesimista ¹² , entre la dirección de la vivienda del paciente y el centro médico respectivo.
Dist ²	Numérica	Variable dependiente correspondiente al cuadrado de la variable <i>Distancia</i> .
Time ²	Numérica	Variable dependiente correspondiente al cuadrado de la variable <i>Tiempo</i> .
Antig	Numérica	Variable dependiente correspondiente a la diferencia en días entre la primera solicitud de cita médica del paciente y la solicitud actual (relativa).
TC	Numérica	Variable dependiente correspondiente a la diferencia en días entre la fecha de atención pasada y la fecha de solicitud actual. Tiene por objetivo capturar el comportamiento del paciente dependiendo del <i>tiempo</i> entre eventos
Preferiado	Dicotómica	Variable dependiente que toma el valor de 1 si la fecha de atención corresponde a un día previo a un feriado, 0 en otro caso.
Feriado	Dicotómica	Variable dependiente que toma el valor de 1 si la fecha de atención corresponde a un día feriado, 0 en otro caso.
Postferiado	Dicotómica	Variable dependiente que toma el valor de 1 si la fecha de atención corresponde a un día siguiente a un feriado, 0 en otro caso.
Manifest	Dicotómica	Variable dependiente que toma el valor de 1 si para el día de la atención de la cita médica ocurre alguna manifestación social en Santiago, 0 en otro caso.
Deporte	Dicotómica	Variable dependiente que toma el valor de 1 si para el día de atención de la cita médica ocurre algún evento deportivo con interés a nivel país, por ejemplo, partidos de la selección nacional Chilena de fútbol, 0 en otro caso.
IPC	Numérica	Variable dependiente que señala la variación mensual del IPC (Índice del Precio del Consumidor).
Temp	Numérica	Variable dependiente que señala la temperatura en grados Celsius promedio del día de la cita.
Precipg	Categoría	Para las precipitaciones diarias, se categoriza como 0 a precipitaciones promedios diarias igual a 0mm, 1 a precipitaciones promedios diarias menores o iguales a 2mm diarias, 2 a precipitaciones promedios diarias mayores a 2mm y menores a 15mm, 3 a precipitaciones promedios diarias mayores a 15mm y menores a 30mm y finalmente como 4 a precipitaciones promedios diarias mayores a 30mm y menores a 60mm. El conjunto para efectos de la ecuación corresponde a <i>PG</i> , con categoría base igual a 0.

(Continúa en la página siguiente)

¹²Corresponde al tiempo de viaje promedio de eventos con mayor congestión registrada para la hora de la cita (ver sección 3.2.1 Selección de pacientes).

(Continúa de la página anterior)

<i>Variable</i>	<i>Tipo</i>	<i>Descripción</i>
Lag ⁽ⁱ⁾	Dicotómica	Variable dependiente que toma el valor de 1 si para el retardo (i) ¹³ , el paciente asiste a su cita médica, 0 en otro caso.
Lagprof ⁽ⁱ⁾	Dicotómica	Variable dependiente que toma el valor de 1 si para el retardo (i) el paciente fue atendido por el mismo profesional que lo atiende en la cita actual (relativa).
Lagesp ⁽ⁱ⁾	Dicotómica	Variable dependiente que toma el valor de 1 si para el retardo (i) el paciente solicita una cita médica para la misma especialidad que la cita actual (relativa).
Lagweb ⁽ⁱ⁾	Dicotómica	Variable dependiente que toma el valor de 1 si para el retardo (i) el paciente solicita una cita médica vía web.
Laglaten	Numérica	Variable dependiente correspondiente a la latencia de la cita médica con retardo 1.
Lagpor	Numérica	Variable dependiente correspondiente al promedio relativo de asistencias del paciente.
Lagfiab	Numérica	Variable dependiente correspondiente al promedio relativo de asistencias del paciente, donde el paciente ha confirmado su asistencia (fiabilidad en su declaración)
Lagscupo	Numérica	Variable dependiente que toma el valor de 1 si la cita anterior a la cita relativa (actual) solicitada por el paciente fue por medio de un sobrecupo.

Tabla 4.3: Definición de variables utilizadas en el modelo predictivo
(Fuente: Elaboración propia)

¹³Por ejemplo, si se quisiera analizar efectos de la cita anterior sobre la actual para el mismo paciente, se utilizaría el retardo 1 (lag1), donde esta variable tomaría el valor de 1 si el paciente asiste a dicha cita médica y 0 si no.

Capítulo 5

Modelamiento

Tal como se introduce en la sección *1.4 justificación del proyecto*, el modelamiento del comportamiento del paciente posee dos motivaciones claras:

La primera motivación corresponde a la interpretación de información para la toma de decisiones a partir de los modelos creados, tales como conocer la probabilidad de asistencia a una cita confirmada y notar diferencias significativas con respecto a citas que no se realiza dicha gestión, eventualmente podría significar oportunidades de mejora para el proceso, extrapolando el beneficio a distintos departamentos, donde se puede citar a marketing con propuestas publicitarias dada la eventual variación en la probabilidad de asistencia de pacientes en edades o géneros determinados.

En cuanto a la estructuración del modelo, se puede dar cuenta de la importancia de cinco líneas de información: La primera línea se encuentra asociada a la información de la cita en si misma; horarios de atención, niveles de latencia, entre otras características. En segundo lugar, se encuentra la información exógena asociada, la cual dependerá del tiempo¹ de acción, por ejemplo, marchas, temperatura, niveles de precipitación, entre otras. Luego, como tercera línea de información se encuentran las características del paciente; edad, género, entre otra información personal. A continuación, se encuentra la información asociada al comportamiento histórico del paciente, tales como, asistencia a su última cita médica, confiabilidad de sus declaraciones (para pacientes confirmados), niveles de asistencia, entre otra información asociado a la trazabilidad del comportamiento. Finalmente, se encuentra la última línea de información, la cual está asociada a componentes estocásticas de heterogeneidad no observable generada por clusters determinados, por ejemplo, si se asume a cada paciente como un cluster, es importante tener en cuenta heterogeneidad no observable entre pacientes y entre las decisiones de cada paciente (entre clases e intra clases).

Para comprender de forma más clara la asociación de líneas de información adoptadas en el presente escrito, se desarrolla el ejemplo de la gestión de confirmación expuesto en el párrafo ante-

¹Claramente dichas variables pueden ser modeladas con asociación a una cita i dado que cada cita se encuentra sujeta a un momento temporal único, sin embargo, se asocia a una variable tiempo para evitar confusiones y asociar el modelo a oportunidades de elección en un momento determinado. Un beneficio de dicha estructuración es que el modelo es extrapolable a momentos de elección múltiples como almacenamiento de estados (cita cancelada, en espera, asistida, eliminada, bloqueada), lo cual aportaría mayor nivel de información en acerca del comportamiento del paciente.

rior: Para tal gestión se pueden observar dos niveles de entendimiento, el primero y más superficial involucra supuestos de comportamiento ideal, es decir, pretender que la gestión de confirmación como tal logra capturar un porcentaje de utilidad del modelo (aborda de forma parcial las líneas de información definidas), y por ende, responde a la pregunta: "¿Qué tan importante es confirmar a los pacientes?", sin embargo, dicha asunción puede estar sesgada por una conclusión general y apresurada. Es aquí donde se involucra parte fundamental del análisis, el cual aborda el modelo bajo los cinco niveles de información, donde la respuesta del paciente se encuentra capturada a partir de su comportamiento histórico parcial o total, e incluso puede poseer efectos de su heterogeneidad no observable, es decir, que mediante la trazabilidad de sus decisiones, la gestión de confirmación se encuentra sujeta a la fiabilidad de las declaraciones del paciente (si el paciente asegura que asistirá a la cita médica al momento de la gestión de confirmación), e incluso podrían existir eventos aleatorios que afecten el comportamiento de pacientes específicos, como, por ejemplo, el comportamiento de un paciente que presenta un alto nivel de fiabilidad en sus declaraciones, confirma su cita médica, pero por razones desconocidas e inexplicables por la información disponible, no asiste a su cita médica.

La segunda motivación corresponde a la utilización de un modelo estructural para la predicción del comportamiento del paciente, con el fin de integrarlo de forma directa a la toma de decisiones (por ejemplo, saber la probabilidad promedio de asistencia de un número determinado de pacientes con el propósito de crear políticas de sobrecupo para dicha agenda, de esta forma, reducir las inasistencias a citas médicas).

5.1. Validación de la muestra

Información histórica de pacientes

Para validar el análisis del comportamiento del paciente, es necesario evaluar la cantidad de información disponible por paciente, de esta forma asegurar el aporte de información por cluster y evitar correlaciones entre rezagos² de primer orden (comportamiento de asistencia de la cita previa) y porcentaje de asistencia del paciente y/o rezagos de nivel superior, de igual forma, podría haber correlación entre porcentaje de asistencia y fiabilidad de la declaración del paciente, lo que sería perjudicial para el desarrollo del modelo. Un ejemplo de ello sería un conjunto de información donde la mayoría de los pacientes tuviera menos de tres citas solicitadas, para este caso, existiría una fuerte correlación entre los rezagos y el nivel de asistencia del paciente. Por este motivo es fundamental trabajar con fuentes de alto volumen de información heterogénea en el comportamiento de los clusters (niveles de recurrencia de atención, antigüedad del paciente, entre otras características).

En el gráfico de la figura 5.1, se puede observar la distribución de frecuencias promedio de pacientes categorizados por cantidad de citas médicas solicitadas, donde se observa un comportamiento descendente en la frecuencia promedio. Se puede notar que el conjunto de datos contiene incluso información de pacientes con más de un centenar de citas médicas asociadas, también existen cerca de 14.000 personas que poseen registros precarios, lo cual se puede asociar a pacientes nuevos (han solicitado sólo una cita médica) que no han solicitado más citas médicas. Cerca de 4.000 pacientes en promedio presentan 10 registros de citas médicas solicitadas. De igual forma

²Para efectos del presente estudio, se utilizará indistintamente el termino *rezago o lag* para referirse al comportamiento de asistencia de la cita anterior para el mismo paciente. Es importante notar que se utiliza sólo la terminología de serie de tiempo, mas no la metodología asociada.

se puede observar la antigüedad promedio para pacientes que solicitan una cantidad de citas establecida, de esta forma, es natural que a medida que aumente la cantidad de citas, aumente la antigüedad en años del paciente, es así como se registran antigüedades promedio mayores a 5 años para pacientes que han solicitado más de 60 citas médicas (en el panel de datos disponible)³

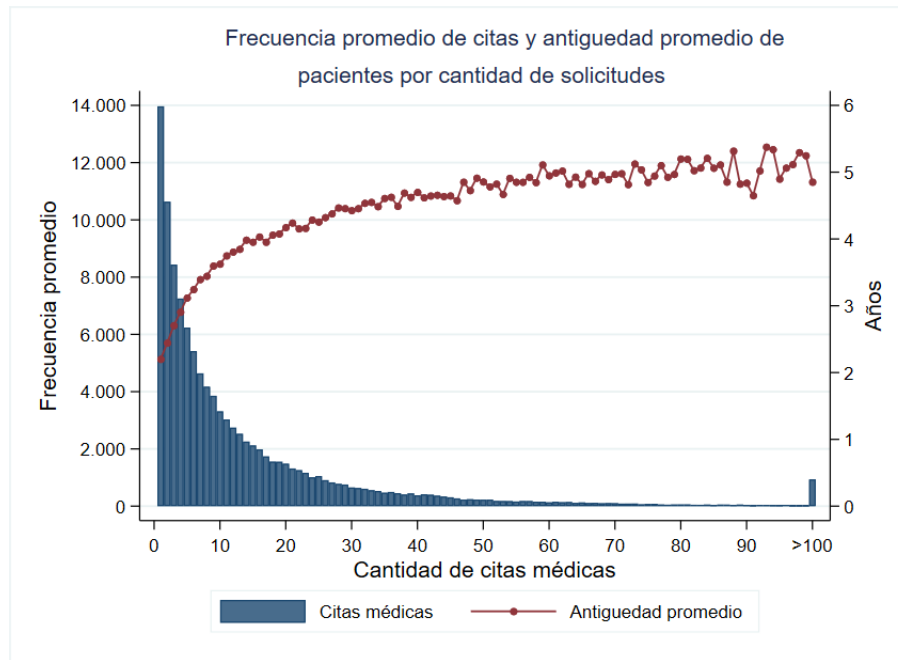


Figura 5.1: Frecuencia promedio de citas médicas y antigüedad promedio de pacientes por cantidad de solicitudes del panel
(Fuente: Elaboración propia)

Focos de investigación

Existen dos motivaciones para la creación e interpretación del modelo predictivo; predecir el comportamiento del paciente y explicar de forma numérica el efecto de los factores involucrados sobre su respuesta. A partir de lo anterior, es fundamental definir dos focos de investigación de acuerdo a sus objetivos: *Foco explicativo* y *foco predictivo*. El primero correspondiente a un modelo teórico, donde se busca que las estimaciones de los coeficientes de la ecuación sean precisas, ya que a partir de ellas se realizarán deducciones acerca de las covariables y sus efectos en la respuesta del paciente. El segundo foco tiene por objetivo obtener predicciones más fiables y acertadas, donde la interpretación de las variables no debe ser necesariamente intuitiva mientras cumpla su propósito (anticipar el comportamiento del paciente).

Selección de la muestra

Para obtener una interpretación precisa de los factores que afectan el comportamiento del paciente, es necesario entender las diferencias de comportamiento entre pacientes y en el paciente (comportamiento). Si bien se puede obtener información de su comportamiento histórico, a menudo existen componentes no observables de la muestra que afectan su comportamiento (cambio de

³En la sección 3.2.1 *selección de registros y transformación de data*, se muestra como fue recolectada la información de rezagos y comportamiento de asistencia de los pacientes.

gustos, eventos fortuitos, etc.), variables difícilmente cuantificables de forma directa, es así como es fundamental incluir dicha heterogeneidad no observable mediante variables estocásticas. Con el fin de incluir heterogeneidad no observable al modelo, es necesario agregar variables estocásticas asociada a una distribución específica, para el caso de estudio se utilizará la distribución normal, la cual se adaptará de acuerdo a la correlación entre clases e intra clases (heterogeneidad entre pacientes y entre el comportamiento de asistencia de cada uno). Dado lo anterior, a medida que se aumente el número de clusters, mayor será el costo⁴ asociado al ajuste del modelo.

A partir de la fórmula de Freeman[22], el número mínimo de registros concernientes a las variables a testear corresponden a aproximadamente 800 (incluyendo interacciones entre variables para otorgar mayor seguridad). Dada las limitaciones del procesamiento del panel de datos (al menos para comparar modelos con efecto aleatorio), se decide extraer una muestra aleatoria sujeta a la máxima capacidad de procesamiento de la tecnología disponible para realizar el presente estudio. El factor que limita el procesamiento corresponde a la cantidad de memoria RAM utilizada (16 Gb), la cual procesa como máximo 300.000 registros para el cálculo de heterogeneidad aleatoria, donde el número de clusters asciende a 77.141 (cantidad de pacientes de la muestra). Finalmente, se acepta la muestra dado que la cantidad de registro es muy superior a su cota inferior establecida.

Correlación entre observaciones

A veces las observaciones sobre la variable de resultado son independientes entre pacientes (clusters), pero no son necesariamente independientes dentro de ellos (dependencia de comportamiento). Para evitar problemas de correlación entre observaciones, se decide explicitar el *ID* del paciente como un *CLUSTER* y de esta forma obtener estimaciones robustas sobre los errores estándar (se observa una variación en los errores estándar y la matriz de varianza-covarianza de los estimadores, pero no sobre los coeficientes estimados). Para el caso del modelo con variable aleatoria, la correlación se resuelve al utilizar el ID del paciente como variable de efectos aleatorios.

Multicolinealidad y ajuste funcional

La importancia del volumen de data se muestra principalmente en el análisis de correlaciones de variables de comportamiento, evidentemente para data precaria, la correlación entre rezagos junto al nivel de asistencia de los pacientes forman un conjunto de variables electas a posibles problemas de multicolinealidad (dado que tienden a correlacionarse), a medida que la data se incrementa, el comportamiento de los pacientes es más heterogéneo por lo que va disminuyendo la correlación entre dichas variables, dentro de esta categoría queda el presente caso de estudio, donde el nivel de información es tan grande que otorga correlaciones entre rezagos menores a 0.4, lo cual es una correlación moderada y abordable con errores estándar robustos asociados. Otras variables a tener en cuenta son las transformaciones funcionales de covariables seleccionadas, tal es el caso de la edad, distancia y tiempo de viaje (cuadrado de las covariables como predictores adicionales⁵). El motivo de aceptar multicolinealidad tiene que ver con la obtención de un mejor ajuste funcional, de esta forma obtener con mayor precisión (de ser significativa) el efecto de la variable con el estado de la cita médica, eventualmente esta relación podría ser no lineal.

⁴El costo se encuentra relacionado a la capacidad del procesador, tiempo de procesamiento y finalmente memoria RAM para mantener matrices de varianza - covarianza activas.

⁵Ver sección 3.4 *Selección de variables*.

5.2. Estimación y evaluación del modelo empírico: Foco explicativo

La estructuración del modelo corresponde a la fase número 5 de la metodología, donde se estima el modelo empírico, se evalúa su ajuste y dependiendo del resultado se replantea el modelo. Para ello, se utilizará la adición de información asociada al comportamiento del paciente, de esta forma determinar diferencias en el impacto que produce sobre la asistencia la información del comportamiento histórico (nivel de asistencia, nivel de fiabilidad, lags o rezagos, etc.) y la información asociada a la cita en un momento determinado (horario, gestiones, especialidad, temperatura, etc.). Para analizar la información mediante efectos fijos, se utilizarán las funciones de enlace LOGIT y PROBIT, de igual forma para analizar efectos fijos y aleatorios (heterogeneidad no observable) se utilizará la función de enlace MIXED LOGIT. La estimación de coeficientes de los modelos con función de enlace LOGIT y PROBIT se realiza mediante el método de *Máxima verosimilitud (ML)*, mientras el modelo de efectos aleatorios que utiliza la función de enlace MIXED LOGIT estima los coeficientes mediante el *Método de Máxima verosimilitud Simulado (MSLE)*.

El proceso de análisis general para el foco explicativo será el siguiente:

1. Evaluar el ajuste de la función de utilidad *sin historial de comportamiento* mediante la función de enlace LOGIT.
2. Evaluar el ajuste de la función de utilidad *con información histórica asociada a 1 rezago* (comportamiento de la cita anterior) mediante la función de enlace LOGIT
3. Evaluar el ajuste de la función de utilidad *con información histórica asociada a 1 y 2 rezagos junto al nivel de asistencia y fiabilidad* mediante la función de enlace LOGIT
4. Evaluar el ajuste de la función *con información histórica completa* mediante la función de enlace PROBIT
5. Evaluar el ajuste de la función *con información histórica completa* agregando *heterogeneidad continua* sobre el paciente mediante la función de enlace MIXED LOGIT

La evaluación de los modelos tendrá tres aristas fundamentales⁶:

1. **Evaluar nivel de significancia:** Se evaluará el nivel de significancia global del modelo. Para ello se utilizará el test de Wald, el cual establece que si el p-valor es menor a 0.05 (nivel de significancia establecido), se rechaza la hipótesis nula de regresores nulos y por ende el modelo posee significancia conjunta.
2. **Evaluar bondad del ajuste:** Para evaluar el ajuste del modelo, se utilizarán dos familias de indicadores complementarios:
 - (a) *Pseudo-R²*: El *Pseudo-R²* tiene como objetivo establecer una métrica similar a *R²* (coeficiente de determinación para modelos ajustados mediante OLS⁷) en funciones li-

⁶El presente resumen de herramientas a testear, sus objetivos y estándares de decisión tiene como propósito clarificar al lector lo que se está evaluando. Para información técnica ver *Capítulo 2, sección 2.2.: Herramientas de ajuste.*

⁷Ajuste mediante Mínimos Cuadrados Ordinarios.

neales) a ajuste de funciones no lineales, en particular, ajustes realizados por el método de máxima verosimilitud (ML). Es importante destacar que es un parámetro aproximado de la medida de ajuste y por ende no se debe interpretar su resultado por sí solo. Para el presente estudio se utilizará el Pseudo R^2 de McFadden, cuyo estimador puede tomar valores de 0 a 1, donde la norma establece que a mayor estimador mayor es el ajuste del modelo. Es importante destacar que McFadden mediante la expresión *Rule of Thumb* señala que si se cumple: $0,20 \leq McFaddenPseudo - R^2 \leq 0,40$, el modelo presenta un excelente ajuste [19]. De igual forma que en el R^2 ajustado en OLS, el Pseudo- R^2_{adjust} corresponde a un ajuste que penaliza la utilización de demasiados predictores, si los predictores son significativos, eventualmente la penalización será pequeña y el Pseudo- R^2_{adjust} tenderá a Pseudo- R^2 .

- (b) *Criterios de información*: Los criterios de información establece un nivel de calidad relativa del modelo, es por ello que tiene por objetivo comparar modelos ajustados. La norma general es escoger el modelo con menor criterio de información. Los dos criterios utilizados en el presente estudio serán: Criterio de Información de Akaike (AIC) y Criterio de Información Bayesiano (BIC).

3. Evaluar poder discriminatorio⁸

- (a) *Dispersión de residuos* Es una medida que muestra la eficacia del modelo, es decir, que tan bien estima a cada observación. En el presente estudio se expondrá de forma gráfica la dispersión de los residuos estandarizados o de Pearson y residuos de desviación, es claro que a mayor dispersión de residuos menor es la eficacia del modelo.
- (b) *Análisis de curva ROC*: Corresponde a la representación gráfica de la sensibilidad (capacidad para detectar asistencia en pacientes que realmente asisten) y la especificidad (capacidad para detectar inasistencia en pacientes inasistentes) del modelo. De esta forma, el modelo que refleje mejor dichas categorías en forma conjunta, será el modelo a escoger. Lo anterior se realiza mediante el análisis del Área bajo la curva ROC (A.U. ROC Curve), la cual indica un ajuste perfecto si el modelo tiene un área igual a 1 (modelo óptimo).
- (c) *Porcentaje de observaciones correctamente clasificadas*: Corresponde al porcentaje de observaciones que fueron predichas correctamente. Lo anterior se realiza mediante la elección del punto de corte (Cutoff), el cual indica bajo que supuesto de probabilidad será la discriminación de comportamiento. Para efectos del presente estudio, se utilizará el Cutoff por defecto igual a 0.5, es decir, si la probabilidad de asistencia predicha por el modelo es mayor a 0.5, dicha cita será clasificada como asistencia, se concluye de igual forma para el caso contrario.

⁸Si bien es cierto el poder discriminatorio es parte de la bondad de ajuste, se expone mediante una tercera arista para reflejar métricas asociadas al nivel de predicción del modelo.

5.2.1. Modelo estructural con efectos fijos

Modelo I: Modelo sin historial de comportamiento

La forma funcional base que se utilizara corresponde a la descrita en la ecuación 5.1, dicha función fue formulada a partir de un proceso iterativo⁹ de ajuste, donde se incluyen variables asociada a la cita i solicitada por un paciente j en un momento determinado t , sin embargo, se excluye la información histórica del paciente. De igual forma se definen los conjuntos: Conjunto de citas (I) y conjunto de pacientes (J). La variable asociada a la utilidad es $\nu_{ijt}^{(n)}$, variable dicotómica que toma el valor de 1 para asistencias y 0 para inasistencias, el exponente de la utilidad (n) se utilizará para identificar a la función. El conjunto de covariables utilizadas se encuentran definidas en la sección 3.4. *Selección de variables* donde las componentes de las sumatorias asociadas corresponden a las variables categóricas como especialidad, día de la semana, entre otras, donde se excluye del conjunto a una elección base para efectos de identificación¹⁰. El conjunto al que pertenece k en las sumatorias se encuentra especificado en la descripción de cada variable categórica en la sección 3.4. del presente estudio.

$$\begin{aligned}
\nu_{ijt}^{(1)} = & \beta_0 + \sum_{k \in ESP \setminus \{m.g.\}} \beta_{1k} esp_{ij}^k + \beta_2 centro_{ij} + \beta_3 scupo_{ij} + \beta_4 web_{ij} + \beta_5 conf_{ij} + \beta_6 gender_j \\
& + \beta_7 edad_{jt} + \beta_8 edad_{jt}^2 + \beta_9 (edad_{jt} \cdot mail_{ij}) + \beta_{10} (gender_j \cdot edad_{jt}) + \beta_{11} ffijo_{ij} + \beta_{12} fmovil_{ij} \\
& + \beta_{13} ftrab_{ij} + \beta_{14} mail_{ij} + \beta_{15} (ffijo_{ij} \cdot fmovil_{ij}) + \beta_{16} (ffijo_{ij} \cdot ftrab_{ij}) + \beta_{17} (ffijo_{ij} \cdot mail_{ij}) \\
& + \beta_{18} (fmovil_{ij} \cdot mail_{ij}) + \beta_{19} (ftrab_{ij} \cdot mail_{ij}) + \beta_{20} (fmovil_{ij} \cdot ftrab_{ij}) \\
& + \beta_{21} (ffijo_{ij} \cdot fmovil_{ij} \cdot mail_{ij}) + \beta_{22} (ffijo_{ij} \cdot fmovil_{ij} \cdot email_{ij}) + \beta_{23} (ffijo_{ij} \cdot ftrab_{ij} \cdot email_{ij}) \\
& + \beta_{24} (fmovil_{ij} \cdot ftrab_{ij} \cdot email_{ij}) + \beta_{25} (ffijo_{ij} \cdot fmovil_{ij} \cdot trab_{ij}) \\
& + \beta_{26} (ffijo_{ij} \cdot fmovil_{ij} \cdot trab_{ij} \cdot email_{ij}) + \beta_{27} manifest_t + \beta_{28} temp_t + \sum_{k \in M \setminus \{ene\}} \beta_{29k} mes_t^k \\
& + \sum_{k \in M \setminus \{ene\}} \beta_{30k} (mes_t^k \cdot temp_t) + \sum_{k \in DS \setminus \{lun\}} \beta_{31k} diasem_t^k + \sum_{k \in HG \setminus \{1\}} \beta_{32k} horag_t^k \\
& + \sum_{k \in DS \setminus \{lun\}, l \in HG \setminus \{1\}} \beta_{33kl} (diasem_t^k \cdot horag_t^l) + \sum_{k \in DG \setminus \{1\}} \beta_{34k} diag_t^k + \beta_{35} tc_{ij} + \beta_{36} dist_j \\
& + \beta_{37} dist_j^2 + \beta_{38} time_{jt}^2 + \beta_{39} deporte_t + \beta_{40} preferiado_t + \beta_{41} feriado_t + \beta_{42} postferiado_t \\
& + \sum_{k \in PG \setminus \{1\}} \beta_{43k} lluviag_t^k + \beta_{44} gescaec_{ij} + \beta_{45} post_t + \beta_{46} laten_{ij} + \beta_{47} ipc_t + \beta_{48} antig_{jt}, \\
& cluster(idpaciente) \quad \forall i \in I, j \in J
\end{aligned} \tag{5.1}$$

5.1: Modelo I: Especificación de utilidad para el modelo estructural de efectos fijos sin historial de comportamiento

(Fuente: Elaboración propia)

A partir de la forma funcional de la utilidad anterior se utiliza la función de enlace LOGIT para establecer la probabilidad de asistencia de un paciente (ver ecuación 5.2). Los coeficientes asociados se encuentran especificados en *Apéndice A.1. Ajuste de coeficientes*. Es importante notar que existen interacciones estadísticamente no significativas, sin embargo, se deben mantener en el

⁹El proceso iterativo se realiza de acuerdo a lo establecido en la fase 5 de la metodología y su variación (ver sección 1.7. Metodología).

¹⁰Para la estructuración de un modelo con variables categóricas, se debe omitir en la ecuación una categoría pivote, la cual servirá para identificar variaciones de las alternativas sobre dicho pivote.

modelo dado que niveles mayores de interacciones¹¹ resultaron significativas, claro es el caso de la interacción de nivel 4 asociada al canal de comunicación (fono fijo, móvil, trabajo y email). Los Odds Ratios calculados para el modelo de la ecuación 5.1 se encuentran en *Apéndice A.2. Odds ratios para modelos con ajuste Logit y Mixed Logit*.

$$P(\text{estado}_{ijt} = 1) = \frac{e^{\nu_{ijt}}}{1 + e^{\nu_{ijt}}} \quad (5.2)$$

5.2: Probabilidad de asistencia utilizando la función de enlace LOGIT para un modelo con utilidad ν_{ijt} (Fuente: Elaboración propia)

<i>Descripción</i>	Logit Modelo I
Observaciones	300000
LL	-126831.143
McFadden's R^2	0.197
McFadden's R^2_{adjust}	0.196
A.U. ROC Curve	0.7772
AIC	253920.3
BIC	255289.2
Correctly Classified (Cutoff 0.5)	82.94 %

Tabla 5.1: Resumen de ajuste Modelo I con función de enlace LOGIT (Fuente: Elaboración propia)

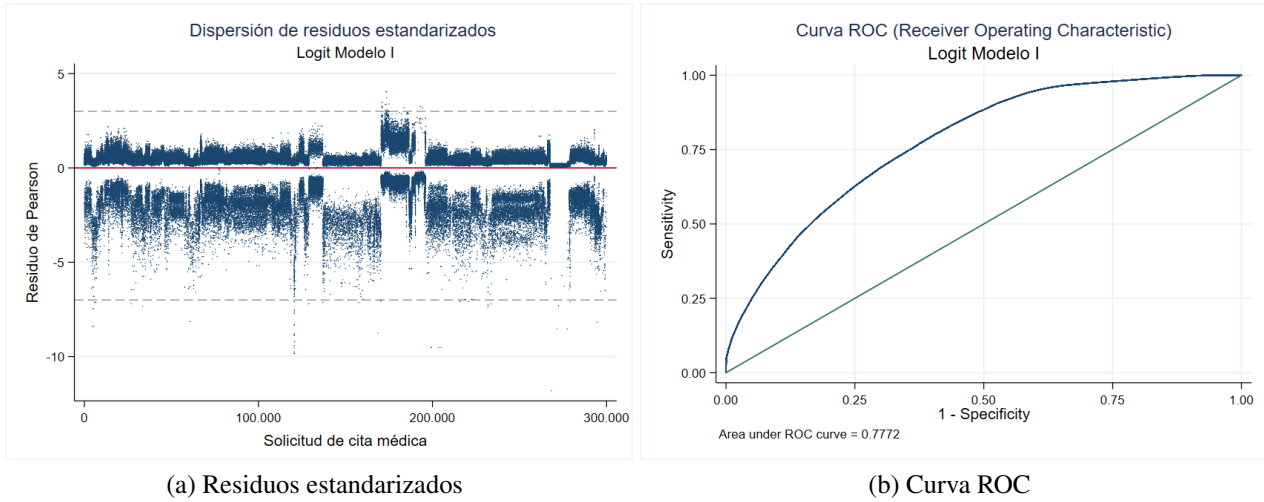
En la tabla 5.1 se puede observar que el Pseudo- R^2 indica un nivel de ajuste de 19.7 %, clasificándolo como un modelo de ajuste bajo, su índice ajustado muestra una reducción de 0.1 punto porcentual, lo que indica que las covariables son predictores significativos para el modelo. Se establecen además los criterios de información base AIC y BIC. En cuanto al poder discriminatorio del modelo, se observa que el 82.94 % de las observaciones fueron correctamente clasificadas. El área bajo la curva ROC es igual a 0.777 (ver figura 5.2a) . En cuanto a los residuos de Pearson (ver figura 5.2b), se puede observar dispersión en su mayoría bordeando el intervalo $[3, -7]$, con varios residuos que se escapan del intervalo mencionado debido a que el modelo no ajusta correctamente a dichas observaciones. Finalmente, el test de Wald indica significatividad conjunta del modelo¹².

Modelo II: Modelo con historial de comportamiento asociado a 1 rezago

Para mejorar el ajuste del Modelo I, se recurre a la adición de información histórica relacionada a 1 rezago (comportamiento del paciente j en la cita $i - 1$). De esta forma, se construye un nuevo modelo llamado Modelo II descrito en la ecuación 5.3, donde la utilidad queda expresada por $\nu_{ijt}^{(2)}$ con la misma interpretación de coeficientes que el modelo anterior. El Modelo II incorpora variables de comportamiento y sus interacciones de nivel. Es importante mencionar que al igual que el Modelo I, la adición de variables de interacción de nivel superior se realiza mediante un proceso iterativo de ajuste, donde se integran sólo interacciones de nivel superior significativas, de

¹¹Se denomina nivel n de interacción a interacciones entre n covariables.

¹²Ver *Anexo A.1. Ajuste de coeficientes*, Logit Model I.



(a) Residuos estandarizados (b) Curva ROC

Figura 5.2: Gráficos de ajuste de Modelo I mediante función de enlace LOGIT
(Fuente: Elaboración propia)

igual forma, pese a que ciertas interacciones de nivel inferior no sean significativas, se deben incluir de igual forma para mantener la coherencia del modelo¹³.

$$\begin{aligned}
 \nu_{ijt}^{(2)} = & \nu_{ijt}^{(1)} + \beta_{49}lagscupo_{ij} + \beta_{50}(scupo_{ij} \cdot lagscupo_{ij}) + \beta_{51}lagweb_{ij} + \beta_{52}(web_{ij} \cdot lagweb_{ij}) \\
 & + \beta_{53}lag_{ij}^{(1)} + \beta_{54}lagprof_{ij}^{(1)} + \beta_{55}lagesp_{ij}^{(1)} + \beta_{56}(lag_{ij}^{(1)} \cdot lagprof_{ij}^{(1)}) + \beta_{57}(lag_{ij}^{(1)} \cdot lagesp_{ij}^{(1)}) \\
 & + \beta_{58}(lagprof_{ij}^{(1)} \cdot lagesp_{ij}^{(1)}) + \beta_{59}(lag_{ij}^{(1)} \cdot lagprof_{ij}^{(1)} \cdot lagesp_{ij}^{(1)}) + \beta_{60}laglaten_{ij} \\
 & + \beta_{61}(lag_{ij}^{(1)} \cdot laglaten_{ij}), \quad cluster(idpaciente) \quad \forall i \in I, j \in J
 \end{aligned}
 \tag{5.3}$$

5.3: Modelo II: Especificación de utilidad para el modelo estructural de efectos fijos con historial de comportamiento asociado a 1 rezago (cita anterior)
(Fuente: Elaboración propia)

De igual forma que el modelo anterior, se utiliza la función de enlace LOGIT para establecer la probabilidad de asistencia de un paciente (ver ecuación 5.2). Los coeficientes asociados se encuentran especificados en *Apéndice A.1. Ajuste de coeficientes*. Los Odds ratios calculados para el modelo de la ecuación 5.3 se encuentran en *Apéndice A.2. Odds ratios para modelos con ajuste Logit y Mixed Logit*.

Al igual que el modelo anterior, el nuevo modelo ajustado presenta significancia conjunta. En la tabla 6.1 se puede observar que al agregar variables de comportamiento de rezago 1, el ajuste mostrado por el $Pseudo-R^2$ es sustancial, mejorando en aproximadamente 7 puntos porcentuales, llegando hasta un 26.8 %. En la gráfica 5.3b, se observa la dispersión de residuos estandarizados, los cuales se encuentran gráficamente más uniformes y en su mayoría dentro del intervalo $[3, -5]$. Se observa además que el nivel de ajuste no es afectado por el número de parámetros dado que son

¹³En el Modelo II se agregan interacciones como sobrecupo y su lag o rezago correspondiente, sin embargo, se agrega sólo su interacción y la variable de rezago asociada, obviando la variable de sobrecupo ya que esta se encuentra incluida en el Modelo I.

<i>Descripción</i>	Logit Modelo I	Logit Modelo II
Observaciones	300000	300000
LL	-126831.143	-115572.791
McFadden's R^2	0.197	0.268
McFadden's R^2_{adjust}	0.196	0.267
A.U. ROC Curve	0.7772	0.8284
AIC	253920.3	231431.6
BIC	255289.2	232949
Correctly Classified (Cutoff 0.5)	82.93 %	83.55 %

Tabla 5.2: Resumen de ajuste Modelo I y II con función de enlace LOGIT
(Fuente: Elaboración propia)

estadísticamente significativos, de esta forma el $Pseudo - R^2_{adjust}$ no presenta mayor variación en comparación al $Pseudo - R^2$. Los criterios de información de AIC y BIC son menores, lo que demuestra que el Modelo II es superior a su predecesor. De la misma forma, el área bajo la curva ROC aumenta hasta 0.8284, mejorando significativamente su nivel de especificidad y sensibilidad para el punto de corte establecido (ver gráfico 5.3a), finalmente el modelo estima el 83.55 % de las observaciones correctamente, aumentando en 0.62 puntos porcentuales en comparación al ajuste del Modelo I.

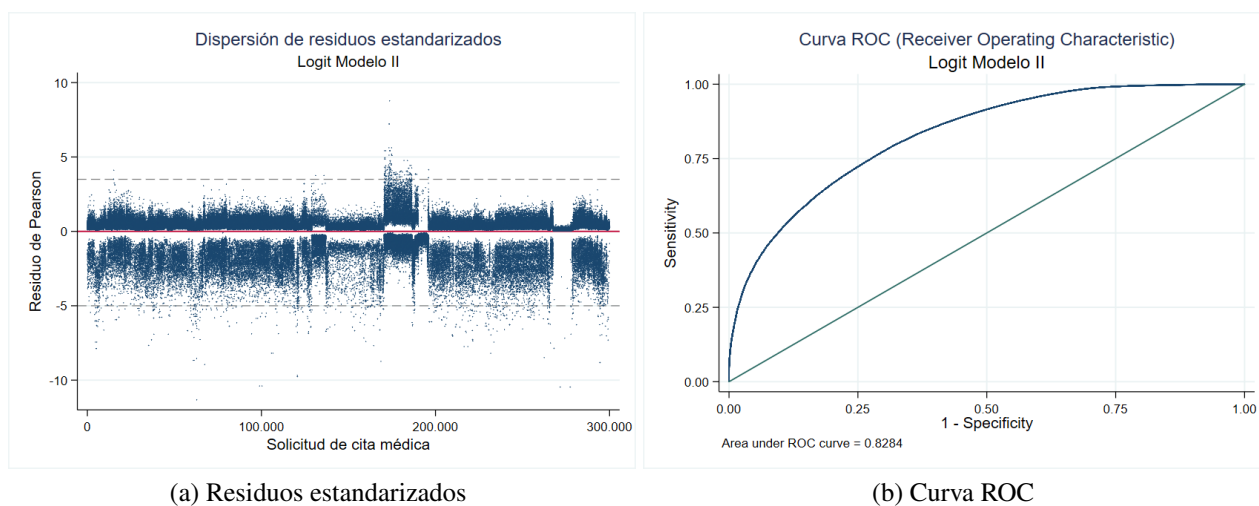


Figura 5.3: Gráficos de ajuste de Modelo II mediante función de enlace LOGIT
(Fuente: Elaboración propia)

Modelo III: Modelo con historial de comportamiento asociado a 1 y 2 rezagos junto al nivel de asistencia y fiabilidad histórico del paciente

Dado que la información histórica del paciente frente a sus elecciones generó una mejora significativa en el modelo, se decide incorporar información histórica asociada a 1 y 2 rezagos, junto con el nivel de asistencia del paciente y su índice de fiabilidad¹⁴. Para ello, se define un nuevo modelo

¹⁴Variables definidas en el capítulo 3, sección 3.4. Selección de variables.

de utilidad cuya variable se define como $\nu_{ijt}^{(3)}$, el cual incorpora nuevas interacciones significativas de orden superior (ver ecuación 5.4). Al igual que el modelo anterior, el ajuste de dichas variables se obtuvo mediante un proceso iterativo de ajuste del modelo. De igual forma, dado que existen variables de interacción de orden superior estadísticamente significativas, se deben incorporar interacciones de orden inferior asociadas para mantener la coherencia del modelo (independiente de su significancia individual).

$$\begin{aligned}
\nu_{ijt}^{(3)} = & \nu_{ijt}^{(2)} + \beta_{63}lag_{ij}^{(2)} + \beta_{64}lag_{ij}^{(3)} + \beta_{65}(lag_{ij}^{(1)} \cdot lag_{ij}^{(2)}) + \beta_{66}(lag_{ij}^{(1)} \cdot lag_{ij}^{(3)}) + \beta_{67}(lag_{ij}^{(2)} \cdot lag_{ij}^{(3)}) \\
& + \beta_{68}(lag_{ij}^{(1)} \cdot lag_{ij}^{(2)} \cdot lag_{ij}^{(3)}) + \beta_{69}lagprof_{ij}^{(2)} + \beta_{70}lagesp_{ij}^{(2)} + \beta_{71}(lag_{ij}^{(2)} \cdot lagprof_{ij}^{(2)}) \\
& + \beta_{72}(lag_{ij}^{(2)} \cdot lagesp_{ij}^{(2)}) + \beta_{73}(lagprof_{ij}^{(2)} \cdot lagesp_{ij}^{(2)}) + \beta_{74}(lag_{ij}^{(2)} \cdot lagprof_{ij}^{(2)} \cdot lagesp_{ij}^{(2)}) \\
& + \beta_{75}lagpor_{ij} + \beta_{76}lagfiab_{ij} + \beta_{77}(lagpor_{ij} \cdot lagfiab_{ij}), \quad cluster(idpaciente) \quad \forall i \in I, j \in J
\end{aligned}
\tag{5.4}$$

5.4: Modelo III: Especificación de utilidad para el modelo estructural de efectos fijos con historial de comportamiento asociado a 1 y 2 rezagos junto al nivel de asistencia y fiabilidad histórico del paciente (Fuente: Elaboración propia)

Para ajustar el modelo, se utiliza nuevamente la función de enlace LOGIT para establecer la probabilidad de asistencia de un paciente (ver ecuación 5.2). De igual forma, los coeficientes asociados se encuentran especificados en *Apéndice A.1. Ajuste de coeficientes* y los Odds ratios se encuentran especificados en *Apéndice A.2. Odds ratios para modelos con ajuste logit y Mixed Logit*.

<i>Descripción</i>	Logit Modelo I	Logit Modelo II	Logit Modelo III
Observaciones	300000	300000	300000
LL	-126831.143	-115572.791	-106955.510
McFadden's R^2	0.197	0.268	0.323
McFadden's R^2_{adjust}	0.196	0.267	0.321
A.U. ROC Curve	0.7772	0.8284	0.8597
AIC	253920.3	231431.6	214227
BIC	255289.2	232949	215903.6
Correctly Classified (Cutoff 0.5)	82.93 %	83.55 %	84.32 %

Tabla 5.3: Resumen de ajuste Modelo I, II y III con función de enlace LOGIT (Fuente: Elaboración propia)

En la tabla 5.3 se observan los resultados del ajuste del Modelo III, una vez más existe significancia conjunta de los parámetros. Al incorporar rezagos de orden superior e información histórica del comportamiento y niveles de confiabilidad del paciente, se genera un aumento significativo en el nivel de ajuste, claro es el caso del $Pseudo - R^2$, el cual aumenta 5.5 puntos porcentuales llegando a 32.3 %, el $Pseudo - R^2_{adjust}$ no presenta mayor variación con respecto al $Pseudo - R^2$ debido a que las nuevas covariables son estadísticamente significativas. El gráfico 5.4b muestra la dispersión de los residuos estandarizados, los cuales presentan mayor homogeneidad, estando en mayor medida dentro del intervalo $[3, -4]$, lo que indica un mejor ajuste. Los criterios de información AIC y BIC son menores al Modelo II, lo que indica superioridad del Modelo III. El área bajo la curva ROC es 0.8597, lo cual indica mejor comportamiento en la sensibilidad y especificidad del

modelo en comparación al Modelo I y II. Finalmente, el modelo presenta un poder predictivo de 84.32 % de observaciones correctamente clasificadas.

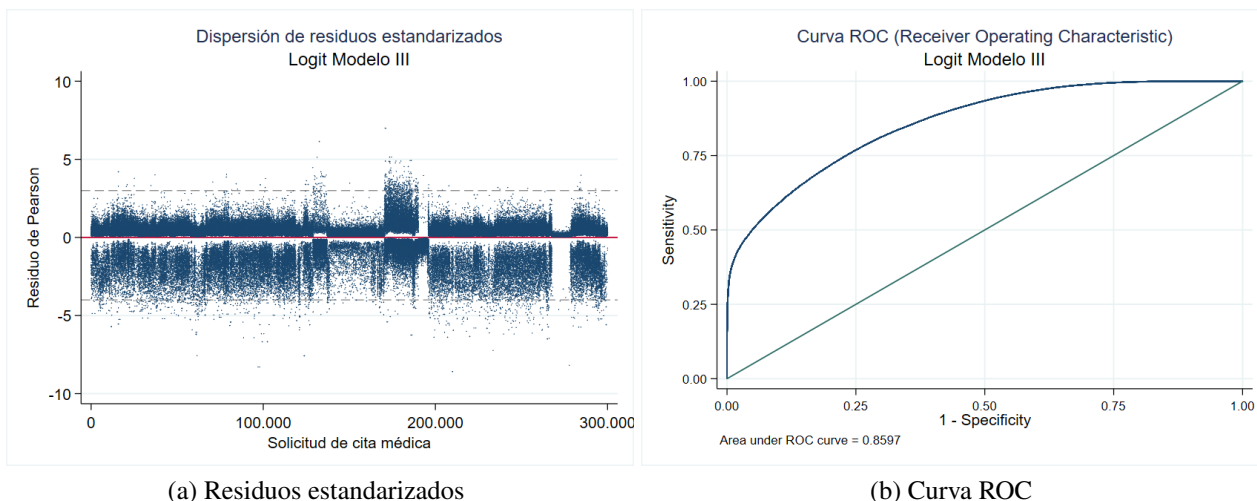


Figura 5.4: Gráficos de ajuste de Modelo III mediante función de enlace LOGIT
(Fuente: Elaboración propia)

Dado que a medida que se fue agregando información con respecto al comportamiento histórico del paciente, el ajuste del modelo y su poder predictivo aumenta abruptamente, lo que indica que dicha información es fundamental para analizar y pronosticar el comportamiento de asistencia a las citas médicas. Dado lo anterior, se decide utilizar una nueva función de enlace para testear el modelo y estudiar si la nueva función de enlace permite ajustar aún más el Modelo III. La función de enlace escogida fue la función PROBIT, cuya expresión de probabilidad asociada a la utilidad $\mu_{ijt}^{(3)}$ del Modelo III queda expresada mediante la ecuación 5.5

$$P(\text{estado}_{ijt} = 1) = \int_{-\infty}^{\nu_{ijt}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (5.5)$$

5.5: Probabilidad de asistencia utilizando la función de enlace PROBIT para un modelo con utilidad ν_{ijt}
(Fuente: Elaboración propia)

Los coeficientes de la estimación del Modelo III mediante la función de enlace PROBIT se encuentran especificados en *Apéndice A.1. Ajuste de coeficientes*.

En el cuadro 5.4, se puede observar que el nivel de ajuste presentado por el $Pseudo - R^2$ es levemente mayor, de igual forma el área bajo la curva ROC denota una imperceptible mejoría. Los resultados correctamente clasificados sufren una disminución con respecto al modelo ajustado mediante la función de enlace LOGIT de 0.01 %, lo cual es despreciable. Es así como se determina que la función de enlace PROBIT es más adecuada para el modelo.

5.2.2. Modelo estructural con efectos fijos y aleatorios

Hasta el momento la información histórica del comportamiento del paciente ha sido fundamental para el nivel de ajuste del modelo. A partir de ello, se decide integrar heterogeneidad no observable

<i>Descripción</i>	Logit Modelo III	Probit Modelo III
Observaciones	300000	300000
LL	-106955.510	-106174.844
McFadden's R^2	0.323	0.328
McFadden's R^2_{adjust}	0.321	0.326
A.U. ROC Curve	0.8597	0.86
AIC	214227	212665.7
BIC	215903.6	214342.3
Correctly Classified (Cutoff 0.5)	84.32 %	84.31 %

Tabla 5.4: Resumen de ajuste del Modelo III con función de enlace LOGIT y PROBIT
(Fuente: Elaboración propia)

al modelo con el motivo de ajustar comportamientos del paciente no explicados por los predictores actuales, este comportamiento representa efectos asociados a la heterogeneidad entre pacientes y grado de homogeneidad en las decisiones tomadas por cada paciente, así lograr un mejor ajuste del modelo. La muestra registra 77141 pacientes, los cuales serán modelados como clusters, de esta forma, para modelar los efectos aleatorios es necesario reducir el número de covariables del modelo con el fin de lograr el ajuste con la tecnología disponible¹⁵. Para reducir el número de covariables, se utiliza el criterio de evaluación de menor impacto, donde se extraer las covariables que generan menores impactos al ajuste del comportamiento del paciente. De esta forma el modelo se define como Modelo IV cuya especificación está dada por la ecuación 5.6.

$$\begin{aligned}
\eta_{ijt} = & \beta_0 + \sum_{k \in ESP \setminus \{m.g.\}} \beta_{1k} esp_{ij}^k + \beta_2 centro_{ij} + \beta_3 scupo_{ij} + \beta_4 web_{ij} + \beta_5 conf_{ij} + \beta_6 gender_j \\
& + \beta_7 edad_{jt} + \beta_8 (gender_j \cdot edad_{jt}) + \beta_9 f fijo_{ij} + \beta_{10} f movil_{ij} + \beta_{11} f trab_{ij} + \beta_{12} mail_{ij} \\
& + \beta_{13} manifest_t + \beta_{14} temp_t + \sum_{k \in M \setminus \{ene\}} \beta_{15k} mes_t^k + \sum_{k \in M \setminus \{ene\}} \beta_{16k} (mes_t^k \cdot temp_t) \\
& + \sum_{k \in DS \setminus \{lun\}} \beta_{17k} diasem_t^k + \sum_{k \in HG \setminus \{1\}} \beta_{18k} horag_t^k + \sum_{k \in DG \setminus \{1\}} \beta_{19k} diag_t^k + \beta_{20} lag_{ij}^{(1)} \\
& + \beta_{21} tc_{ij} + \beta_{22} (lag_{ij}^{(1)} \cdot tc_{ij}) + \beta_{23} dist_j + \beta_{24} time_{jt} + \beta_{25} lag_{ij}^{(1)} + \beta_{26} lagprof_{ij}^{(1)} + \beta_{27} lagesp_{ij}^{(1)} \\
& + \beta_{28} (lag_{ij}^{(1)} \cdot lagprof_{ij}^{(1)}) + \beta_{29} (lag_{ij}^{(1)} \cdot lagesp_{ij}^{(1)}) + \beta_{30} (lagprof_{ij}^{(1)} \cdot lagesp_{ij}^{(1)}) \\
& + \beta_{31} (lag_{ij}^{(1)} \cdot lagprof_{ij}^{(1)} \cdot lagesp_{ij}^{(1)}) + \beta_{32} lagpor_{ij} + \beta_{33} lagfiab_{ij} + \beta_{34} (lagpor_{ij} \cdot lagfiab_{ij}) \\
& + \beta_{35} deporte_t + \beta_{36} preferiado_t + \beta_{37} feriado_t + \beta_{38} postferiado_t + \sum_{k \in PG \setminus \{1\}} \beta_{39k} lluviag_t^k \\
& + \beta_{40} gescaec_{ij} + \beta_{41} post_i + \beta_{42} laten_{ij} + \beta_{43} ipc_t + \beta_{44} antig_{jt} + \zeta_j, \quad \forall i \in I, j \in J, \zeta_j \in N(0, \psi)
\end{aligned} \tag{5.6}$$

5.6: Modelo IV: Especificación de utilidad para el modelo estructural de efectos aleatorios sobre cada paciente

(Fuente: Elaboración propia)

El Modelo IV corresponde a un modelo con intercepto aleatorio, donde la utilidad se expresa mediante η_{ijt} , cuyo término estocástico (heterogeneidad no observable) corresponde a ζ_j para cada

¹⁵ver sección 4.1. Validación de la muestra, apartado Selección de la muestra.

paciente j . El término estocástico distribuye normal con media 0 y varianza ψ . La función de enlace utilizada en el modelo corresponde a la extensión de LOGIT integrando la componente estocástica mediante la función de densidad de la distribución normal ϕ (ver ecuación 5.7) a la cual en adelante se le denominara MIXED LOGIT.

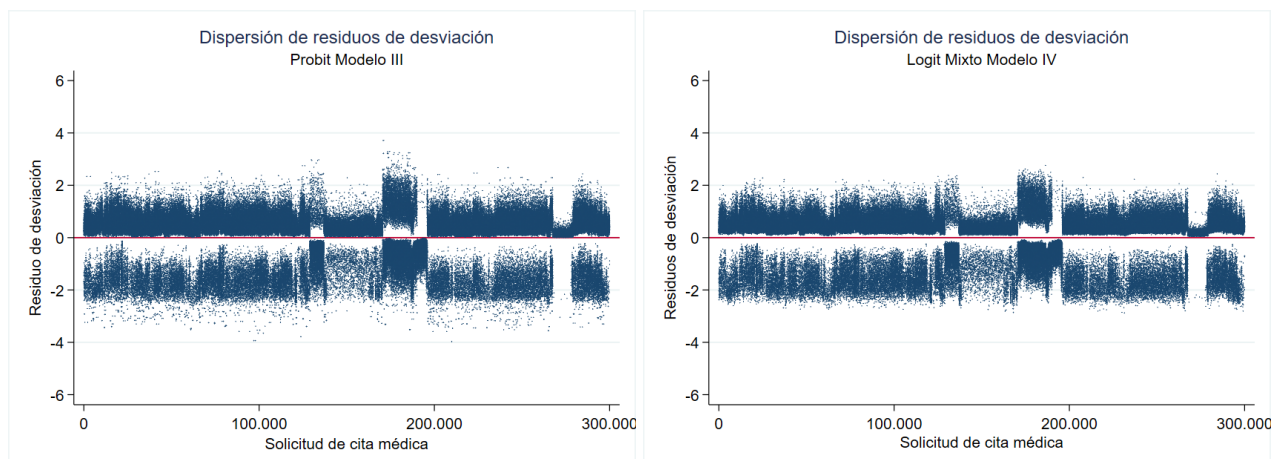
$$P(\text{estado}_{ijt} = 1 | \zeta_j) = \int \frac{e^{\eta_{ijt}}}{1 + e^{\eta_{ijt}}} \phi(\zeta, \psi) d\zeta \quad (5.7)$$

5.7: Probabilidad de asistencia utilizando la función de enlace MIXED LOGIT para un modelo con utilidad

η_{ijt}
(Fuente: Elaboración propia)

Los coeficientes del Modelo IV ajustado mediante la función de enlace MIXED LOGIT se encuentran especificados en *Apéndice A.1. Ajuste de coeficientes* y los Odds ratios se encuentran especificados en *Apéndice A.2. Odds ratios para modelos con ajuste logit y Mixed Logit*.

Para comparar el ajuste del Modelo III mediante la función de enlace PROBIT y el modelo IV mediante la función de enlace MIXED LOGIT se analiza de forma gráfica los residuos de desviación (ver figura 5.5). Se puede observar que el Modelo IV ajustado por MIXED LOGIT posee menor dispersión, de esta forma genera predicciones más precisas acerca del comportamiento del paciente.



(a) Residuos de desviación PROBIT Modelo III

(b) Residuos de desviación MIXED LOGIT Modelo IV

Figura 5.5: Gráficos de ajuste de Modelo III mediante la función de enlace PROBIT y del Modelo IV mediante la función de enlace MIXED LOGIT

(Fuente: Elaboración propia)

Se observa además que el Modelo IV con función de enlace MIXED LOGIT posee un poder predictivo superior al del Modelo III con función de enlace PROBIT, llegando a un 85.42 % considerando Cutoff igual a 0.5. En el gráfico 5.6, se observa una curva de relación sensibilidad/especificidad más suavizada para el modelo Mixto (ver gráfico 5.6b), además el área bajo la curva del modelo mixto asciende a 0.8879, superior a 0.86 del Modelo PROBIT.

Para evaluar la correlación intraclase, se utiliza el ratio Intraclass Correlation (ICC), el cual mantiene como parámetros el valor 0 para expresar que todos los pacientes tienen las mismas tendencias

de comportamiento y la variabilidad se encuentra en cada solicitud de cita médica y 1 para indicar que toda la variabilidad del comportamiento del paciente es explicada por la heterogeneidad entre pacientes (Ver figura 5.8). Se puede observar que el 47.03 % de la variabilidad del comportamiento en la asistencia médica puede explicarse por las diferencias en las características de cada paciente (heterogeneidad entre pacientes). Dicha variabilidad presenta un error estándar de 0.0038.

$$ICC = \frac{Var(idpaciente)}{Var(idpaciente) + Var(residuos)} = 0,4703 \quad (5.8)$$

5.8: Ratio de correlación intra-clases
(Fuente: Elaboración propia)

Dados los antecedentes de ajuste del modelo, se escoge el modelo IV con función de enlace MIXED LOGIT para desarrollar el foco explicativo del modelo.

En la figura

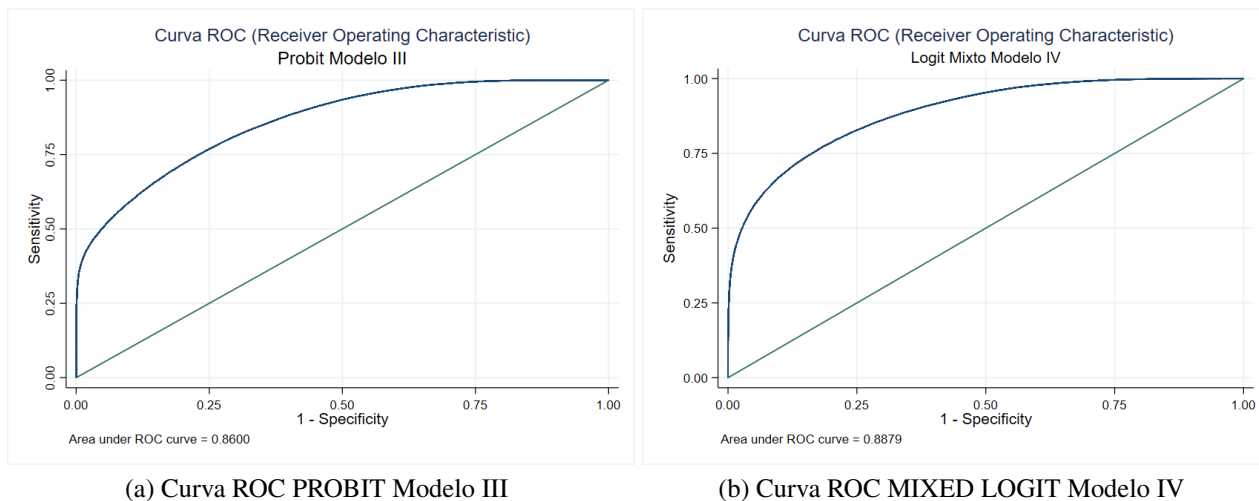


Figura 5.6: Curva ROC del Modelo III y IV on funciones de enlace PROBIT y MIXED LOGIT respectivamente

(Fuente: Elaboración propia)

5.3. Análisis del modelo

5.3.1. Análisis del modelo con foco explicativo y validación de hipótesis

Una vez comprobado el modelo que presenta mejor ajuste, se procede a estimar computacionalmente las probabilidades de asistencia para cada solicitud de cita médica. Mediante el modelo, es posible analizar contrafactuales¹⁶, de esta forma se calculan probabilidades bajo el supuesto de variación en las características específicas de las covariables (edad, género, horario, etc.) se utilizarán los *Average Adjusted Predictions (AAP's)* y *Average Margin Effects (AME)*, el primero estima el

¹⁶Se denomina contrafactual a todo acontecimiento que no necesariamente ha ocurrido, pero que eventualmente podría haber ocurrido o existe la posibilidad que ocurra en el futuro.

promedio de las probabilidades predichas (para todas las combinaciones de características específicas) mediante la función de utilidad con la función de enlace MIXED LOGIT considerando el resto de las covariables constantes en cada observación, además se calculará un intervalo de confianza al 95 % de fiabilidad para el promedio de la variable de estudio. El segundo método se utilizará para obtener el impacto de las variaciones marginales de las covariables sobre la probabilidad de comportamiento (efecto diferencial).

Probabilidad promedio de asistencia De forma general, se puede afirmar que, en promedio, la probabilidad de asistencia de un paciente a una cita médica en la institución corresponde a 78.7 % con un intervalo de confianza asociado a la muestra de [0,7862, 0,788], se puede notar un margen de error muy bajo con un nivel de confianza de 95 %.

Evaluación de edad y género Se observa que el parámetro del modelo (ver Apéndice A.1.) para la variable género no es significativa por sí sola a un nivel de confianza de 95 %, sin embargo, se incluye una interacción entre la variable género y edad, cuyo parámetro asociado posee un nivel de significancia de 99 %. Lo anterior indica que existen variaciones en las probabilidades de asistencia por edad y género respectivo, sin embargo, no se aprecian variaciones de género por sí sola a un nivel de significancia deseado.

En el gráfico 5.7 se puede observar que existen diferencias en la probabilidad de asistencia promedio por género, diferencias que se incrementan progresivamente a medida que el paciente posee mayor edad. De esta forma, para pacientes menores a 5 años de edad la probabilidad promedio de asistencia entre hombres y mujeres es similar (76 %), a partir de los 5 años de edad se observa que el género femenino posee mayor probabilidad promedio de asistencia que el género masculino, diferencia que se acentúa levemente con la edad. La diferencia promedio entre la variación marginal por año entre hombres y mujeres corresponde a 21.7 %, es decir, en cada año de edad, la mujer posee en promedio 21.7 % más de probabilidades de asistencia que la del hombre.

Validación de Hipótesis N°1 y N°2

Dada la información anterior, se validan las hipótesis 1 y 2, notando que existen leves diferencias en la probabilidad promedio de asistencia entre hombres y mujeres, las cuales se incrementan de forma paulatina con la edad. El gráfico 5.7 muestra el análisis de casos para cada edad y género de estudio. Para efectos del análisis se incluyen sólo personas con edades menores o iguales a 100 años de edad.

Evaluación de las especialidades Se puede observar que la significancia para la validación en las probabilidades predichas para las especialidades juega un papel fundamental al momento de concluir variaciones marginales (ver Apéndice A.1.).

En el gráfico 5.8 se puede observar la probabilidad promedio de especialidades cuyos factores fueron significativos al menos a un 95 % de confianza. La probabilidad promedio de la categoría *Laboratorio* fue la más alta llegando a un 89.8 % de asistencia; dentro de los tratamientos asociados a dicha categoría se pueden mencionar: tratamientos anticoagulantes, espirometría, curva de insulina/glicemia, entre otros. La segunda categoría con mayor probabilidad (85.9 %) corresponde a *Dolor*, cuyos tratamientos asociados a dicha categoría son variados: Dolor fisiatría, Dolor psiquiatría, Dolor psicología, Dolor terapia grupal, Dolor traumatología, entre otros. De esta forma,

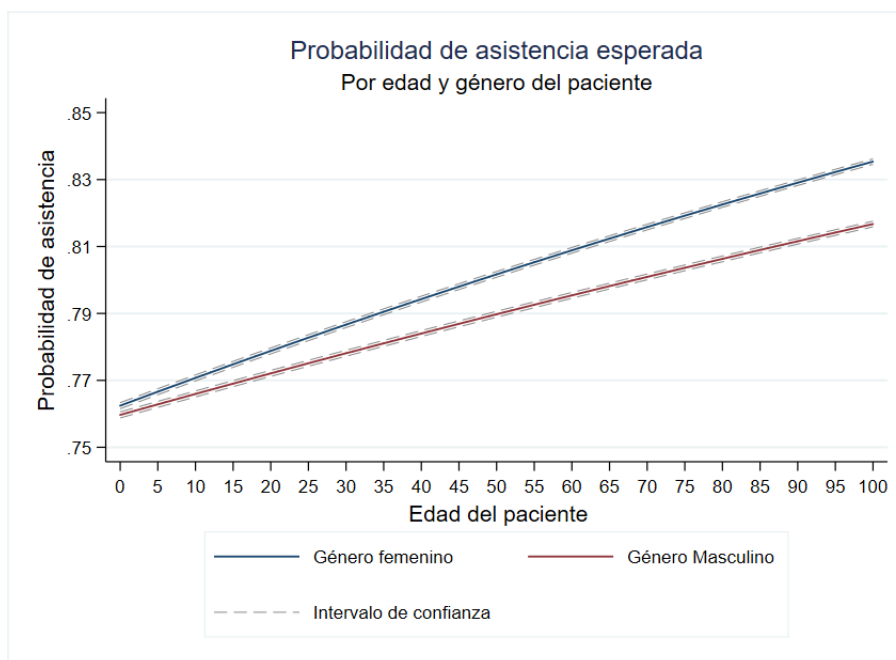


Figura 5.7: Variación de probabilidad promedio asociada a la cita médica por edad y género (Fuente: Elaboración propia)

la categoría dolor se encuentra a nivel transversal, incorporando asociación física y mental del paciente. En tercer lugar, de probabilidad promedio más alta se encuentra la especialidad *Medicina física y rehabilitación*, cuya probabilidad promedio corresponde a 84.4 %. La alta probabilidad de asistencia demuestra la importancia que el paciente percibe sobre la terapia asociada.

Las especialidades con menor estimación de probabilidad de asistencia promedio corresponden a *Investigación clínica* con un 42.8 % los cuales corresponden principalmente a proyectos. Luego con un 62.2 % de probabilidades de asistencia se encuentra la categoría *Dentomaxilofacial*, la cual involucra tratamientos como odontología general, prótesis, endodoncia, entre otras. Se destaca además la categoría *Respiratorio*, cuya probabilidad de asistencia promedio corresponde a un 67.8 %, categoría que involucra las subespecialidades Bronco pulmonar adulto y procedimientos respiratorios. Finalmente, se destaca la categoría medicina general, cuyo índice de probabilidad promedio de asistencia asciende a 82.6 %. Las categorías con índice de probabilidad promedio entre 70 % y 85 % se dividen en 3 niveles:

1. Nivel medio superior: Corresponde a categorías asociadas a una probabilidad de asistencia promedio entre 80 % y 85 %, dentro de las cuales se pueden mencionar: Inmunología y alergias, Pediatría, Medicina física, Neurología, Oftalmología, entre otras.
2. Nivel medio: Corresponde a categorías asociadas a una probabilidad de asistencia promedio entre 75 % y 80 %, dentro de las cuales se pueden mencionar: Cardiología, Obesidad y nutrición, Genética, Otorrinolaringología, entre otras.
3. Nivel medio inferior: Corresponde a categorías asociadas a una probabilidad de asistencia promedio entre 70 % y 75 %, dentro de las cuales se pueden mencionar: Oncología, Fonoaudiología, Cirugía, Radiología, entre otras.

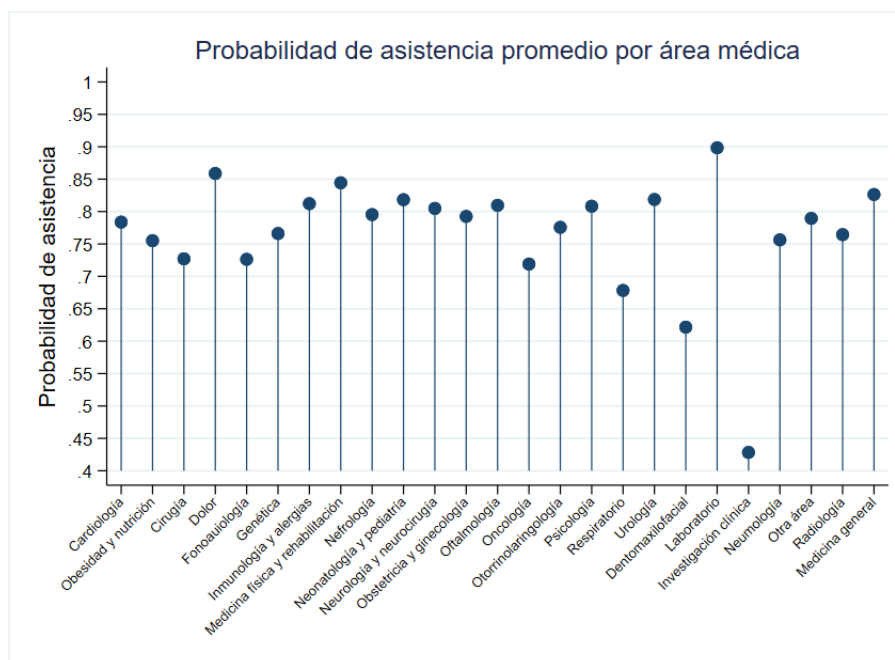


Figura 5.8: Probabilidad esperada asociada a la cita médica por especialidad (Fuente: Elaboración propia)

Validación de Hipótesis N°3

Dada la información anterior, se valida parcialmente la hipótesis N°3, ya que sólo tres especialidades mantienen mayores probabilidades promedio de asistencia que medicina general, estas son: Laboratorio, dolor y medicina física y rehabilitación.

Evaluación del canal de agendamento Se puede observar que la probabilidad de asistencia promedio para citas médicas agendadas vía web corresponde a un 77.92 %, mientras que el agendamento realizado por otro canal mantiene una probabilidad de asistencia promedio de 78.88 %.

Validación de Hipótesis N°4

Dada la información anterior, se valida la hipótesis N°4, donde el canal web genera menor probabilidad de asistencia promedio a citas médicas. Sin embargo, es importante destacar que la variación en probabilidad es cercana al 1 %, por lo que la hipótesis se valida con precaución.

Evaluación de sobrecupos La diferencia marginal promedio de probabilidades asociada a sobrecupos es cercana a 5 %. Donde la probabilidad promedio de asistencia a una cita médica solicitada como sobrecupo corresponde a 83.12 %, probabilidad superior al 78.88 % estimado para citas médicas agendadas como citas de agenda regular.

Validación de Hipótesis N°5

Se valida la hipótesis N°5, la cual señala que los sobrecupos afectan positivamente la probabilidad de asistencia del paciente.

Evaluación de la gestión de confirmación La variación marginal promedio de la probabilidad de la gestión de confirmación sobre la probabilidad de asistencia del paciente es cercana a 6.1 %, donde la probabilidad promedio de asistencia de pacientes confirmados asciende a 82.72 %, superior a la probabilidad promedio de asistencia estimada para pacientes no confirmados, la cual asciende a 76.63 %.

Validación de Hipótesis N°6

Se valida la hipótesis N°6, la cual señala que la confirmación genera un impacto positivo en la asistencia del paciente. Dicho impacto corresponde a un aumento marginal de 6.1 %.

Evaluación de un eventual aumento de latencia ¹⁷ Para analizar el efecto que genera una política de extensión de agenda es importante analizar el comportamiento que posee el paciente que solicita una cita médica para el corto, mediano y largo plazo. De acuerdo a la definición de *Latencia* propuesta en la sección 3.3. *Análisis descriptivo*, el análisis del comportamiento del paciente frente a variaciones marginales de la latencia responde al contrafactual de inasistencia frente a periodos de tiempos definidos. El gráfico 5.9 muestra claramente que a medida que aumenta la latencia de la cita médica, la probabilidad de asistencia disminuye a razón de 5 %. Se destaca que la probabilidad de asistencia esperada para citas con latencia menor o igual a 7 días sobrepasa el 80 %. En términos marginales, por cada día de diferencia entre la solicitud de la cita y el día de atención, se estima una variación marginal de la probabilidad estimada de $-0,2\%$ en las probabilidades de asistir a la cita médica agendada. Para reconstruir sin la estimación de probabilidades ajustadas la variación marginal de la probabilidad dada una variación marginal en la variable latencia, se puede aproximar mediante un factor inicial pivote $L_0 = 0,808517$ y aplicar un decrecimiento sostenido mediante un factor $\varepsilon_0 = 0,00222487$. finalmente se obtendrá una aproximación de probabilidades \hat{P}_l para testear políticas asociadas a la latencia (ver capítulo 5).

Validación de Hipótesis N°7

Dado el análisis anterior se rechaza la hipótesis acerca de la relación proporcional directa entre latencia y asistencia

Evaluación de horarios de atención De acuerdo a la figura 5.10, la probabilidad promedio más alta de asistencia se encuentra en los días miércoles y jueves, mientras que la probabilidad promedio más baja se puede encontrar los días sábados. En cuanto a los bloques horarios de atención, se destaca la atención de 10 AM a 12 PM, con una probabilidad máxima de 79.6 % para el día jueves mientras que la probabilidad más baja para este bloque, se encuentra en los días sábado con un 76.9 % de probabilidades de asistencia promedio. La probabilidad promedio más baja de asistencia se obtiene en las citas entre las 08:00 AM y 09:00 AM, con una máxima de 78 % para los días Lunes y probabilidad mínima de 76.8 % los días sábado

¹⁷Corresponde al aumento de la disponibilidad de citas por agenda aumentando el tiempo de actividad de la misma. A modo de ejemplo, si antes de la extensión un paciente tenía la posibilidad de agendar una cita dentro de los siguientes dos meses, luego de la extensión de agenda a T meses, el paciente podría eventualmente escoger una cita disponible dentro de los siguientes T meses.

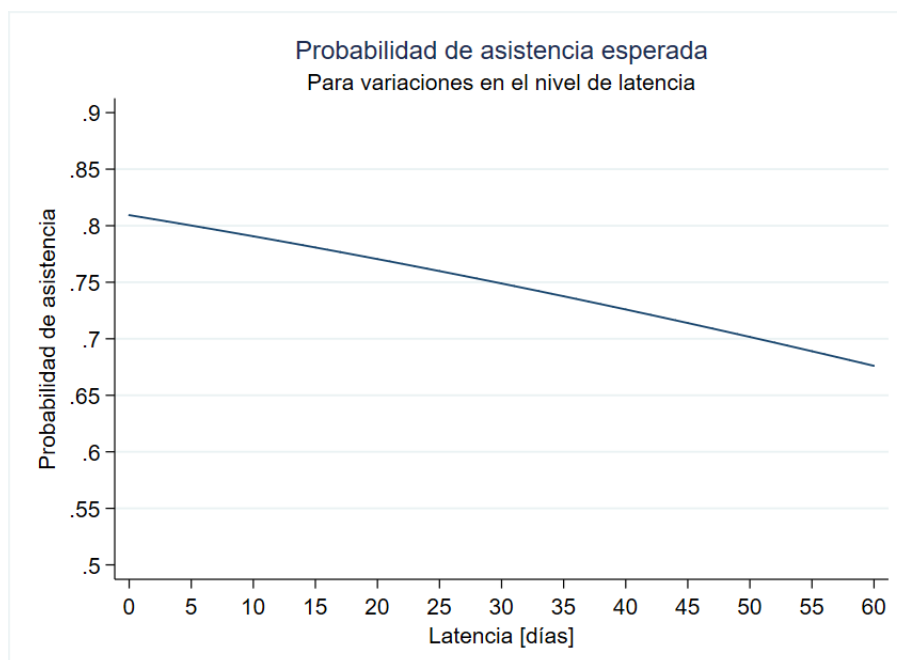


Figura 5.9: Probabilidad esperada bajo variaciones en el nivel de latencia de la cita médica (Fuente: Elaboración propia)

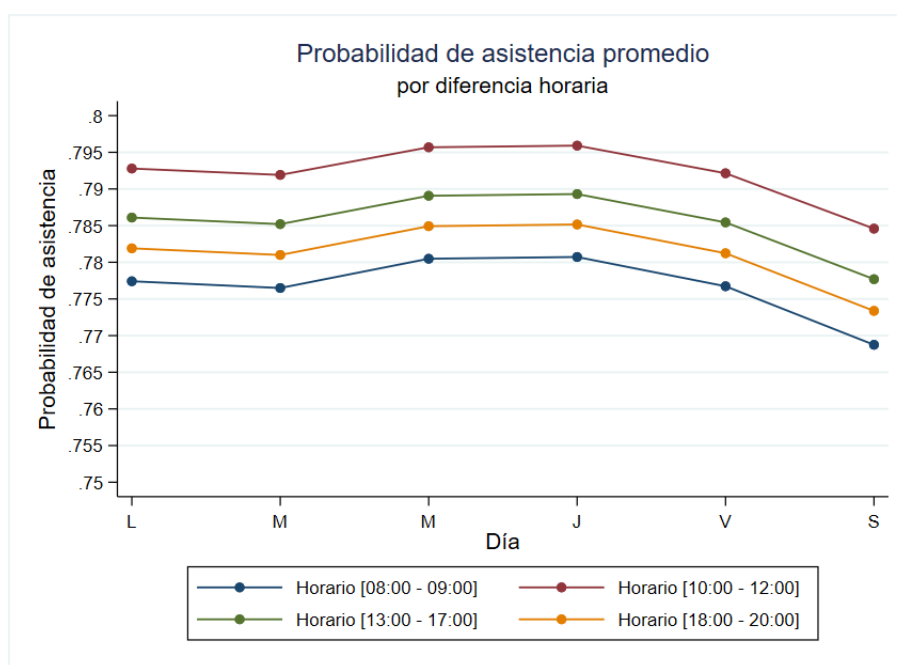


Figura 5.10: Probabilidad de asistencia promedio por horarios de atención (Fuente: Elaboración propia)

Validación de Hipótesis N°8 y N°9

Se valida la hipótesis N°8 ya que el día sábado es el día con menor probabilidad de asistencia, con una probabilidad promedio ajustada de 77.61 %. De igual forma, se rechaza la hipótesis

Nº9, debido a que el bloque horario con menor probabilidad de asistencia corresponde a [08:00 - 09:00], con una probabilidad promedio ajustada de 77.68 %.

Evaluación de la confiabilidad en las declaraciones del paciente La fiabilidad de la declaración del paciente al confirmar la cita trasciende eventualidades y da cuenta de un comportamiento persistente. Existen cuatro casos de variabilidad destacables:

- Pacientes que confirman y asisten a la cita médica
- Pacientes que confirman y no asisten a la cita médica
- Pacientes no confirman (sin contacto) y asisten a su cita médica
- Pacientes que no confirman (sin contacto) y no asisten a su cita médica

En el gráfico de la figura 5.11 se puede observar que la probabilidad de asistencia promedio para pacientes con un nivel de asistencia histórica de un 80 % sobrepasa el 80 % (Punto de corte). Es importante notar que los niveles extremos de asistencia (0 % y 100 %) solo son valores referenciales proyectados y no poseen mayor interpretación. De acuerdo a lo anterior, es posible segmentar tres comportamientos diferentes, para ello se utilizarán los niveles de asistencia referenciales 20 % y 80 %:

1. Para un paciente que presenta un nivel de asistencia de 20 % y una fiabilidad de declaración de 20 %, es un paciente que habitualmente es confirmado, pero no asiste a la mayoría de las citas. Posee aproximadamente una probabilidad promedio de asistencia de 65 %.
2. Para un paciente que presenta un nivel de asistencia de 20 % y una fiabilidad de 80 % es un paciente no es confirmado con frecuencia (sin gestión), sin embargo, las veces que ha sido confirmado, ha asistido a la cita médica. Posee probabilidad de asistencia de 73 %
3. Para un paciente con nivel de asistencia de 70 % la fiabilidad no tiene mayor incidencia, debido a que un paciente que ha asistido al 70 % de sus citas médicas debe poseer índice de fiabilidad alto y de no ser así, es por efectos de no confirmación (sin gestión).

Validación de Hipótesis Nº10

Se valida la hipótesis Nº10 ya que, si se deja constante el nivel de asistencia del paciente, existe variación en la probabilidad promedio de asistencia bajo distintos niveles de fiabilidad, diferencias que podrían sobrepasar un 10 % en diferencias de probabilidad esperada de asistencia. Lo anterior es muy relevante para el modelo, ya que frente a variaciones marginales de fiabilidad (comportamiento histórico), genera variaciones significativas en su probabilidad de asistencia estimada.

Evaluación efectos de continuidad de la cita Los niveles de fidelidad o acompañamiento podrían otorgar un impacto sobre las probabilidades de asistencia. A partir de lo anterior, se evalúa la probabilidad de asistencia para las siguientes ocasiones particulares:

1. Pacientes que asistieron a su última cita médica, a la misma especialidad con distinto profesional. La probabilidad de asistencia esperada corresponde a 79.96 % con un nivel de confianza de 90 %.

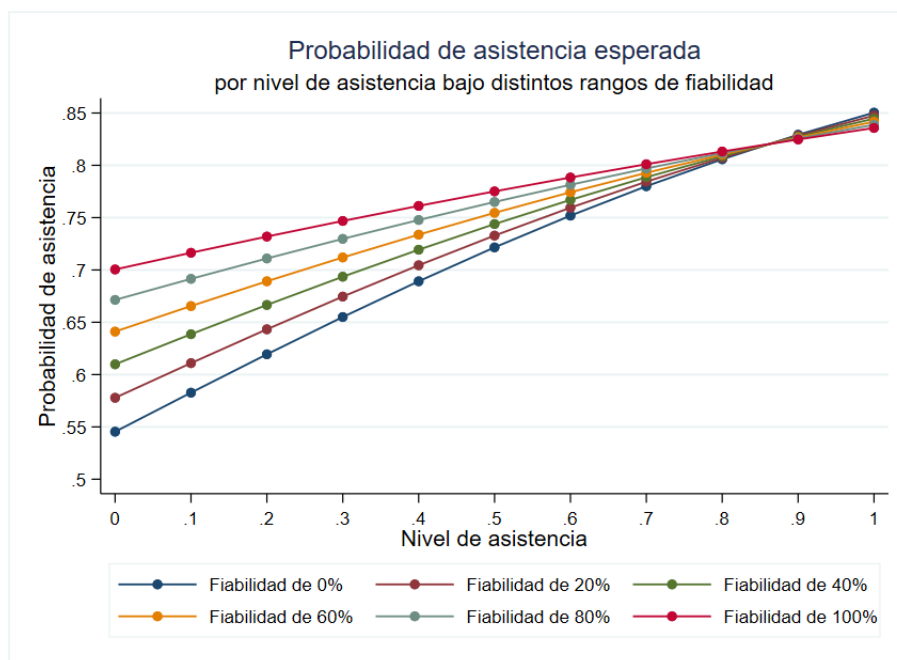


Figura 5.11: Probabilidad de asistencia promedio por nivel de fiabilidad de declaraciones del paciente y nivel de asistencia histórico
(Fuente: Elaboración propia)

2. Pacientes que asistieron a su última cita médica, a la misma especialidad con el mismo profesional. La probabilidad de asistencia esperada corresponde a 84.2 % con un nivel de confianza de 95 %.
3. Pacientes que asistieron a su última cita médica, a una especialidad diferente. La probabilidad de asistencia esperada corresponde a 78.5 % con un nivel de confianza de 95 %.
4. Pacientes que no asistieron a su última cita médica y esta última no tenía relación con la solicitud de especialidad actual. La probabilidad de asistencia esperada corresponde a 77.8 % con un nivel de confianza de 95 %.

Se puede observar que la variación marginal entre probabilidades esperadas para asistencia de la cita cuya cita anterior varia de asistida y no asistida tiene una diferencia aproximada de 1 %, lo mismo ocurre para citas asistidas que fueron agendadas para la misma especialidad en comparación a una cita asistida con agendamiento de una especialidad distinta. Se concluye que la especialidad de la cita anterior (asistida o no) no afecta en gran medida el comportamiento de asistencia de la cita actual.

Al comparar la probabilidad promedio de asistencia para una cita cuya cita anterior no fue asistida en comparación a una cita cuya cita anterior fue agendada con el mismo profesional y asistida, se obtiene una diferencia marginal de probabilidad esperada de aproximadamente 4 %, de esta forma se concluye que existe un efecto importante en la probabilidad de asistencia promedio generada por la fidelidad del paciente con su médico de tratamiento.

Validación de Hipótesis N°11

Dado el análisis anterior se verifica la hipótesis N°11, la cual señala que existen tendencias a asistir en mayor medida a citas médicas con el mismo profesional de la última cita a la cual asistió. Se demuestra además que la diferencia en probabilidad promedio asciende a aproximadamente 4 %

Evaluar distancia y tiempo de viaje Esta es una de las mayores interrogantes a resolver, ya que la distancia utilizada corresponde a la medida por vías permitidas en dos puntos y el tiempo utilizado corresponde al tiempo de viaje por transporte público. Se obtuvo que la probabilidad marginal promedio para una variación de 50 km corresponde a 1 %, es decir, la distancia entre el hogar del paciente y el centro de atención afecta positivamente la probabilidad de asistencia (a mayor distancia, mayor probabilidad de asistir a la cita médica). Sin embargo, la variación es bastante pequeña considerando la cantidad de kilómetros que representa.

El efecto del tiempo de tránsito por transporte público resultó ser una variable significativa a un 90 % de confianza, al igual que la distancia, el tiempo afecta positivamente la probabilidad de asistencia del paciente. La probabilidad marginal promedio al variar una unidad (estimada como 100 min), aumenta la probabilidad de asistencia en aproximadamente 1 %. Si bien dichas variables no se encuentran correlacionadas, mantienen un efecto similar sobre la variación en la probabilidad.

Validación de Hipótesis N°12 y N°13

Dado el análisis anterior se rechaza la hipótesis N°12, la cual señala que la distancia de viaje y la probabilidad de asistencia promedio son inversamente proporcionales. Sin embargo, es importante notar que la variación sobre la probabilidad es muy baja, llegando a 1 % cada 50 km, lo cual es casi imperceptible.

De igual forma, se rechaza la hipótesis N°13 para la variable tiempo, se concluye que no es significativa. Sin embargo, pese a que se acepte el nivel de confianza al 90 %, la variación marginal de la probabilidad esperada es cercana a 1 % por cada unidad de tiempo, reflejada esta última como 100 minutos, por lo que al igual que la variable distancia, su efecto marginal sobre la probabilidad marginal de asistencia promedio es casi imperceptible.

Variaciones de temperatura y precipitaciones La variable temperatura no es significativa al 95 % de confianza por lo que no se puede interpretar, sin embargo, la variable precipitaciones es significativa al 95 % de confianza. Se puede observar que para precipitaciones débiles (menores a 2mm diarios), la probabilidad de asistencia esperada es 78.4 %. Para precipitaciones moderada-baja (2mm a 15 mm), la probabilidad de asistencia esperada es 78.24 %. Para precipitaciones moderada-alta (15mm a 30 mm), la probabilidad esperada de asistencia asciende a 77.42 % mientras que para precipitaciones intensas (>30mm), la probabilidad promedio de asistencia corresponde a 66.37 % (ver figura 5.12).

Validación de Hipótesis N°14

Dado el análisis anterior, no se puede validar ni rechazar la hipótesis asociada a variación de

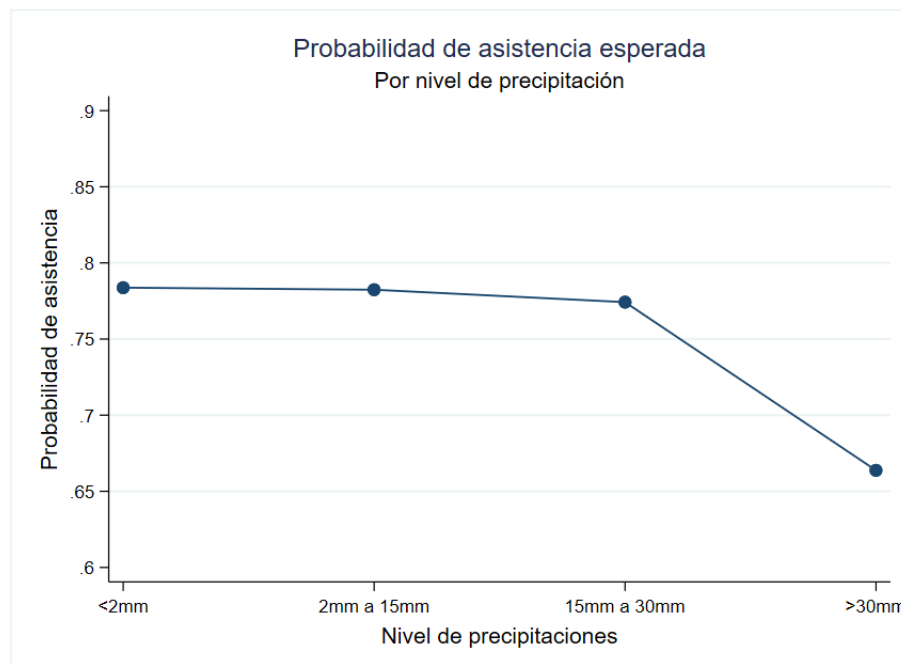


Figura 5.12: Probabilidad de asistencia esperada por nivel de precipitación
(Fuente: Elaboración propia)

probabilidad marginal sujeta a variaciones de temperatura debido a problemas de significancia de la variable.

En cuanto a la hipótesis al nivel de precipitaciones, existe una variación en la probabilidad de asistencia por lo que la hipótesis es validada de forma parcial. Se destaca además que la probabilidad esperada para precipitaciones intensas corresponde a 66.37 %, es decir, existe una reducción de probabilidad promedio de 12 % con respecto a la probabilidad esperada a precipitaciones débiles.

Efectos de manifestaciones sociales El efecto de manifestaciones sociales en Santiago sobre el nivel de asistencia del paciente es estadísticamente significativo a un nivel de confianza de 95 %. Sin embargo, la probabilidad esperada corresponde a 78.7 % para ausencia de manifestaciones en Santiago y un 78.3 % para días donde se generen manifestaciones. Es importante destacar que la asignación de marchas queda sujeta a horarios establecidos por lo que si bien es cierto la covariable resulta significativa, ciertos efectos causado por las asistencias a citas médicas en horarios pre y post manifestación, pueden reducir la capacidad explicativa del factor estimado.

Validación de Hipótesis N°15

Se valida la hipótesis de efectos de manifestaciones sobre la probabilidad estimada de asistencia. Se destaca que la variación de probabilidad promedio es 0.4 % lo que la hace un factor secundario a la hora de predecir el comportamiento del paciente.

Efectos de eventos deportivos a nivel nacional Se puede observar que la probabilidad esperada de asistencia para un día de evento deportivo es 78.4 %, mientras que para un día sin evento deportivo asciende a 78.7 %. De esta forma, se establece que existe una variación marginal sobre

la probabilidad estimada de asistencia de aproximadamente 0.3 % debido a un evento deportivo a nivel nacional. Es importante destacar que al igual que en manifestaciones sociales, no existe la precisión para detectar el inicio y término del evento durante el día (sensación de termino con el fin de normalizar el comportamiento del paciente), de esta forma se estructura la probabilidad con asistencias pre y/o post evento.

Validación de Hipótesis N°16

Se valida la hipótesis de efectos de eventos deportivos a nivel nacional sobre la probabilidad de asistencia del paciente. La variación marginal de probabilidad promedio corresponde a 0.3 % sobre la ocurrencia del evento.

Efectos de citas médicas post operatorias Se observa una variación marginal en la probabilidad promedio estimada de aproximadamente 6 % frente a una variación en la categoría. De esta forma una cita médica asociada a post operatorio posee una probabilidad estimada de 84.6 %, mientras que una consulta médica sin dicha especificación posee una probabilidad estimada de 78.6 %. La interpretación del aumento en la probabilidad de asistencia tiene que ver con la necesidad del paciente de mantener en evaluación su condición.

Validación de Hipótesis N°17

Se valida la hipótesis de efectos positivos sobre la asistencia del paciente para categorías asociadas a consultas post operatorias. La variación marginal de la probabilidad esperada asciende a 6 %, de esta forma se justifica su contribución debido a la importancia que involucra el tratamiento.

5.3.2. Análisis del modelo con foco predictivo

El modelo estructural con foco predictivo, tiene por objetivo principal ser eficaz en la predicción de asistencia del paciente, además debe ser simple y de costo computacional reducido. Dada las características anteriores, se propone utilizar el Modelo III, con función de enlace LOGIT por los siguientes motivos:

- El modelo III maneja información de comportamiento histórico, característica que lo convierte en un modelo dinámico.
- La función de enlace LOGIT no posee una formula cerrada, a diferencia del LOGIT MIXTO y PROBIT (modelos integrados), por lo que el modelo sería fácilmente interpretable.
- El Modelo III con función de enlace LOGIT posee un 84.3 % de las observaciones correctamente clasificada, al igual que el modelo PROBIT y MIXED LOGIT (ver tabla 5.4).
- El costo computacional asociado al modelo LOGIT es menor y no requiere de mayor capacidad de memoria para mantener datos activos (forma funcional sencilla).

Dada la propuesta anterior, para asegurar la validez del modelo, se decide mejorar su ajuste mediante la utilización de la totalidad de la data disponible, esto es el set de datos compuestos por 1.5 millones de observaciones. El ajuste del modelo estructural para la extensión de data junto con sus Odds Ratios respectivos se encuentra en el *Apéndice A.3. Ajuste del Modelo III mediante función de enlace LOGIT con extensión de data.*

Se puede observar que pese a agregar mayor data, el modelo clasifica el 84.44 % de las observaciones correctamente, el nivel de ajuste no manifiesta una variación significativa, lo cual indica que el modelo mejora su poder de predicción (ver tabla 5.5).

<i>Descripción</i>	<i>Logit Modelo III (muestra)</i>	<i>Logit Modelo III (expansión de data)</i>
Observaciones	300000	1585098
LL	-106955.510	-567249.086
McFadden's R^2	0.323	0.322
McFadden's R^2_{adjust}	0.321	0.321
A.U. ROC Curve	0.8597	0.8577
Correctly Classified (Cutoff 0.5)	84.32 %	84.44 %

Tabla 5.5: Resumen de ajuste del Modelo III con función de enlace LOGIT para la muestra y extensión de data

(Fuente: Elaboración propia)

Evaluación de la aplicación de un modelo predictivo en la gestión de sobrecupos

En la tabla 5.6 se pueden observar los parámetros de validación de la clasificación de asistencia e inasistencia para el modelo predictivo escogido y ajustado sobre la expansión de data. Para analizar la prueba de validación de forma intrínseca, se recurre al análisis de sensibilidad y especificidad. Se puede observar que el modelo predice en un 95.65 % a los pacientes asistentes que eventualmente asistirían a su cita médica (sensitividad), mientras que mantiene un 44.84 % de especificidad, lo que se traduce en su capacidad para excluir a pacientes inasistentes. Es aquí la importancia de la aplicación del Modelo III ajustado en políticas de sobrecupo, ya que permite predecir asistencia de pacientes por cota superior, por ejemplo, es capaz de predecir con un alto nivel de confianza a pacientes asistentes, dicho de otro modo, es altamente probable que no asistan más pacientes de los que se han previsto, de esta forma al generar políticas de sobrecupo, se estaría asegurando con bastante confianza una cota superior de pacientes a atender y por ende el beneficio se traduce en atención adicional para cubrir las inasistencias con una cota superior de pacientes que controlan la capacidad de atención del profesional para la agenda asignada. De forma complementaria se puede observar la tasa de falsos negativos, la cual representa a los pacientes que asistirían definitivamente a la cita médica, pero fueron predichos como inasistentes. Se puede observar que la probabilidad de error (que asistan más pacientes que los estimados) es bastante baja, asegurando el control de la capacidad de la agenda.

Validación de Hipótesis N°18

Dada la información anterior, se valida la hipótesis y se concluye que es posible predecir el comportamiento de un paciente a partir de un modelo estructural. Se demuestra además que, si se aplica de forma empírica a la gestión de sobrecupos, su utilización asegura el control sobre la capacidad de la agenda.

<i>Descripción</i>	<i>Logit Modelo III (expansión de data)</i>
Sensitividad	95.65 %
Especificidad	44.84 %
Tasa de falsos negativo	4.35 %
Tasa de falso positivo	55.16 %
Correctly Classified (Cutoff 0.5)	84.44 %

Tabla 5.6: Resumen de clasificación de predicciones: sensibilidad y especificidad
(Fuente: Elaboración propia)

Estimación de resultados al integrar el modelo predictivo en el proceso

El siguiente apartado tiene como propósito estimar los resultados al aplicar el modelo predictivo en la gestión de citas médicas. Para ello se utilizan los siguientes supuestos:

- Se establece que aproximadamente 400.000 citas médicas son asistidas anualmente [7].
- No se considera la sensibilidad del modelo con el fin de evaluar un modelo con mayores dificultades de predicción (realista).
- El costo de una consulta médica ambulatoria se generaliza de acuerdo a la especialidad medicina general, la cual asciende a \$33.520 pesos [8].
- No se considera la diferencia de probabilidad de asistencia entre citas asociadas a cupos y citas extras por concepto de predicción, las que eventualmente debieran agendarse como sobrecupos y, por ende, tener mayor probabilidad de asistencia.
- Se considera que el modelo es aplicado a todas las citas médicas gestionadas (todas las especialidades).

Dado que el 84.44 % de los valores fueron correctamente clasificados bajo un Cutoff de 0.5, existe un error de estimación de 15.56 %. Si se asume que el error de estimación corresponde a errores de falso positivo (es el peor caso, ya que se espera que asistan más personas de las que realmente asistirán), entonces para un año promedio, en el cual asisten 400.000 pacientes a sus citas médicas (78.3 %), bajo el supuesto anterior, el modelo predeciría una asistencia de $400000(1 + 0,1556) = 462240$ pacientes asistentes. Además, el total de pacientes agendados (excluye bloqueos de agenda) corresponde a 510.856 (dada la estimación de pacientes asistentes por año), entonces, de acuerdo al modelo existen 48.616 cupos disponibles (debido a que los pacientes no asistirán a su cita médica). Si se reagendan dichos cupos y que, al igual que el grupo anterior, el 78.3 % de los pacientes asista a su cita médica, existe un grupo de 38.066 pacientes que asistirán de forma adicional. Así los ingresos aumentarán en 9.5 % de forma anual por concepto de predicción de cita y aplicación de cita extra. En términos monetarios, considerando la el costo de la cita médica a \$33.520 pesos, el ingreso adicional anual correspondería a 1276 millones de pesos. De igual forma, se puede interpretar en términos de reducción en la tasa de inasistencias, la cual pasa de 21.7 % a 14.25 %. Lo anterior es una clara demostración que el modelo podría causar un gran impacto en el ingreso si es utilizado con políticas de gestión adecuadas.

Revisión bibliográfica asociada a modelos predictivos y su relación con políticas de sobrecupo

LaGanga & Lawrence (2007) [20] muestra que el sobrecupo puede mejorar de forma significativa el performance clínico, mejorando el acceso de pacientes a salud de calidad y mejorando la productividad de las instituciones. Señala que el exceso de reservas proporciona mayor utilidad cuando las clínicas atienden a un mayor número de pacientes, las tasas de inasistencia son más altas y la variabilidad del servicio es menor. Destaca que incluso en instituciones con tiempos de servicio altamente variables, se obtendrán resultados netos positivos mediante el sobrecupo. Berg (2013) [3] concluye que el sobrecupo puede ayudar a mitigar el impacto de las inasistencias en la ganancia neta esperada de una institución y tiene un costo de implementación más bajo que otras estrategias de intervención. Zacharias & Pinedo (2014) [23] estudia modelo de sobrecupo para programar las llegadas a un centro médico bajo una conducta de inasistencia, con pacientes con diferentes probabilidades de inasistencia y diferentes ponderaciones. Los resultados fueron positivos en la performance donde la tasa de inasistencias y la heterogeneidad de los pacientes tienen un impacto significativo en el programa óptimo. Daggy (2010) [6] señala en su estudio que al utilizar el modelo predictivo para aumentar el número de asistencias, se obtuvo un 13 % más de utilización de médicos y un 50 % menos de horas extraordinarias formales en la clínica. Señala además que en general el número de pacientes atendidos aumento en 12 %.

En cuanto a los factores a tener en cuenta a la hora de establecer políticas de sobrecupo asociado a modelos predictivos, se destacan:

- LaGanga & Lawrences (2007) [20] señala que uno de los factores a tener en cuenta a la hora de la toma de decisiones es el aumento de carga de trabajo del médico (Horas extraordinarias).
- Daggy (2010) [6] señala que la mala implementación de un sistema predictivo de citas médicas podría incrementar los tiempos de espera de los pacientes y eventualmente aumentar las siguientes inasistencias
- Berg (2013) [3] demuestra que los tiempos de espera presentan altas correlaciones con las inasistencias.

Hoy en día, existe una empresa estadounidense llamada PINNACLE SOLUTIONS INC., empresa de software que ha incluido un sistema llamado PINNACLE PATIENT NO-SHOWS PREDICTORTM, software que gestiona la data del paciente e incorpora información de tipo endógena y exógena para predecir el comportamiento de asistencia del paciente. No señala el tipo de modelo predictivo que utiliza.

5.3.3. Discusión

Información adicional y adaptación tecnológica

Si bien el modelo se ajusta correctamente a la información, es posible mejorarlo mediante el testeo de nuevas variables que eventualmente afecten en el comportamiento del paciente, por ejemplo, morbilidad del paciente (enfermedades latentes, requerimientos de atención continua), características socioculturales (religión, nivel de educación, etnia, núcleo familiar), características asociadas a la atención (reclamos, niveles de espera, bloqueos de agenda no informados), entre otras. De igual forma, es importante capturar información acerca de las posibles causas reales de las inasistencias, así evaluar nuevas vías de análisis asociada al comportamiento del paciente y su realidad. En cuanto a la adaptación tecnológica, se recomienda crear nuevas variables en la base de datos asociadas

al comportamiento histórico, de esta forma el costo de procesamiento por consultas (queries) por paciente se reduciría de forma significativa.

Continuidad e integralidad de la información

La evaluación acerca del decaimiento del poder predictivo del modelo es fundamental para la continuidad de su eventual utilización. A medida que transcurra el tiempo, claramente el modelo se verá afectado debido a los cambios de comportamiento del paciente o a la modificación de políticas asociadas a la institución (cambiaran levemente los factores ajustados). Dado lo anterior, el modelo se deberá ajustar cada cierto periodo de tiempo para mantener su poder predictivo (debido a que las variables que afectan el comportamiento ya se encuentran definidas, bastaría sólo con ajustar el modelo utilizando data más reciente, lo cual se puede realizar de forma sencilla en un lapso de tiempo reducido). Para mantener información robusta y veraz se recomienda asegurar políticas asociada a la integralidad de información ingresada o modificada en el sistema (mantener información actualizada).

Factores a tener en cuenta en su implementación

De acuerdo al estudio bibliográfico anterior, se puede dar cuenta de la necesidad de establecer políticas de gestión adecuadas al momento de la implementación de un modelo predictivo en la gestión de data. Es así como establecer las capacidades de agenda de forma adecuada es un criterio fundamental para mantener el beneficio potencial del modelo. Con respecto a la utilización de las horas médicas, es importante mantener la integralidad del tiempo asociado a la consulta, ya que este factor podría ser un riesgo potencial si no se establecen las políticas adecuadas para la gestión de sobrecupos. Finalmente, un último aspecto a tener en cuenta son los bloqueos de agenda. Se deberá crear un sistema que mitigue las solicitudes de bloqueo por parte del profesional médico ya que, bajo el supuesto de implementación del modelo, el número de pacientes a comunicar el bloqueo de su cita será mayor, lo que ocasionaría eventualmente sobrecarga de trabajo y una visión de institución desfavorable.

Futuras mejoras en el ajuste

Una variable que eventualmente podría mejorar el ajuste del modelo corresponde al porcentaje de asistencia histórico. Si bien actualmente dicha variable se incluye en el ajuste, a medida que se almacene mayor cantidad de información, es importante asignar mayor nivel de importancia a comportamientos recientes en comparación al comportamiento pasado. Si el objetivo es reajustar el modelo involucrando mayor cantidad de información por paciente se recomienda asignar algún tipo de suavización exponencial temporal a la variable señalada, con el fin de adaptar mejor su resultado frente a cambios en el comportamiento en el corto plazo (ponderar de cierta forma la información, tal de asignar mayor importancia a cambios en el comportamiento reciente).

Capítulo 6

Aplicación del modelo estructural sobre políticas de gestión

La importancia de la información extraída del modelo estructural con foco explicativo subyace en la capacidad para entender de forma teórica las variaciones marginales que genera una variable de estudio sobre la probabilidad de asistencia del paciente. Una de las aplicaciones más potentes del foco explicativo corresponde a la capacidad de evaluación de políticas al comparar eventos sujetos a variaciones de parámetros (contrafactuales). Mediante el análisis de la data desarrollada en el presente estudio, es posible obtener impactos asociados a variaciones teóricas sujeto a contextos, por ejemplo:

- Conocer el impacto que generaría la no confirmación de citas médicas producto de problemas técnicos asociados a llamadas outbound. Entender este problema, ayudaría a generar políticas preventivas.
- Conocer la variación en la asistencia del paciente si las citas médicas que fueron agendadas vía web sean confirmadas a través de emails. Mediante el análisis de las citas web, edad de pacientes y canales de comunicación se puede estimar el impacto de dicha política y por ende implementar políticas asociadas a la heterogeneidad de los pacientes.
- Entender el impacto que se genera en la probabilidad de asistencia si se decide establecer un programa de acompañamiento del paciente, sistema que mejora la fidelización de pacientes a la institución. Esta política podría ser evaluada por medio de contrafactuales, tal como los problemas anteriores.

Finalmente, a modo de abordar un problema latente en el departamento, se mostrará como el ajuste de un modelo estructural con foco explicativo otorga las herramientas para evaluar políticas de gestión.

6.1. Evaluación de política de extensión de agenda

La extensión de agenda es una política abordada por el departamento. Su desarrollo responde a la necesidad por capturar mayor cantidad de pacientes y así asegurar agendas completas. Es importante notar que la motivación nace a partir del costo de oportunidad, es decir, responde a la

interrogante: ¿Cómo evitar que el paciente i llame a otra institución debido a que la agenda para la especialidad solicitada se encuentra copada por j pacientes en una extensión de tiempo t ? A partir de ello se ha generado un interés por aumentar el tiempo t a nt de tal forma que se pueda capturar la solicitud.

El problema se abordará por medio de la maximización de valor esperado. Se puede observar en la figura 6.1 que el diagrama (1) corresponde al modelo actual, donde para un evento de oportunidad de agendamiento T se solicita un número determinado de citas para distintas latencias l . Se recuerda al lector que se define latencia a la diferencia de tiempo entre la fecha de atención y la fecha de solicitud t . El modelo (2) de la misma figura corresponde a la variación de la extensión de la agenda, donde crece en razón de n , de esta forma la nueva agenda quedaría disponible mediante nt tiempo.

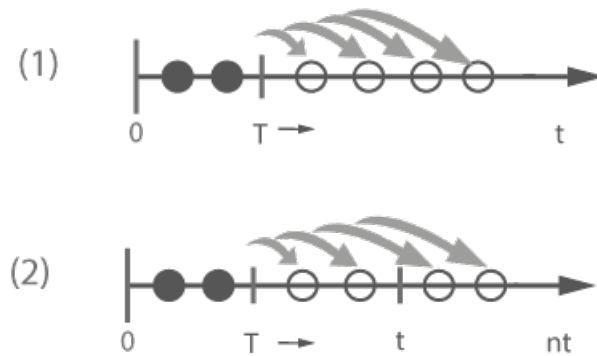


Figura 6.1: Descripción del proceso de extensión de citas
(Fuente: Elaboración propia)

Dado que el conjunto de pacientes agendados conforman una cantidad k para una latencia l en un periodo T , se define: k_{lT} . Junto con ello se debe definir para cada k_{lT} una probabilidad de asistencia P_{lT} . Se debe notar que la probabilidad varía de acuerdo a la latencia y al periodo de selección, dado que no se puede aceptar una cita más allá del límite de tiempo otorgado por la agenda. Para expresar de forma visual la asignación de variables se estructura una matriz triangular descrita en la figura 6.2.

$$\begin{pmatrix} P_{00}k_{00} & P_{01}k_{01} & \dots & P_{0,t-1}k_{0,t-1} & P_{0t}k_{0t} \\ P_{10}k_{10} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & P_{1,t-1}k_{1,t-1} & 0 \\ \dots & P_{t-1,1}k_{t-1,1} & \dots & 0 & 0 \\ P_{t0}k_{t0} & \dots & 0 & 0 & 0 \end{pmatrix} \quad (6.1)$$

6.1: Matriz de pacientes asistentes
(Fuente: Elaboración propia)

Sea C_i el costo de la especialidad i , donde se especifica una reducción del conjunto de especialidades debido a la extensión de la agenda (restricción de especialidades que no pudiesen eventualmente extender sus agendas al periodo nt) se define la desigualdad:

$$\sum_{i \in ESP} \sum_{l=0}^t \sum_{T=0}^{t-l} P_{lT} K_{lT} C_i \leq \sum_{i,j \in ESP} \sum_{j \leq i} \sum_{l=0}^{nt} \sum_{T=0}^{nt-l} P_{lT} K_{lT} C_j \quad \forall K_{0t} + K_{1,t-1} \leq R_i \quad (6.2)$$

6.2: Desigualdad asociada a la hipótesis de investigación

(Fuente: Elaboración propia)

Donde R_i es la restricción de límite de pacientes por agenda (se asume una agenda por especialidad). De acuerdo a lo anterior y a lo propuesto en la política, la diferencia de utilidad entre dos periodos debiera ser positiva.

Dado que la estructura es demasiado compleja para ser testeada mediante análisis teórico, se especifica un conjunto de holguras para resolver el problema:

1. Sea el costo C_i y C_j constantes (asumiendo que todos los pacientes escogieron la misma especialidad para distintos niveles de latencia en distintos periodos).
2. Se establece una probabilidad promedio P_l dependiente sólo del nivel de latencia de la cita (excluye variabilidades del modelo estructural ajustado en el capítulo 4).
3. Se agrega holgura al cupo de la agenda, dado que luego se simplificará la forma de capturar los cupos de acuerdo a una latencia dada.
4. Dado que se trabajan con dimensiones de t y nt respectivamente (ver matriz triangular), se estima que las probabilidades poseen dimensiones similares (probabilidad para cada periodo).

De esta forma, la matriz se descompone en una matriz de probabilidades promedios como muestra la ecuación 6.3

$$\begin{pmatrix} \hat{P}_0 & 0 & \dots & 0 & 0 \\ 0 & \hat{P}_1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & \hat{P}_{t-1} & 0 \\ 0 & 0 & \dots & 0 & \hat{P}_t \end{pmatrix} \begin{pmatrix} k_{00} & k_{01} & \dots & k_{0,t-1} & k_{0t} \\ k_{10} & \dots & \dots & k_{1,t-1} & 0 \\ \dots & \dots & \dots & 0 & 0 \\ \dots & k_{t-1,1} & 0 & 0 & 0 \\ k_{t0} & 0 & 0 & 0 & 0 \end{pmatrix} \quad (6.3)$$

6.3: Descomposición de estructura

(Fuente: Elaboración propia)

Hasta el momento, el modelo se estructura como muestra la ecuación 6.4, donde se puede observar que se poseen todos los parámetros para extraer la utilidad del modelo:

- \hat{P}_l corresponde a la matriz diagonal de probabilidades asociadas a cada nivel de latencia l , la diagonal de la matriz fue obtenido en el capítulo 4 mediante la extracción de la probabilidad marginal promedio de los pacientes para cada latencia l . En dicho capítulo se explica cómo obtener una aproximación de probabilidades para distintos valores de latencia sin evaluar

el modelo estructural (dado que se ha ajustado mediante heterogeneidad aleatoria de los pacientes (Mixed Logit), cada estimación debe ponderarse por dicha variabilidad, de esta forma, para facilitar los cálculos se recomienda estimar la aproximación por efecto marginal sobre la probabilidad).

- La matriz triangular superior asociada a la cantidad de citas agendadas, cada día, para cada nivel de latencia, se puede simular mediante un testeo.

$$\sum_{l=0}^t \sum_{T=0}^{t-l} P_l K_{lT} \leq \sum_{l=0}^{nt} \sum_{T=0}^{nt-l} P_l K_{lT} \quad (6.4)$$

6.4: Reducción del modelo bajo supuestos de holgura
(Fuente: Elaboración propia)

Para evitar un testeo inicial de las cantidades K_{lT} , se pueden ajustar mediante una distribución predefinida. Si no se conoce la distribución de las cantidades, y dado que lo que se quiere abordar es la probabilidad total del evento se puede simplificar accediendo a sólo la diagonal de la matriz triangular, teóricamente sería iniciar regresivamente desde la cantidad de citas solicitadas en el periodo final e ir seleccionando la cantidad correspondiente para periodos inferiores, de esta forma se acotaría la agenda (ya que se procede de forma regresiva). Finalmente, la ecuación queda expresada en la ecuación 6.5.

$$\sum_{l=0}^t P_t K_{0,l-t} \leq \sum_{l=0}^{nt} P_t K_{0,l-nt} \quad (6.5)$$

6.5: Reducción del modelo
(Fuente: Elaboración propia)

A partir de este punto el análisis, se formula por casos:

1. Si la agenda actual bajo el periodo t no se completa totalmente, entonces eventualmente $K_{0,l-nt} > K_{0,l-t}$, por lo que es necesaria una estimación de la cantidad para analizar cómo la probabilidad asociada a la asistencia producida por un aumento en la latencia es compensada por el aumento de utilización de la agenda.
2. Para el caso en el cual las agendas médicas se encuentran semejantemente completas, es decir $K_{0,l-nt} = K_{0,l-t} + \varepsilon$ con $\varepsilon \rightarrow 0$, entonces las cantidades no podrán compensar el efecto de la latencia sobre la probabilidad, de esta forma se compara la probabilidad marginal para cada n periodos, así evaluar el impacto de reducción de la probabilidad sobre el exceso de latencia en la agenda.

En la tabla 6.1 se puede observar que, si las agendas se realizan cada dos semanas, la reducción sobre la probabilidad de asistencia es 1.51 % para la extensión de las primeras dos semanas. A partir de 30 días, el aumento de dos semanas adicionales se duplica al anterior (3.13 %), produciendo un perjuicio creciente, dependiendo de la latencia de la agenda inicial y su periodo de extensión. Si se considera 1 mes como base, la reducción en la probabilidad de asistencia promedio al extender en un mes la agenda genera un impacto de 3.3 % en la

probabilidad de asistencia. El análisis de las probabilidades puras es el mejor caso, ya que, si la cantidad de citas por agenda son desiguales, el factor de reducción en la probabilidad debe ser compensado por el aumento en la utilización de las agendas (completar agenda), si esto no ocurre, el efecto asociado a la reducción en probabilidad, más una agenda precaria, causaría costos de gestión que sobrepasan el costo de oportunidad y por ende la política sería un completo fracaso.

<i>n</i>	<i>Periodo</i>	<i>Prob. promedio n periodos iguales</i>	<i>Prob. promedio extensión a n periodos</i>
15 días	15 días	79.53 %	79.53 %
	30 días	79.53 %	78.02 %
	45 días	79.53 %	76.40 %
	60 días	79.53 %	74.68 %
30 días	1 mes	78.02 %	78.02 %
	2 meses	78.02 %	74.68 %

Tabla 6.1: Variación de la probabilidad de asistencia frente a variación marginal de la latencia (Fuente: Elaboración propia)

Validación de Hipótesis de aplicación del modelo: foco explicativo N°1

Se rechaza la hipótesis debido a que la política depende del estado actual de la cita (distribución de latencias, cupos por agenda, nivel de utilización) y la extensión del tiempo de actividad. Si se utiliza el modelo anterior, se puede observar que el modelo se convierte en una política exitosa si cumple las siguientes condiciones:

- Si el cupo de la agenda es robusto, es decir, no tiene problemas para coparse, la probabilidad asociada a la extensión de agenda causará una reducción de pacientes asistentes por lo que se deberán establecer necesidades justificadas junto con políticas adicionales asociadas al aumento de las citas médicas. Una situación justificable puede ser una agenda que se copa fácilmente, sin embargo, las gestiones de apertura ocasionan costos adicionales para el hospital. Dado que la cantidad de citas es similar a una velocidad de utilización promedio, la política sería una mejora en la gestión.
- Si el cupo de la agenda es precario y su velocidad de ocupación es lenta, extender la agenda con débil ocupación de la misma realizaría un perjuicio al proceso, dado que reduciría la probabilidad promedio de asistencia y no habría un aumento de cupo suficiente para lograr una mejora. Es así como la reducción a una agenda robusta (alta velocidad y capacidad de ocupación) adecuada es una motivación clara para estudiar la latencia en las políticas de acción.

En conclusión, el aumento de la latencia en agendas medicas es un perjuicio asegurado, que puede ser cubierto por un aumento en la utilización y políticas de gestión adecuadas, bajo una necesidad justificada. Para obtener mayor precisión acerca del efecto asociado al contrafactual, es necesario simular las ponderaciones de la matriz triangular superior sujeto a una capacidad de agendamiento y adicionar el costo asociado a la cita (costo de gestión), de esta forma estructurar una utilidad esperada. La asunción del vector de probabilidades promedio, no afectara en

gran medida el cálculo real, por lo que se podría seguir utilizando. Si el objetivo es el contrario, es decir reducir periodo de agendamiento se debe tener precaución del costo asociado a la apertura de agendas. Fuera de ello, es más factible y justificable convertir agendas extensas con baja utilización y velocidad de llenado a agendas promedios robustas.

6.2. Evaluación de aplicación del modelo con foco predictivo sobre la gestión de citas médicas en Call Center

De acuerdo a lo establecido en el capítulo 4, el modelo predictivo es una solución factible y viable. Se inicia el estudio enunciando problemáticas asociadas a un departamento o institución, sin embargo, la potencialidad de establecer un modelo mediante un foco predictivo y explicativo es extraer información cuantificable a partir de la estimación de variables para predecir un comportamiento de un suceso futuro y adquirir una visión general de la variación en el comportamiento frente a variaciones ficticias de variables relacionadas, todo lo anterior mediante un ajuste adecuado (realista), donde se agregan componentes aleatorias para complementar las variaciones no observables entre grupos.

A partir de los resultados, el modelo prevé un 84.4 % del comportamiento correctamente. Dado que el modelo predice en un 95.65 % a los pacientes que eventualmente asistirán a la cita (sensitividad), sería adecuado establecer medidas asociadas al sobrecupo, ya que con el nivel de predicción de los asistentes, sólo existe un 4.4 % de probabilidad de asista un paciente adicional (no esperado), lo que se traduce en un controlador natural de capacidad de agenda. Es así como se recomienda fuertemente establecer la cantidad de sobrecupos como un seguro asociado al aumento en la asistencia. Dado lo anterior, si se establece un límite claro de sobrecupos frente al nivel de predicción del modelo se podría reducir efectivamente el número de inasistencias.

6.2.1. Aspectos a tener en cuenta en la implementación

1. En primer lugar, establecer políticas asociadas a la calidad del servicio. En el capítulo 4 se enuncia la correlación entre las inasistencias y el nivel de espera del paciente, de esta forma el modelo podría predecir correctamente, sin embargo, el aumento en los tiempos de espera podría ser contraproducente.
2. Se debe establecer el nivel de sobrecupo ajustado al tiempo de apertura de agenda. Se demostró en la sección anterior que podría eventualmente haber ineficiencias a la hora de extender el periodo de agendamiento para agendas con alta velocidad de llenado y agendas con baja velocidad.
3. Se deben corregir políticas como confirmaciones ilimitadas, debido a que existen malas prácticas utilizadas por los agentes debido a brechas de acción como aumento de sueldo variable mediante llamados reiterativos a números telefónicos inexistentes. Una alternativa de solución simple y rápida se deriva del estudio del modelo estructural con foco explicativo. Donde se estima una probabilidad de asistencia bajo variaciones en la gestión de confirmación cercana a 82.72 % frente a una cita no confirmada con probabilidad de 76.63 %. Además, mediante análisis exploratorio se obtuvo que el 93 % de las confirmaciones se realizan realizando hasta 4 intentos, de esta forma se valida la confirmación como gestión y se propone establecer una

cota de confirmación de 4 intentos por paciente. Es importante mencionar que esta cota no es una solución relacionada a las malas prácticas de los operadores, sin embargo, valida el problema y lo soluciona de forma rápida y eficiente.

Validación de Hipótesis N°2 del modelo

Se valida la hipótesis de malas prácticas en el sistema de gestión, las cuales fueron causadas por generalizar la confirmación como un aumento de probabilidad de asistencia proporcional al número de intentos, es así como se corrige gracias a la información extraída del modelo estructural con foco explicativo, entregando una correcta información acerca del efecto de la confirmación sobre la probabilidad de asistencia y corrigiendo la política de confirmaciones ilimitadas.

4. Si el modelo es aplicado de forma adecuada, las inasistencias a citas médicas se suplirán por sobrecupos, lo que eventualmente aumentaría la utilización de los médicos, es decir, si se reducen las inasistencias a citas médicas (objetivo del modelo estructural con foco predictivo), se reducirán los tiempos no utilizados por médicos en horas institucionales (contractuales).
5. Finalmente, para las horas no institucionales, la reducción de inasistencias asociadas al modelo se traducirá eventualmente en aumentos en el sueldo variable de los médicos para este tipo de agenda (esta implicancia es directa dado que la utilidad del médico para horas no institucionales corresponde a un porcentaje establecido de la cita pagada). Un efecto adicional de la aplicación del modelo se asocia a la motivación por proponer agendas no institucionales, aumentando eventualmente la oferta médica y por consiguiente el porcentaje de retención de profesionales en la institución (dado que una de las problemáticas colaterales asociadas a las ausencias era la elección de los profesionales a no presentar oferta médica en la institución).

Validación de Hipótesis N°3 del modelo

Finalmente, dada la información bibliográfica y la composición del departamento de Call Center, en conjunto con el análisis de ajuste y probabilidad de predicción realizado en el capítulo 4, se valida la hipótesis.

Conclusión

La importancia de analizar el foco explicativo del modelo estructural subyace en la necesidad de abordar contrafactuales, es decir, estudiar mediante mecanismos u herramientas posibilidades latentes en el comportamiento de una variable predicha bajo variaciones de otra variable implicante. La profundidad interpretativa que arroja el modelo va más allá de estudios descriptivos, ya que entrega la probabilidad de ocurrencia frente a variaciones del medio. Esta última información es fundamental a la hora de evaluar cambios en los procesos o políticas aplicadas, de esta forma, lo anterior otorga al tomador de decisiones un abanico de posibilidades de acción frente a eventualidades del medio, sin necesidad de testear el sistema para orientar la solución.

La principal política abordada en el estudio corresponde a la extensión de agenda médica. A partir del análisis del sistema, probabilidades promedio y probabilidades marginales, se concluye que al extender las agendas se produce un perjuicio en la tasa de asistencia de los pacientes, debido a que la probabilidad marginal de latencia por cada día adicional disminuye en promedio en -0.2% . De la misma forma, se establecen los efectos marginales para distintas variables testeadas, endógenas y exógenas, con el fin de evaluar futuras políticas de gestión respaldadas de forma numérica y sostenibles en el tiempo.

En cuanto al ajuste del modelo con foco predictivo, se obtuvo un modelo robusto, con 84.44% de predicciones correctamente clasificadas bajo un cutoff de 0.5 y una sensibilidad de 95.65% , lo que se traduce en un muy buen predictor de pacientes asistentes. Los factores que generaron mayor nivel explicativo en el modelo corresponden a factores de comportamiento y fiabilidad. Si se considera un comportamiento similar al histórico, bajo niveles de asistencia promedio, se estima que al aplicar el modelo anterior en la gestión de citas médicas se obtendrá un aumento de 9.5% en los ingresos del hospital por concepto de utilización de citas no asistentes, cuyo número de citas recuperadas asciende a aproximadamente 38.000 citas anuales. En términos de tasas, la tasa de inasistencia se reduciría de 21.7% a 14.5% , cumpliéndose por completo el objetivo general del presente estudio.

Bibliografía

- [1] El mercurio [en línea]. <https://www.emol.com>. [consulta: 18 de Marzo de 2019].
- [2] Smith B., Leimkuhler J., and Darrow R. *Yield management at american airlines*. Dallas /Fort Worth Airport/Texas 75261-9616, 1992.
- [3] Berg B.P., Murr M., Chermak D., Woodall J., Pignone M., Sandler R.S., and Denton B.T. Estimating the cost of no-shows and evaluating the effects of mitigation strategies. *Medical decision making : an international journal of the Society for Medical Decision Making*, 33(8), 11 2013.
- [4] Banco Central. Base de datos estadísticos [en línea]. <https://si3.bcentral.cl/Siete/secure/cuadros/home.aspx>. [consulta: 18 de Marzo de 2019].
- [5] Dirección General De Aeronáutica Civil. Dirección meteorológica de Chile - servicios climáticos [en línea]. <https://climatologia.meteochile.gob.cl/application/index/productos/RE7003>. [consulta: 01 de Septiembre de 2018].
- [6] Joanne Daggy, Mark Lawley, Deanna Willis, Debra Thayer, Christopher Suelzer, Po-Ching DeLaurentis, Ayten Turkcan, Santanu Chakraborty, and Laura Sands. Using no-show modeling to improve clinic performance. *Health informatics journal*, 16:246–59, 12 2010.
- [7] Universidad de Chile. Memoria de la Universidad de Chile 2014 - 2018 [en línea]. <http://www.uchile.cl/publicaciones/143433/memoria-de-la-universidad-de-chile-2014-2018>. [consulta: 17 de Junio de 2019].
- [8] Hospital Clínico de la Universidad de Chile. Arancel hcuch año 2019 [en línea]. <https://www.redclinica.cl/informacion-al-paciente/aranceles-hcuch.aspx>. [consulta: 01 de Junio de 2019].
- [9] Hospital Clínico de la Universidad de Chile. Especialidades médicas [en línea]. <https://www.redclinica.cl/plantilla/especialidades.aspx>. [consulta: 18 de Marzo de 2019].
- [10] Superintendencia de salud. Cobertura Adicional para Enfermedades Catastróficas (CAEC) [en línea]. <http://www.supersalud.gob.cl/difusion/665/w3-propertyvalue-1964.html>. [consulta: 20 de Marzo de 2019].

- [11] Superintendencia de salud. Garantías Explícitas en Salud (GES) [en línea]. <http://www.supersalud.gob.cl/difusion/665/w3-propertyvalue-1962.html>. [consulta: 20 de Marzo de 2019].
- [12] Woodcock E. *Mastering Patient Flow: More Ideas to Increase Efficiency and Earnings*. MG-MA, 2003.
- [13] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 1996.
- [14] Feriados.cl. Feriados de carácter general [en línea]. <https://www.feriados.cl/>. [consulta: 18 de Marzo de 2019].
- [15] M Goic. Apunte de curso Marketing II, 2015.
- [16] Google. Google Cloud [en línea]. <https://cloud.google.com>. [consulta: 18 de Marzo de 2019].
- [17] Nigro, Héctor, Xodo, Daniel, Corti, Gabriel, and Terren, Damián. *KDD (Knowledge Discovery in Databases): Un proceso centrado en el usuario*. 2004. In VI Workshop de Investigadores en Ciencias de la Computación.
- [18] Joseph M. Hilbe. *Logistic Regression Models*. 1. ed edition, 2009. CRC Press.
- [19] Wolfgang Langer. The assessment of fit in the class of logistic regression models: A pathway out of the jungle of pseudo- r^2 s, 2000.
- [20] Linda R. LaGanga Stephen R. Lawrence. Clinic overbooking to improve patient access and increase provider productivity. *Decision Sciences. A journal of the Decision Sciences Institute*, 38, 05 2007.
- [21] Szumilas, Magdalena. Explaining odds ratios. 2010.
- [22] Ortega C., Manuel and Cayuela D. ,Aurelio. Regresión logística no condicionada y tamaño de muestra: una revisión bibliográfica. 2002.
- [23] Christos Zacharias Michael Pinedo. Appointment scheduling with no-shows and overbooking. *POMS-Productions and Operations Management*, 04 2013.
- [24] Solofutbol.cl. Partidos seleccion chilena de fútbol [en línea]. <http://www.solofutbol.cl/seleccion%20chilena/partidos/partidos%20seleccion%20chilena.htm>. [consulta: 18 de Marzo de 2019].
- [25] Kenneth E. Train. *Métodos de elección discreta con simulación*. 2. ed edition, 2009.
- [26] Matthew Turland. *phplarchitect's Guide to Web Scraping with PHP*. MarcoTabini, 28 Bombay Ave. Toronto, ONM3H 1B7 Canada, 1. ed edition, 2010.
- [27] Fernández Alarcón, V. *Relaciones encontradas entre las dimensiones de las estructuras organizativas y los componentes del constructo capacidad de absorción: El caso de empresas*

ubicadas en el territorio español. 2004. Tesi doctoral, UPC, Departament d'Organització d'Empreses.

Apéndice A

Anexo metodológico

A.1. Modelos de elección discreta y maximización de utilidad aleatoria

Los modelos de elección discreta [25] se obtienen habitualmente bajo el supuesto de que el decisor se comporta de forma que maximiza la utilidad que percibe, a esto se le denomina modelos de utilidad aleatoria (Random Utility Models, RUMs). El modelo de utilidad aleatoria se obtiene de la siguiente forma: Sea un decisor n se enfrenta a una elección entre j alternativas en el tiempo t , donde cada alternativa reflejara una utilidad al decisor llamada U_{njt} , dado que el decisor es racional, escoge la alternativa con mayor utilidad: Escoge i si y sólo si $U_{nit} > U_{njt} \forall i \neq j$. Como es natural el investigador no observa la utilidad del decisor, sólo posee atributos de las alternativas que maneja el decisor denotado como x_{njt} y algunos atributos del decisor denotados por s_{njt} , de esta forma se confirma la utilidad representativa denotada por $V_{njt} = V(x_{njt}, s_{nt})$. Dado que existen aspectos que el investigador no puede observar (no existe conocimiento absoluto), es claro que $U_{njt} \neq V_{ijt}$, de esta forma se puede descomponer la utilidad mediante $U_{njt} = V_{njt} + \varepsilon_{njt}$ donde ε_{njt} captura factores no observables de cada decisión. Como la variable ε_{ijt} es desconocida, es tratada como una variable aleatoria para cada decisión j .

Sea el vector aleatorio $\varepsilon'_{nt} = \langle \varepsilon_{n1t}, \dots, \varepsilon_{njt} \rangle$, se denota como $f(\varepsilon_{nt})$ a la densidad de probabilidad conjunta del vector aleatorio ε'_{nt} . Con esta densidad, el investigador puede hacer afirmaciones probabilísticas acerca de la elección del decisor. La probabilidad de que el decisor n elija la alternativa i en t es:

$$\begin{aligned} P_{nit} &= \text{Prob}(U_{nit} > U_{njt} \forall i \neq j) \\ &= \text{Prob}(V_{nit} + \varepsilon_{nit} > V_{njt} + \varepsilon_{njt} \quad \forall i \neq j) \\ &= \text{Prob}(\varepsilon_{njt} - \varepsilon_{nit} < V_{nit} - V_{njt} \quad \forall i \neq j) \end{aligned} \tag{A.1}$$

A.1: Probabilidad de elección del decisor

La probabilidad de A.1 es una distribución acumulada. Si se integra sobre la densidad $f(\varepsilon_{nt})$, la probabilidad acumulada se reescribe como se muestra en A.2, donde $I(\cdot)$ es la función indicatriz, igual a 1 cuando la expresión dentro del paréntesis es verdadera y 0 en caso contrario.

$$\begin{aligned}
P_{nit} &= Prob(\varepsilon_{njt} - \varepsilon_{nit} < V_{nit} - V_{njt} \quad \forall i \neq j) \\
&= \int_{\varepsilon} I(\varepsilon_{njt} - \varepsilon_{nit} < V_{nit} - V_{njt} \quad \forall i \neq j) f(\varepsilon_{nit}) d\varepsilon_{nit}
\end{aligned} \tag{A.2}$$

A.2: Probabilidad de elección del decisor utilizando la función de densidad de probabilidad de la componente no observada

Dependiendo de la especificación de $f(\cdot)$, la integral tendrá una forma abierta (cálculo de la integral mediante métodos numéricos) o cerrada (se reduce a una expresión simple). La expresión con función de enlace Logit posee una expresión cerrada debido a que se asume que la parte no observada de la utilidad se distribuye de acuerdo a una distribución de tipo valor extremo independiente e idénticamente distribuida (iid). La función de enlace Probit posee una expresión abierta dado que se asume que $f(\cdot)$ es una normal multivariada. Mientras que la función de enlace Mixed Logit posee una forma abierta debido a que se basa en la asunción de que la parte no observada de la utilidad se compone por una componente que distribuye de acuerdo a la función de distribución especificada por el investigador y más una componente que es de tipo valor extremo iid.

Apéndice B

Ajuste del modelo estructural: Foco explicativo

B.1. Ajuste de coeficientes

A continuación, se presentan los coeficientes de cada modelo testado:

- **Modelo I:** Corresponde a la especificación de utilidad para un modelo estructural de efectos fijos sin historial de comportamiento.
- **Modelo II:** Corresponde a la especificación de utilidad para un modelo estructural de efectos fijos con historial de comportamiento asociado a 1 rezago (cita anterior).
- **Modelo III:** Corresponde a la especificación de utilidad para un modelo estructural de efectos fijos con historial de comportamiento asociado a 1 y 2 rezagos junto al nivel de asistencia y fiabilidad.
- **Modelo IV:** Corresponde a la especificación de utilidad para un modelo estructural de efectos fijos y aleatorios sobre cada paciente.

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
1.Cardiología/Cardiología	-0.743*** (0.0428)	-0.572*** (0.0420)	-0.430*** (0.0423)	-0.244*** (0.0235)	-0.541*** (0.0452)
2.Obesidad y nutrición	-0.917*** (0.0461)	-0.781*** (0.0461)	-0.740*** (0.0463)	-0.428*** (0.0260)	-0.863*** (0.0497)
3.Cirugía	-1.422*** (0.0364)	-1.041*** (0.0362)	-0.897*** (0.0371)	-0.521*** (0.0207)	-1.155*** (0.0395)
4.Dermatología	0.275*** (0.0756)	0.101 (0.0689)	-0.0335 (0.0654)	-0.0190 (0.0348)	-0.00156 (0.0706)
5.Diabetes	-0.294***	-0.171*	-0.139	-0.0815	-0.146

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
	(0.100)	(0.102)	(0.101)	(0.0541)	(0.105)
6.Dolor	0.357***	0.529***	0.454***	0.229***	0.487***
	(0.110)	(0.114)	(0.112)	(0.0577)	(0.123)
7.Endocrinología	0.0102	0.0401	0.0182	0.0102	0.113
	(0.0722)	(0.0716)	(0.0697)	(0.0376)	(0.0717)
8.Fonoaudiología	-1.147***	-1.106***	-1.192***	-0.694***	-1.163***
	(0.0834)	(0.0842)	(0.0845)	(0.0462)	(0.0906)
9.Gastroenterología	-0.0975**	-0.0502	-0.0532	-0.0284	-0.0474
	(0.0451)	(0.0450)	(0.0455)	(0.0251)	(0.0483)
10.Genética	-0.991***	-0.819***	-0.702**	-0.394**	-0.743**
	(0.273)	(0.298)	(0.287)	(0.164)	(0.303)
11.Geriatría	-0.115	-0.0686	-0.00633	-0.0109	0.222
	(0.137)	(0.137)	(0.141)	(0.0747)	(0.146)
12.Hematología	0.189*	0.190*	-0.0384	-0.0495	0.0418
	(0.109)	(0.111)	(0.107)	(0.0553)	(0.112)
13.Infectología	-0.0930*	-0.0998*	-0.00213	0.000529	-0.00842
	(0.0558)	(0.0574)	(0.0597)	(0.0333)	(0.0637)
14.Inmunología y alergias	-0.387***	-0.263***	-0.227***	-0.119***	-0.191***
	(0.0676)	(0.0653)	(0.0632)	(0.0354)	(0.0683)
15.Medicina familiar	-0.286	-0.238	-0.206	-0.114	-0.136
	(0.499)	(0.561)	(0.591)	(0.329)	(0.696)
16.Medicina física y rehabilitación	0.479***	0.274***	0.134***	0.0979***	0.260***
	(0.0586)	(0.0529)	(0.0490)	(0.0262)	(0.0509)
18.Medicina interna	-0.0564	-0.0333	-0.0318	-0.0221	-0.0254
	(0.0560)	(0.0571)	(0.0581)	(0.0319)	(0.0618)
21.Nefrología	-0.556***	-0.408***	-0.339***	-0.197***	-0.403***
	(0.0458)	(0.0464)	(0.0470)	(0.0262)	(0.0502)
22.Neonatología/Pediatría	-0.206***	-0.244***	-0.364***	-0.206***	-0.109***
	(0.0424)	(0.0414)	(0.0408)	(0.0226)	(0.0406)
23.Neurología/Neurocirugía	-0.220***	-0.252***	-0.309***	-0.173***	-0.284***
	(0.0849)	(0.0751)	(0.0704)	(0.0381)	(0.0823)
24.Obstetricia/Ginecología	-0.427***	-0.368***	-0.385***	-0.223***	-0.438***
	(0.0352)	(0.0352)	(0.0355)	(0.0197)	(0.0381)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
25.Oftalmología	-0.376*** (0.0390)	-0.234*** (0.0391)	-0.161*** (0.0396)	-0.0951*** (0.0218)	-0.224*** (0.0421)
26.Oncología	-0.904*** (0.0684)	-0.819*** (0.0693)	-0.996*** (0.0673)	-0.566*** (0.0369)	-1.236*** (0.0734)
27.Otorrinolaringología	-0.759*** (0.0398)	-0.640*** (0.0399)	-0.581*** (0.0405)	-0.335*** (0.0226)	-0.635*** (0.0433)
28.Psicología	-0.101* (0.0530)	-0.230*** (0.0524)	-0.363*** (0.0514)	-0.213*** (0.0285)	-0.241*** (0.0551)
29.Psiquiatría	0.145*** (0.0542)	0.0656 (0.0534)	-0.0275 (0.0522)	-0.0238 (0.0283)	-0.0254 (0.0554)
30.Respiratorio	-2.165*** (0.0690)	-1.735*** (0.0708)	-1.404*** (0.0759)	-0.818*** (0.0439)	-1.624*** (0.0800)
31.Reumatología	0.106* (0.0616)	0.107* (0.0603)	-0.00487 (0.0597)	-0.0124 (0.0321)	-0.00244 (0.0634)
32.Traumatología	0.0983* (0.0510)	0.166*** (0.0512)	0.169*** (0.0522)	0.0798*** (0.0279)	0.105* (0.0558)
33.Urología	-0.127** (0.0511)	-0.0837* (0.0503)	-0.100** (0.0502)	-0.0633** (0.0271)	-0.106** (0.0537)
34.Dentomaxilofacial	-2.761*** (0.0391)	-2.323*** (0.0386)	-1.786*** (0.0396)	-1.043*** (0.0222)	-2.127*** (0.0429)
35.Laboratorio	1.680*** (0.0974)	1.466*** (0.0919)	1.056*** (0.0862)	0.527*** (0.0397)	1.231*** (0.0904)
36.Investigación clínica	-3.538*** (0.459)	-3.378*** (0.438)	-3.387*** (0.431)	-1.985*** (0.257)	-3.740*** (0.461)
37.Anestesiología	0.166 (0.419)	0.239 (0.428)	0.266 (0.451)	0.140 (0.250)	0.296 (0.481)
38.Banco de sangre	0.971 (1.038)	1.159 (1.043)	1.104 (1.050)	0.493 (0.471)	1.020 (1.081)
39.Neumología	-0.971*** (0.0994)	-0.838*** (0.102)	-0.771*** (0.103)	-0.442*** (0.0570)	-0.850*** (0.110)
40.Transplante	0.359* (0.205)	0.291 (0.208)	0.0523 (0.206)	-0.0160 (0.103)	-0.0284 (0.216)
41.Otra	-0.356***	-0.346***	-0.418***	-0.250***	-0.474***

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
	(0.0698)	(0.0701)	(0.0711)	(0.0382)	(0.0763)
42.Radiología	-0.927***	-0.652*	-0.614*	-0.386**	-0.762**
	(0.302)	(0.354)	(0.357)	(0.192)	(0.383)
6.HCUCH	1.018***	0.781***	0.603***	0.351***	0.755***
	(0.0175)	(0.0166)	(0.0157)	(0.00874)	(0.0168)
1.sobrecupo	0.467***	0.873***	0.938***	0.539***	0.684***
	(0.0199)	(0.0250)	(0.0272)	(0.0149)	(0.0230)
1.lag_sobrecupo		-0.276***	-0.280***	-0.169***	
		(0.0215)	(0.0227)	(0.0125)	
1.sobrecupo#1.lag_sobrecupo		-0.717***	-0.732***	-0.395***	
		(0.0431)	(0.0474)	(0.0264)	
1.web	-0.137***	-0.0291	-0.0575**	-0.0361***	-0.123***
	(0.0189)	(0.0240)	(0.0240)	(0.0134)	(0.0184)
1.lag_web		0.0764***	0.0749***	0.0401***	
		(0.0263)	(0.0260)	(0.0144)	
1.web#1.lag_web		-0.240***	-0.186***	-0.100***	
		(0.0377)	(0.0380)	(0.0212)	
1.confirm	0.764***	0.745***	0.748***	0.429***	0.812***
	(0.0145)	(0.0145)	(0.0146)	(0.00792)	(0.0154)
1.fono_fijo	1.834***	1.643***	1.484***	0.768***	0.0902***
	(0.480)	(0.427)	(0.520)	(0.259)	(0.0139)
1.fono_celu	0.784***	0.670***	0.502**	0.288**	-0.0541
	(0.249)	(0.245)	(0.231)	(0.126)	(0.0658)
1.fono_fijo#1.fono_celu	-1.766***	-1.603***	-1.435***	-0.738***	
	(0.482)	(0.428)	(0.521)	(0.260)	
1.fono_trab	0.921***	0.815***	0.804***	0.452***	-0.00810
	(0.320)	(0.312)	(0.310)	(0.166)	(0.0142)
1.fono_fijo#1.fono_trab	-2.002***	-1.808***	-1.761***	-0.913***	
	(0.532)	(0.480)	(0.569)	(0.285)	
1.fono_celu#1.fono_trab	-0.955***	-0.850***	-0.827***	-0.465***	
	(0.322)	(0.314)	(0.311)	(0.166)	
1.fono_fijo#1.fono_celu#1.fono_trab	1.988***	1.821***	1.789***	0.928***	
	(0.535)	(0.482)	(0.570)	(0.286)	

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
1.email	0.725** (0.355)	0.729** (0.348)	0.597* (0.363)	0.360* (0.196)	0.0181 (0.0139)
1.fono_fijo#1.email	-1.915** (0.816)	-1.985*** (0.759)	-2.250*** (0.773)	-1.206*** (0.418)	
1.fono_celu#1.email	-0.572 (0.356)	-0.626* (0.349)	-0.540 (0.364)	-0.327* (0.196)	
1.fono_fijo#1.fono_celu#1.email	1.876** (0.817)	1.968*** (0.761)	2.217*** (0.774)	1.188*** (0.419)	
1.fono_trab#1.email	-0.544 (0.459)	-0.579 (0.446)	-0.701 (0.469)	-0.436* (0.251)	
1.fono_fijo#1.fono_trab#1.email	1.751** (0.887)	1.815** (0.829)	2.328*** (0.846)	1.265*** (0.456)	
1.fono_celu#1.fono_trab#1.email	0.498 (0.462)	0.547 (0.449)	0.663 (0.471)	0.414 (0.252)	
1.fono_fijo#1.fono_celu #1.fono_trab#1.email	-1.635* (0.890)	-1.728** (0.831)	-2.265*** (0.848)	-1.230*** (0.457)	
1.marcha	-0.0510*** (0.0185)	-0.0502** (0.0197)	-0.0516** (0.0207)	-0.0281** (0.0114)	-0.0456** (0.0213)
2.month	0.741** (0.335)	0.582* (0.350)	0.255 (0.371)	0.131 (0.205)	0.272 (0.385)
3.month	1.214*** (0.285)	0.843*** (0.299)	0.516 (0.315)	0.289* (0.174)	0.473 (0.326)
4.month	0.656** (0.255)	0.442* (0.268)	0.244 (0.284)	0.143 (0.157)	0.252 (0.293)
5.month	1.096*** (0.236)	0.887*** (0.248)	0.624** (0.263)	0.351** (0.146)	0.676** (0.272)
6.month	0.863*** (0.228)	0.569** (0.239)	0.267 (0.253)	0.151 (0.140)	0.313 (0.262)
7.month	0.831*** (0.230)	0.636*** (0.241)	0.375 (0.256)	0.216 (0.142)	0.386 (0.265)
8.month	1.035*** (0.235)	0.734*** (0.246)	0.415 (0.260)	0.234 (0.144)	0.475* (0.269)
9.month	0.835***	0.609**	0.320	0.179	0.326

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
	(0.233)	(0.245)	(0.259)	(0.144)	(0.268)
10.month	1.176***	0.942***	0.643**	0.350**	0.646**
	(0.237)	(0.249)	(0.264)	(0.146)	(0.273)
11.month	1.016***	0.849***	0.702**	0.398**	0.779***
	(0.256)	(0.269)	(0.284)	(0.157)	(0.293)
12.month	0.261	0.164	-0.0637	-0.0254	-0.0679
	(0.302)	(0.319)	(0.339)	(0.187)	(0.349)
temp	0.0428***	0.0314***	0.0194*	0.0110*	0.0214*
	(0.00972)	(0.0102)	(0.0108)	(0.00598)	(0.0112)
2.month#c.temp	-0.0368**	-0.0303*	-0.0159	-0.00845	-0.0173
	(0.0154)	(0.0161)	(0.0171)	(0.00945)	(0.0177)
3.month#c.temp	-0.0524***	-0.0345**	-0.0203	-0.0113	-0.0175
	(0.0138)	(0.0145)	(0.0152)	(0.00838)	(0.0158)
4.month#c.temp	-0.0227*	-0.0145	-0.00678	-0.00420	-0.00704
	(0.0132)	(0.0139)	(0.0147)	(0.00812)	(0.0152)
5.month#c.temp	-0.0526***	-0.0449***	-0.0337**	-0.0188**	-0.0368**
	(0.0125)	(0.0132)	(0.0139)	(0.00770)	(0.0144)
6.month#c.temp	-0.0367***	-0.0205	-0.00516	-0.00276	-0.00776
	(0.0126)	(0.0132)	(0.0140)	(0.00774)	(0.0145)
7.month#c.temp	-0.0318**	-0.0283**	-0.0191	-0.0111	-0.0171
	(0.0133)	(0.0140)	(0.0148)	(0.00815)	(0.0153)
8.month#c.temp	-0.0490***	-0.0336**	-0.0179	-0.00979	-0.0215
	(0.0132)	(0.0138)	(0.0145)	(0.00801)	(0.0150)
9.month#c.temp	-0.0369***	-0.0284**	-0.0151	-0.00843	-0.0145
	(0.0119)	(0.0125)	(0.0132)	(0.00730)	(0.0136)
10.month#c.temp	-0.0561***	-0.0468***	-0.0333***	-0.0179**	-0.0328**
	(0.0116)	(0.0122)	(0.0129)	(0.00711)	(0.0134)
11.month#c.temp	-0.0478***	-0.0412***	-0.0359***	-0.0203***	-0.0395***
	(0.0123)	(0.0129)	(0.0136)	(0.00749)	(0.0140)
12.month#c.temp	-0.0159	-0.0116	-0.00113	-0.00118	-0.00143
	(0.0143)	(0.0151)	(0.0160)	(0.00886)	(0.0165)
2.dia_sem	-0.0879**	-0.106***	-0.108***	-0.0577**	-0.0116
	(0.0388)	(0.0401)	(0.0411)	(0.0224)	(0.0196)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
3.dia_sem	0.0529 (0.0384)	0.0511 (0.0401)	0.0700* (0.0419)	0.0412* (0.0228)	0.0388** (0.0197)
4.dia_sem	0.0213 (0.0407)	0.0325 (0.0416)	0.0515 (0.0428)	0.0309 (0.0234)	0.0418** (0.0199)
5.dia_sem	-0.00933 (0.0431)	0.0353 (0.0449)	0.0784* (0.0467)	0.0458* (0.0255)	-0.00856 (0.0210)
6.dia_sem	-0.336*** (0.0627)	-0.391*** (0.0646)	-0.387*** (0.0660)	-0.221*** (0.0370)	-0.107*** (0.0397)
2.hourg	0.107*** (0.0365)	0.104*** (0.0375)	0.129*** (0.0387)	0.0742*** (0.0212)	0.199*** (0.0180)
3.hourg	-0.00826 (0.0358)	0.0133 (0.0362)	0.0573 (0.0371)	0.0343* (0.0203)	0.111*** (0.0180)
4.hourg	0.0103 (0.0525)	0.0636 (0.0548)	0.116** (0.0565)	0.0643** (0.0313)	0.0569** (0.0282)
2.dia_sem#2.hourg	0.0328 (0.0488)	0.0796 (0.0508)	0.0946* (0.0524)	0.0483* (0.0287)	
2.dia_sem#3.hourg	0.132*** (0.0477)	0.158*** (0.0491)	0.165*** (0.0503)	0.0873*** (0.0275)	
2.dia_sem#4.hourg	-0.0334 (0.0700)	-0.0235 (0.0740)	-0.0128 (0.0776)	-0.00911 (0.0429)	
3.dia_sem#2.hourg	-0.0223 (0.0485)	0.0196 (0.0508)	0.0175 (0.0533)	0.00775 (0.0291)	
3.dia_sem#3.hourg	-0.0542 (0.0470)	-0.0507 (0.0490)	-0.0655 (0.0511)	-0.0374 (0.0280)	
3.dia_sem#4.hourg	-0.0396 (0.0688)	-0.0783 (0.0738)	-0.104 (0.0780)	-0.0566 (0.0432)	
4.dia_sem#2.hourg	0.0148 (0.0506)	0.0300 (0.0522)	0.0308 (0.0540)	0.0162 (0.0296)	
4.dia_sem#3.hourg	0.0134 (0.0494)	-0.00662 (0.0503)	-0.0315 (0.0518)	-0.0175 (0.0284)	
4.dia_sem#4.hourg	-0.0711 (0.0708)	-0.127* (0.0743)	-0.150* (0.0772)	-0.0830* (0.0429)	
5.dia_sem#2.hourg	-0.0633	-0.0626	-0.0743	-0.0418	

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
	(0.0525)	(0.0548)	(0.0572)	(0.0314)	
5.dia_sem#3.hourg	-0.0187	-0.0599	-0.0962*	-0.0555*	
	(0.0520)	(0.0540)	(0.0561)	(0.0308)	
5.dia_sem#4.hourg	-0.137*	-0.201**	-0.283***	-0.160***	
	(0.0754)	(0.0804)	(0.0853)	(0.0473)	
6.dia_sem#2.hourg	0.361***	0.395***	0.381***	0.215***	
	(0.0752)	(0.0792)	(0.0817)	(0.0457)	
6.dia_sem#3.hourg	0.599***	0.574***	0.569***	0.320***	
	(0.137)	(0.151)	(0.155)	(0.0858)	
2.dayg	-0.0737***	-0.0787***	-0.0863***	-0.0500***	-0.0954***
	(0.0132)	(0.0140)	(0.0147)	(0.00809)	(0.0152)
3.dayg	-0.0464***	-0.0486***	-0.0574***	-0.0349***	-0.0630***
	(0.0139)	(0.0146)	(0.0154)	(0.00846)	(0.0160)
1.lag1		0.564***	0.276***	0.180***	0.169***
		(0.0253)	(0.0387)	(0.0223)	(0.0237)
1.lag2			0.0380	0.0390*	
			(0.0354)	(0.0205)	
1.lag1#1.lag2			-0.294***	-0.186***	
			(0.0433)	(0.0250)	
1.lag3			0.333***	0.206***	
			(0.0314)	(0.0181)	
1.lag1#1.lag3			-0.131***	-0.0841***	
			(0.0423)	(0.0243)	
1.lag2#1.lag3			-0.158***	-0.0998***	
			(0.0425)	(0.0245)	
1.lag1#1.lag2#1.lag3			0.463***	0.262***	
			(0.0551)	(0.0316)	
1.lag1#c.tc		-	-	-	-
		0.000774***	0.000796***	0.000472***	0.000933***
		(7.77e-05)	(7.92e-05)	(4.41e-05)	(8.43e-05)
tc	0.000299***	0.000703***	0.000738***	0.000434***	0.000781***
	(4.00e-05)	(6.64e-05)	(6.78e-05)	(3.78e-05)	(7.16e-05)
d	0.00361***	0.00240***	0.00163***	0.000863***	0.000562**

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
	(0.000792)	(0.000696)	(0.000621)	(0.000335)	(0.000285)
d2	-3.78e-06***	-2.44e-06***	-1.61e-06**	-8.47e-07**	
	(9.24e-07)	(8.39e-07)	(7.10e-07)	(3.89e-07)	
t2	8.68e-09	7.90e-09	1.61e-08**	8.91e-09**	
	(7.39e-09)	(6.23e-09)	(8.09e-09)	(4.32e-09)	
1.lag_prof		-0.346***	-0.473***	-0.286***	
		(0.108)	(0.106)	(0.0628)	
1.lag1#1.lag_prof		0.453***	0.560***	0.349***	
		(0.122)	(0.122)	(0.0710)	
1.lag_espe		-0.200***	-0.185***	-0.108***	
		(0.0270)	(0.0280)	(0.0161)	
1.lag1#1.lag_espe		0.414***	0.326***	0.188***	
		(0.0325)	(0.0327)	(0.0185)	
1.lag_prof#1.lag_espe		-0.255**	-0.112	-0.0641	
		(0.111)	(0.110)	(0.0647)	
1.lag1#1.lag_prof#1.lag_espe		0.723***	0.445***	0.235***	0.759***
		(0.125)	(0.126)	(0.0730)	(0.0193)
1.lag_espe2			-0.205***	-0.119***	
			(0.0285)	(0.0163)	
1.lag2#1.lag_espe2			0.406***	0.233***	
			(0.0330)	(0.0186)	
1.lag_prof2			-0.780***	-0.463***	
			(0.113)	(0.0656)	
1.lag2#1.lag_prof2			1.041***	0.616***	
			(0.129)	(0.0739)	
1.lag_espe2#1.lag_prof2			0.197*	0.116*	
			(0.117)	(0.0680)	
1.lag2#1.lag_espe2#1.lag_prof2			0.181	0.0931	
			(0.134)	(0.0766)	
lagfiabilidad			1.672***	0.956***	1.314***
			(0.121)	(0.0686)	(0.125)
lagpor			2.163***	1.267***	3.027***

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
			(0.0427)	(0.0240)	(0.0377)
c.lagfiabilidad#c.lagpor			-2.109***	-1.188***	-1.528***
			(0.158)	(0.0876)	(0.164)
1.futbol	-0.0346*	-0.0434**	-0.0453**	-0.0254**	-0.0422**
	(0.0177)	(0.0187)	(0.0196)	(0.0108)	(0.0203)
1.preferiado	-0.0622**	-0.0597**	-0.0506*	-0.0271	-0.0515*
	(0.0271)	(0.0286)	(0.0298)	(0.0165)	(0.0308)
1.feriado	0.0364	0.0405	0.0328	0.0168	0.0370
	(0.0464)	(0.0482)	(0.0509)	(0.0280)	(0.0527)
1.postferiado	-0.0193	-0.0295	-0.0350	-0.0198	-0.0300
	(0.0233)	(0.0245)	(0.0259)	(0.0142)	(0.0267)
edad	-0.0281***	-0.0242***	-0.0146***	-	0.00746***
				0.00790***	
	(0.00173)	(0.00160)	(0.00148)	(0.000802)	(0.000436)
3.gender	-0.0528*	-0.0429	-0.0341	-0.0172	0.0326
	(0.0313)	(0.0292)	(0.0269)	(0.0147)	(0.0276)
3.gender#c.edad	0.00438***	0.00413***	0.00330***	0.00179***	0.00254***
	(0.000696)	(0.000647)	(0.000592)	(0.000320)	(0.000597)
1.email#c.edad	-0.00164**	-0.00121*	-0.000152	-8.34e-05	
	(0.000709)	(0.000657)	(0.000597)	(0.000324)	
edad2	0.000440***	0.000387***	0.000251***	0.000139***	
	(1.87e-05)	(1.72e-05)	(1.58e-05)	(8.51e-06)	
1.lluviag	-0.0245	-0.0443	-0.0579*	-0.0336*	-0.0499
	(0.0313)	(0.0330)	(0.0345)	(0.0191)	(0.0358)
2.lluviag	-0.0453*	-0.0543*	-0.0616**	-0.0364**	-0.0671**
	(0.0265)	(0.0280)	(0.0295)	(0.0162)	(0.0306)
3.lluviag	-0.153***	-0.159***	-0.141**	-0.0879**	-0.170**
	(0.0570)	(0.0611)	(0.0649)	(0.0359)	(0.0664)
4.lluviag	-1.053***	-1.179***	-1.279***	-0.728***	-1.359***
	(0.180)	(0.187)	(0.191)	(0.109)	(0.200)
1.gescaec	0.583***	0.616***	0.510***	0.265***	0.556***
	(0.100)	(0.100)	(0.101)	(0.0509)	(0.104)
1.post	0.960***	0.920***	0.819***	0.455***	0.876***

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
	(0.0869)	(0.0885)	(0.0879)	(0.0452)	(0.0946)
laten	-0.0220***	-0.0248***	-0.0264***	-0.0151***	-0.0255***
	(0.000549)	(0.000558)	(0.000574)	(0.000319)	(0.000577)
lag_laten	-	0.00647***	0.000582	0.000276	
		(0.000904)	(0.000926)	(0.000532)	
1.lag1#c.lag_laten		0.0222***	0.0122***	0.00683***	
		(0.00112)	(0.00112)	(0.000627)	
ipc	-0.114***	-0.116***	-0.0806***	-0.0412***	-0.0785***
	(0.0221)	(0.0224)	(0.0232)	(0.0128)	(0.0242)
antigüedad	3.66e-05***	1.92e-05*	6.12e-05***	3.49e-05***	8.48e-05***
	(1.15e-05)	(1.08e-05)	(1.11e-05)	(6.13e-06)	(1.18e-05)
t					8.10e-05*
					(4.72e-05)
0b.lag1#0b.lag_prof#1.lag_espe					-0.230***
					(0.0290)
0b.lag1#1.lag_prof#0b.lag_espe					-0.535***
					(0.113)
0b.lag1#1.lag_prof#1.lag_espe					-0.934***
					(0.0267)
1.lag1#0b.lag_prof#1.lag_espe					0.178***
					(0.0200)
1.lag1#1.lag_prof#0b.lag_espe					0.128*
					(0.0667)
Constant	-0.452	-0.555*	-1.784***	-1.049***	-1.916***
	(0.331)	(0.337)	(0.338)	(0.186)	(0.263)
Var(Constant)				0.43496	
				(0.01515)	
Observations	300,000	300,000	300,000	300,000	300,000
LR χ^2	24924	41598	55898	67874	44887
Prob $<\chi^2$	0	0	0	0	0
McFadden's R^2	0.197	0.268	0.323	0.328	
McFadden's R^2 Adjust	0.196	0.267	0.321	0.326	

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Probit Modelo III</i>	<i>Mixed Logit Modelo IV</i>
Mckelvey and Zavoina's R^2	0.317	0.408	0.489	0.512	
LL	-126831	-115573	-106956	-106175	-110140
Clusters	77141	77141	77141	77141	77141
Number of groups					77,141

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

B.2. Odds ratios para modelos con ajuste Logit y Mixed Logit (efectos fijos y aleatorios)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
1. Cardiología/Cardiología	0.476*** (0.0204)	0.565*** (0.0237)	0.650*** (0.0275)	0.582*** (0.0263)
2. Obesidad y nutrición	0.400*** (0.0184)	0.458*** (0.0211)	0.477*** (0.0221)	0.422*** (0.0210)
3. Cirugía	0.241*** (0.00878)	0.353*** (0.0128)	0.408*** (0.0151)	0.315*** (0.0125)
4. Dermatología	1.316*** (0.0995)	1.106 (0.0762)	0.967 (0.0633)	0.998 (0.0705)
5. Diabetes	0.745*** (0.0746)	0.843* (0.0863)	0.870 (0.0875)	0.864 (0.0903)
6. Dolor	1.429*** (0.158)	1.697*** (0.194)	1.575*** (0.177)	1.627*** (0.200)
7. Endocrinología	1.010 (0.0730)	1.041 (0.0745)	1.018 (0.0709)	1.120 (0.0803)
8. Fonoaudiología	0.318*** (0.0265)	0.331*** (0.0279)	0.304*** (0.0256)	0.313*** (0.0283)
9. Gastroenterología	0.907** (0.0409)	0.951 (0.0428)	0.948 (0.0431)	0.954 (0.0461)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
10.Genética	0.371*** (0.101)	0.441*** (0.131)	0.496** (0.142)	0.476** (0.144)
11.Geriatría	0.891 (0.122)	0.934 (0.128)	0.994 (0.140)	1.249 (0.182)
12.Hematología	1.209* (0.132)	1.209* (0.134)	0.962 (0.103)	1.043 (0.117)
13.Infectología	0.911* (0.0509)	0.905* (0.0519)	0.998 (0.0596)	0.992 (0.0632)
14.Inmunología y alergias	0.679*** (0.0459)	0.769*** (0.0502)	0.797*** (0.0504)	0.826*** (0.0565)
15.Medicina familiar	0.751 (0.375)	0.788 (0.442)	0.814 (0.481)	0.873 (0.608)
16.Medicina física y rehabilitación	1.614*** (0.0946)	1.316*** (0.0695)	1.143*** (0.0560)	1.297*** (0.0660)
18.Medicina interna	0.945 (0.0529)	0.967 (0.0552)	0.969 (0.0563)	0.975 (0.0602)
21.Nefrología	0.573*** (0.0262)	0.665*** (0.0308)	0.712*** (0.0335)	0.669*** (0.0336)
22.Neonatología/Pediatría	0.814*** (0.0345)	0.783*** (0.0324)	0.695*** (0.0284)	0.896*** (0.0364)
23.Neurología/Neurocirugía	0.802*** (0.0681)	0.777*** (0.0584)	0.734*** (0.0517)	0.753*** (0.0620)
24.Obstetricia/Ginecología	0.652*** (0.0230)	0.692*** (0.0244)	0.680*** (0.0241)	0.645*** (0.0246)
25.Oftalmología	0.686*** (0.0268)	0.791*** (0.0310)	0.851*** (0.0337)	0.799*** (0.0336)
26.Oncología	0.405*** (0.0277)	0.441*** (0.0305)	0.369*** (0.0249)	0.291*** (0.0213)
27.Otorrinolaringología	0.468*** (0.0186)	0.527*** (0.0210)	0.560*** (0.0227)	0.530*** (0.0229)
28.Psicología	0.904* (0.0479)	0.794*** (0.0416)	0.696*** (0.0358)	0.786*** (0.0433)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
29.Psiquiatría	1.157*** (0.0627)	1.068 (0.0570)	0.973 (0.0508)	0.975 (0.0540)
30.Respiratorio	0.115*** (0.00793)	0.176*** (0.0125)	0.245*** (0.0186)	0.197*** (0.0158)
31.Reumatología	1.112* (0.0685)	1.112* (0.0671)	0.995 (0.0594)	0.998 (0.0633)
32.Traumatología	1.103* (0.0562)	1.181*** (0.0605)	1.184*** (0.0618)	1.111* (0.0620)
33.Urología	0.881** (0.0450)	0.920* (0.0462)	0.905** (0.0454)	0.900** (0.0483)
34.Dentomaxilofacial	0.0632*** (0.00247)	0.0979*** (0.00378)	0.168*** (0.00664)	0.119*** (0.00512)
35.Laboratorio	5.368*** (0.523)	4.332*** (0.398)	2.874*** (0.248)	3.425*** (0.310)
36.Investigación clínica	0.0291*** (0.0133)	0.0341*** (0.0149)	0.0338*** (0.0146)	0.0238*** (0.0109)
37.Anestesiología	1.180 (0.494)	1.270 (0.544)	1.305 (0.588)	1.345 (0.646)
38.Banco de sangre	2.640 (2.741)	3.187 (3.325)	3.016 (3.168)	2.773 (2.999)
39.Neumología	0.379*** (0.0377)	0.432*** (0.0443)	0.462*** (0.0475)	0.427*** (0.0469)
40.Transplante	1.432* (0.293)	1.337 (0.279)	1.054 (0.217)	0.972 (0.210)
41.Otra	0.700*** (0.0489)	0.707*** (0.0496)	0.658*** (0.0468)	0.622*** (0.0475)
42.Radiología	0.396*** (0.120)	0.521* (0.185)	0.541* (0.193)	0.467** (0.179)
6.HCUCH	2.767*** (0.0484)	2.184*** (0.0362)	1.828*** (0.0287)	2.127*** (0.0356)
1.sobrecupo	1.596*** (0.0318)	2.393*** (0.0599)	2.555*** (0.0694)	1.983*** (0.0455)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
1.lag_sobrecupo		0.759*** (0.0163)	0.756*** (0.0171)	
1.sobrecupo#1.lag_sobrecupo		0.488*** (0.0211)	0.481*** (0.0228)	
1.web	0.872*** (0.0164)	0.971 (0.0233)	0.944** (0.0227)	0.884*** (0.0163)
1.lag_web		1.079*** (0.0284)	1.078*** (0.0281)	
1.web#1.lag_web		0.787*** (0.0297)	0.830*** (0.0316)	
1.confirm	2.147*** (0.0311)	2.106*** (0.0305)	2.112*** (0.0309)	2.251*** (0.0347)
1.fono_fijo	6.262*** (3.008)	5.173*** (2.208)	4.409*** (2.295)	1.094*** (0.0153)
1.fono_celu	2.191*** (0.545)	1.954*** (0.478)	1.652** (0.381)	0.947 (0.0624)
1.fono_fijo#1.fono_celu	0.171*** (0.0824)	0.201*** (0.0862)	0.238*** (0.124)	
1.fono_trab	2.511*** (0.803)	2.260*** (0.705)	2.234*** (0.693)	0.992 (0.0141)
1.fono_fijo#1.fono_trab	0.135*** (0.0719)	0.164*** (0.0787)	0.172*** (0.0978)	
1.fono_celu#1.fono_trab	0.385*** (0.124)	0.427*** (0.134)	0.437*** (0.136)	
1.fono_fijo#1.fono_celu#1.fono_trab	7.304*** (3.906)	6.177*** (2.978)	5.984*** (3.413)	
1.email	2.066** (0.733)	2.073** (0.721)	1.817* (0.660)	1.018 (0.0141)
1.fono_fijo#1.email	0.147** (0.120)	0.137*** (0.104)	0.105*** (0.0815)	
1.fono_celu#1.email	0.564 (0.201)	0.535* (0.187)	0.583 (0.212)	

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
1.fono_fijo#1.fono_celu#1.email	6.525** (5.333)	7.159*** (5.446)	9.183*** (7.110)	
1.fono_trab#1.email	0.580 (0.266)	0.560 (0.250)	0.496 (0.233)	
1.fono_fijo#1.fono_trab#1.email	5.762** (5.112)	6.142** (5.089)	10.25*** (8.678)	
1.fono_celu#1.fono_trab#1.email	1.646 (0.760)	1.728 (0.775)	1.941 (0.914)	
1.fono_fijo#1.fono_celu#1.fono_trab#1.email	0.195* (0.174)	0.178** (0.148)	0.104*** (0.0881)	
1.marcha	0.950*** (0.0176)	0.951** (0.0187)	0.950** (0.0196)	0.955** (0.0203)
2.month	2.098** (0.703)	1.789* (0.626)	1.290 (0.479)	1.313 (0.506)
3.month	3.367*** (0.961)	2.324*** (0.695)	1.675 (0.528)	1.605 (0.523)
4.month	1.927** (0.491)	1.556* (0.416)	1.276 (0.362)	1.287 (0.377)
5.month	2.991*** (0.706)	2.427*** (0.603)	1.866** (0.491)	1.966** (0.534)
6.month	2.371*** (0.540)	1.766** (0.421)	1.306 (0.331)	1.367 (0.358)
7.month	2.295*** (0.528)	1.890*** (0.456)	1.456 (0.373)	1.471 (0.389)
8.month	2.814*** (0.661)	2.084*** (0.512)	1.514 (0.394)	1.608* (0.433)
9.month	2.304*** (0.537)	1.839** (0.450)	1.377 (0.357)	1.386 (0.372)
10.month	3.242*** (0.770)	2.564*** (0.639)	1.902** (0.502)	1.907** (0.521)
11.month	2.761*** (0.708)	2.336*** (0.627)	2.018** (0.573)	2.179*** (0.639)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
12.month	1.298 (0.393)	1.179 (0.376)	0.938 (0.318)	0.934 (0.326)
temp	1.044*** (0.0101)	1.032*** (0.0105)	1.020* (0.0110)	1.022* (0.0114)
2.month#c.temp	0.964** (0.0149)	0.970* (0.0156)	0.984 (0.0168)	0.983 (0.0174)
3.month#c.temp	0.949*** (0.0131)	0.966** (0.0140)	0.980 (0.0149)	0.983 (0.0155)
4.month#c.temp	0.978* (0.0129)	0.986 (0.0137)	0.993 (0.0146)	0.993 (0.0151)
5.month#c.temp	0.949*** (0.0119)	0.956*** (0.0126)	0.967** (0.0135)	0.964** (0.0138)
6.month#c.temp	0.964*** (0.0122)	0.980 (0.0130)	0.995 (0.0139)	0.992 (0.0144)
7.month#c.temp	0.969** (0.0129)	0.972** (0.0136)	0.981 (0.0145)	0.983 (0.0150)
8.month#c.temp	0.952*** (0.0126)	0.967** (0.0133)	0.982 (0.0143)	0.979 (0.0147)
9.month#c.temp	0.964*** (0.0115)	0.972** (0.0121)	0.985 (0.0130)	0.986 (0.0135)
10.month#c.temp	0.945*** (0.0110)	0.954*** (0.0116)	0.967*** (0.0124)	0.968** (0.0129)
11.month#c.temp	0.953*** (0.0117)	0.960*** (0.0123)	0.965*** (0.0131)	0.961*** (0.0135)
12.month#c.temp	0.984 (0.0141)	0.988 (0.0150)	0.999 (0.0160)	0.999 (0.0165)
2.dia_sem	0.916** (0.0355)	0.900*** (0.0361)	0.897*** (0.0369)	0.988 (0.0194)
3.dia_sem	1.054 (0.0405)	1.052 (0.0422)	1.073* (0.0449)	1.040** (0.0205)
4.dia_sem	1.022 (0.0416)	1.033 (0.0430)	1.053 (0.0451)	1.043** (0.0208)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
5.dia_sem	0.991 (0.0427)	1.036 (0.0465)	1.082* (0.0506)	0.991 (0.0208)
6.dia_sem	0.714*** (0.0448)	0.676*** (0.0437)	0.679*** (0.0448)	0.898*** (0.0357)
2.hourg	1.112*** (0.0406)	1.110*** (0.0416)	1.138*** (0.0440)	1.220*** (0.0220)
3.hourg	0.992 (0.0355)	1.013 (0.0367)	1.059 (0.0393)	1.117*** (0.0202)
4.hourg	1.010 (0.0531)	1.066 (0.0584)	1.123** (0.0635)	1.059** (0.0298)
2.dia_sem#2.hourg	1.033 (0.0505)	1.083 (0.0550)	1.099* (0.0576)	
2.dia_sem#3.hourg	1.141*** (0.0544)	1.171*** (0.0574)	1.180*** (0.0593)	
2.dia_sem#4.hourg	0.967 (0.0677)	0.977 (0.0723)	0.987 (0.0766)	
3.dia_sem#2.hourg	0.978 (0.0474)	1.020 (0.0518)	1.018 (0.0542)	
3.dia_sem#3.hourg	0.947 (0.0445)	0.951 (0.0466)	0.937 (0.0479)	
3.dia_sem#4.hourg	0.961 (0.0661)	0.925 (0.0683)	0.901 (0.0702)	
4.dia_sem#2.hourg	1.015 (0.0513)	1.030 (0.0538)	1.031 (0.0557)	
4.dia_sem#3.hourg	1.014 (0.0501)	0.993 (0.0500)	0.969 (0.0502)	
4.dia_sem#4.hourg	0.931 (0.0659)	0.881* (0.0654)	0.861* (0.0665)	
5.dia_sem#2.hourg	0.939 (0.0493)	0.939 (0.0515)	0.928 (0.0531)	
5.dia_sem#3.hourg	0.981 (0.0511)	0.942 (0.0509)	0.908* (0.0510)	

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
5.dia_sem#4.hourg	0.872* (0.0658)	0.818** (0.0658)	0.753*** (0.0643)	
6.dia_sem#2.hourg	1.434*** (0.108)	1.485*** (0.118)	1.464*** (0.120)	
6.dia_sem#3.hourg	1.820*** (0.249)	1.776*** (0.269)	1.766*** (0.274)	
2.dayg	0.929*** (0.0123)	0.924*** (0.0129)	0.917*** (0.0135)	0.909*** (0.0138)
3.dayg	0.955*** (0.0133)	0.953*** (0.0139)	0.944*** (0.0145)	0.939*** (0.0150)
1.lag1		1.758*** (0.0444)	1.318*** (0.0510)	1.184*** (0.0281)
1.lag2			1.039 (0.0368)	
1.lag1#1.lag2			0.745*** (0.0323)	
1.lag3			1.395*** (0.0438)	
1.lag1#1.lag3			0.877*** (0.0371)	
1.lag2#1.lag3			0.854*** (0.0363)	
1.lag1#1.lag2#1.lag3			1.589*** (0.0875)	
1.lag1#c.tc		0.999*** (7.76e-05)	0.999*** (7.92e-05)	0.999*** (8.42e-05)
tc	1.000*** (4.00e-05)	1.001*** (6.64e-05)	1.001*** (6.78e-05)	1.001*** (7.17e-05)
d	1.004*** (0.000795)	1.002*** (0.000697)	1.002*** (0.000622)	1.001** (0.000285)
d2	1.000*** (9.24e-07)	1.000*** (8.39e-07)	1.000** (7.10e-07)	

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
t2	1.000 (7.39e-09)	1.000 (6.23e-09)	1.000** (8.09e-09)	
1.lag_prof		0.708*** (0.0765)	0.623*** (0.0664)	
1.lag1#1.lag_prof		1.573*** (0.191)	1.751*** (0.214)	
1.lag_espe		0.818*** (0.0221)	0.831*** (0.0232)	
1.lag1#1.lag_espe		1.513*** (0.0491)	1.386*** (0.0453)	
1.lag_prof#1.lag_espe		0.775** (0.0861)	0.894 (0.0982)	
1.lag1#1.lag_prof#1.lag_espe		2.060*** (0.258)	1.561*** (0.197)	2.137*** (0.0413)
1.lag_espe2			0.815*** (0.0232)	
1.lag2#1.lag_espe2			1.502*** (0.0495)	
1.lag_prof2			0.459*** (0.0516)	
1.lag2#1.lag_prof2			2.832*** (0.364)	
1.lag_espe2#1.lag_prof2			1.217* (0.142)	
1.lag2#1.lag_espe2#1.lag_prof2			1.198 (0.160)	
lagfiabilidad			5.324*** (0.643)	3.722*** (0.466)
lagpor			8.699*** (0.372)	20.63*** (0.777)
c.lagfiabilidad#c.lagpor			0.121*** (0.0191)	0.217*** (0.0357)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
1.futbol	0.966* (0.0171)	0.958** (0.0179)	0.956** (0.0187)	0.959** (0.0194)
1.preferiado	0.940** (0.0254)	0.942** (0.0270)	0.951* (0.0283)	0.950* (0.0292)
1.feriado	1.037 (0.0481)	1.041 (0.0502)	1.033 (0.0526)	1.038 (0.0547)
1.postferiado	0.981 (0.0228)	0.971 (0.0238)	0.966 (0.0250)	0.970 (0.0260)
edad	0.972*** (0.00168)	0.976*** (0.00156)	0.985*** (0.00146)	1.007*** (0.000439)
3.gender	0.949* (0.0297)	0.958 (0.0280)	0.966 (0.0260)	1.033 (0.0285)
3.gender#c.edad	1.004*** (0.000699)	1.004*** (0.000650)	1.003*** (0.000594)	1.003*** (0.000598)
1.email#c.edad	0.998** (0.000708)	0.999* (0.000656)	1.000 (0.000596)	
edad2	1.000*** (1.87e-05)	1.000*** (1.72e-05)	1.000*** (1.58e-05)	
1.lluviag	0.976 (0.0306)	0.957 (0.0316)	0.944* (0.0326)	0.951 (0.0341)
2.lluviag	0.956* (0.0253)	0.947* (0.0265)	0.940** (0.0278)	0.935** (0.0286)
3.lluviag	0.858*** (0.0489)	0.853*** (0.0521)	0.868** (0.0563)	0.844** (0.0561)
4.lluviag	0.349*** (0.0628)	0.308*** (0.0576)	0.278*** (0.0532)	0.257*** (0.0513)
1.gescaec	1.792*** (0.180)	1.852*** (0.186)	1.665*** (0.168)	1.744*** (0.181)
1.post	2.612*** (0.227)	2.510*** (0.222)	2.267*** (0.199)	2.402*** (0.227)
laten	0.978*** (0.000537)	0.976*** (0.000544)	0.974*** (0.000559)	0.975*** (0.000563)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
lag_laten		0.994*** (0.000898)	1.001 (0.000926)	
1.lag1#c.lag_laten		1.022*** (0.00115)	1.012*** (0.00114)	
ipc	0.892*** (0.0197)	0.891*** (0.0200)	0.923*** (0.0214)	0.925*** (0.0224)
antigüedad	1.000*** (1.15e-05)	1.000* (1.08e-05)	1.000*** (1.11e-05)	1.000*** (1.18e-05)
t				1.000* (4.72e-05)
0b.lag1#0b.lag_prof#1.lag_espe				0.794*** (0.0230)
0b.lag1#1.lag_prof#0b.lag_espe				0.586*** (0.0665)
0b.lag1#1.lag_prof#1.lag_espe				0.393*** (0.0105)
1.lag1#0b.lag_prof#1.lag_espe				1.194*** (0.0239)
1.lag1#1.lag_prof#0b.lag_espe				1.136* (0.0758)
Constant	0.636 (0.211)	0.574* (0.193)	0.168*** (0.0568)	0.147*** (0.0388)
Observations	300,000	300,000	300,000	300,000
LR χ^2	24924	41598	55898	44887
Prob $<\chi^2$	0	0	0	0
McFadden's R^2	0.197	0.268	0.323	
McFadden's R^2 Adjust	0.196	0.267	0.321	
Mckelvey and Zavoina's R^2	0.317	0.408	0.489	
LL	-126831	-115573	-106956	-110140
Clusters	77141	77141	77141	77141

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Modelo I</i>	<i>Logit Modelo II</i>	<i>Logit Modelo III</i>	<i>Mixed Logit Modelo IV (Fixed effects)</i>
Number of groups				77,141

Note: Constant estimates baseline odds (conditional on zero random effects).

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

B.3. Ajuste del Modelo III con función de enlace LOGIT con extensión de data.

<i>Variable</i>	<i>Logit Model III (Muestra)</i>	<i>Logit Model III (Expansión de data)</i>
1.Cardiología/Cardiología	-0.162*** (0.0202)	0.850*** (0.0172)
2.Obesidad y nutrición	-0.789*** (0.0208)	0.454*** (0.00945)
3.Cirugía	-0.486*** (0.0181)	0.615*** (0.0111)
4.Dermatología	-0.366*** (0.0196)	0.693*** (0.0136)
5.Diabetes	-0.163*** (0.0461)	0.849*** (0.0392)
6.Dolor	0.173*** (0.0518)	1.189*** (0.0615)
7.Endocrinología	0.321*** (0.0457)	1.379*** (0.0630)
8.Fonoaudiología	-1.262*** (0.0379)	0.283*** (0.0107)
9.Gastroenterología	-0.108*** (0.0216)	0.897*** (0.0194)
10.Genética	-0.669*** (0.124)	0.512*** (0.0633)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Model III (Muestra)</i>	<i>Logit Model III (Expansión de data)</i>
11.Geriatría	-0.000263 (0.0617)	1.000 (0.0617)
12.Hematología	-0.0904* (0.0516)	0.914* (0.0472)
14.Inmunología y alergias	-0.270*** (0.0294)	0.763*** (0.0225)
15.Medicina familiar	-0.451* (0.247)	0.637* (0.158)
16.Medicina física y rehabilitación	-0.584*** (0.0224)	0.557*** (0.0125)
18.Medicina interna	0.0817*** (0.0256)	1.085*** (0.0278)
21.Nefrología	-0.195*** (0.0486)	0.823*** (0.0400)
22.Neonatología/Pediatría	-0.313*** (0.0194)	0.732*** (0.0142)
23.Neurología/Neurocirugía	-0.607*** (0.0556)	0.545*** (0.0303)
24.Obstetricia/Ginecología	-0.405*** (0.0159)	0.667*** (0.0106)
25.Oftalmología	-0.207*** (0.0180)	0.813*** (0.0146)
26.Oncología	-1.038*** (0.0347)	0.354*** (0.0123)
27.Otorrinolaringología	-0.584*** (0.0178)	0.558*** (0.00995)
28.Psicología	-0.368*** (0.0229)	0.692*** (0.0159)
29.Psiquiatría	0.153*** (0.0296)	1.165*** (0.0345)
30.Respiratorio	-0.979*** (0.0429)	0.376*** (0.0161)
31.Reumatología	-0.0426 (0.0296)	0.958 (0.0284)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Model III (Muestra)</i>	<i>Logit Model III (Expansión de data)</i>
32.Traumatología	0.0931*** (0.0188)	1.098*** (0.0206)
33.Urología	-0.0731*** (0.0228)	0.930*** (0.0212)
34.Dentomaxilofacial	-1.956*** (0.0193)	0.141*** (0.00273)
35.Laboratorio	1.760*** (0.0796)	5.815*** (0.463)
36.Investigación clínica	-3.192*** (0.263)	0.0411*** (0.0108)
37.Anestesiología	0.0930 (0.219)	1.097 (0.240)
38.Banco de sangre	-0.135 (0.310)	0.874 (0.271)
39.Neumología	-0.789*** (0.0470)	0.454*** (0.0213)
40.Transplante	-0.288*** (0.0794)	0.750*** (0.0596)
41.Otra	-0.431*** (0.0327)	0.650*** (0.0213)
42.Radiología	-0.479*** (0.142)	0.620*** (0.0882)
6.HCUCH	0.673*** (0.00845)	1.960*** (0.0166)
1.sobrecupo	0.894*** (0.0126)	2.446*** (0.0309)
1.lag_sobrecupo	-0.251*** (0.0100)	0.778*** (0.00780)
1.sobrecupo#1.lag_sobrecupo	-0.793*** (0.0226)	0.452*** (0.0102)
1.web	-0.115*** (0.0115)	0.891*** (0.0102)
1.lag_web	0.0398*** (0.0114)	1.041*** (0.0118)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Model III (Muestra)</i>	<i>Logit Model III (Expansión de data)</i>
1.web#1.lag_web	-0.194*** (0.0165)	0.823*** (0.0136)
1.confirm	0.647*** (0.00721)	1.910*** (0.0138)
1.fono_fijo	0.709*** (0.274)	2.033*** (0.557)
1.fono_celu	0.369*** (0.100)	1.446*** (0.145)
1.fono_fijo#1.fono_celu	-0.652** (0.275)	0.521** (0.143)
1.fono_trab	0.301** (0.142)	1.351** (0.191)
1.fono_fijo#1.fono_trab	-0.566* (0.296)	0.568* (0.168)
1.fono_celu#1.fono_trab	-0.314** (0.143)	0.730** (0.104)
1.fono_fijo#1.fono_celu#1.fono_trab	0.596** (0.298)	1.815** (0.540)
1.email	0.291* (0.173)	1.338* (0.231)
1.fono_fijo#1.email	-0.639 (0.423)	0.528 (0.223)
1.fono_celu#1.email	-0.310* (0.173)	0.734* (0.127)
1.fono_fijo#1.fono_celu#1.email	0.612 (0.424)	1.843 (0.781)
1.fono_trab#1.email	-0.242 (0.223)	0.785 (0.175)
1.fono_fijo#1.fono_trab#1.email	0.488 (0.455)	1.629 (0.741)
1.fono_celu#1.fono_trab#1.email	0.231 (0.224)	1.259 (0.282)
1.fono_fijo#1.fono_celu #1.fono_trab#1.email	-0.452 (0.456)	0.636 (0.290)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Model III (Muestra)</i>	<i>Logit Model III (Expansión de data)</i>
1.marcha	-0.0298** (0.0127)	0.971** (0.0124)
2.month	0.510*** (0.150)	1.666*** (0.250)
3.month	0.425*** (0.133)	1.530*** (0.203)
4.month	0.172 (0.118)	1.188 (0.140)
5.month	0.405*** (0.110)	1.499*** (0.165)
6.month	0.226** (0.106)	1.253** (0.133)
7.month	0.166 (0.107)	1.181 (0.126)
8.month	0.326*** (0.109)	1.385*** (0.152)
9.month	0.145 (0.108)	1.156 (0.125)
10.month	0.296*** (0.111)	1.344*** (0.149)
11.month	0.188 (0.120)	1.206 (0.144)
12.month	-0.111 (0.138)	0.895 (0.124)
temp	0.0145*** (0.00452)	1.015*** (0.00458)
2.month#c.temp	-0.0310*** (0.00691)	0.969*** (0.00670)
3.month#c.temp	-0.0182*** (0.00641)	0.982*** (0.00630)
4.month#c.temp	-0.00791 (0.00608)	0.992 (0.00603)
5.month#c.temp	-0.0246*** (0.00585)	0.976*** (0.00571)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Model III (Muestra)</i>	<i>Logit Model III (Expansión de data)</i>
6.month#c.temp	-0.0165*** (0.00597)	0.984*** (0.00587)
7.month#c.temp	-0.00815 (0.00609)	0.992 (0.00604)
8.month#c.temp	-0.0182*** (0.00611)	0.982*** (0.00600)
9.month#c.temp	-0.00749 (0.00549)	0.993 (0.00545)
10.month#c.temp	-0.0157*** (0.00544)	0.984*** (0.00535)
11.month#c.temp	-0.0102* (0.00573)	0.990* (0.00567)
12.month#c.temp	-0.00153 (0.00652)	0.998 (0.00651)
2.dia_sem	-0.148*** (0.0194)	0.862*** (0.0167)
3.dia_sem	0.0353** (0.0179)	1.036** (0.0185)
4.dia_sem	0.0529*** (0.0195)	1.054*** (0.0205)
5.dia_sem	0.0212 (0.0199)	1.021 (0.0204)
6.dia_sem	-0.364*** (0.0287)	0.695*** (0.0199)
2.hourg	0.129*** (0.0168)	1.138*** (0.0191)
3.hourg	0.0588*** (0.0165)	1.061*** (0.0175)
4.hourg	0.0740*** (0.0237)	1.077*** (0.0255)
2.dia_sem#2.hourg	0.144*** (0.0239)	1.155*** (0.0276)
2.dia_sem#3.hourg	0.188*** (0.0234)	1.207*** (0.0283)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Model III (Muestra)</i>	<i>Logit Model III (Expansión de data)</i>
2.dia_sem#4.hourg	0.109*** (0.0339)	1.115*** (0.0378)
3.dia_sem#2.hourg	-0.00686 (0.0226)	0.993 (0.0224)
3.dia_sem#3.hourg	0.00264 (0.0218)	1.003 (0.0219)
3.dia_sem#4.hourg	-0.0143 (0.0322)	0.986 (0.0317)
4.dia_sem#2.hourg	0.00609 (0.0239)	1.006 (0.0241)
4.dia_sem#3.hourg	0.00521 (0.0235)	1.005 (0.0236)
4.dia_sem#4.hourg	-0.0951*** (0.0333)	0.909*** (0.0303)
5.dia_sem#2.hourg	-0.0395 (0.0245)	0.961 (0.0236)
5.dia_sem#3.hourg	-0.0211 (0.0239)	0.979 (0.0234)
5.dia_sem#4.hourg	-0.0984*** (0.0354)	0.906*** (0.0321)
6.dia_sem#2.hourg	0.337*** (0.0344)	1.401*** (0.0481)
6.dia_sem#3.hourg	0.430*** (0.0697)	1.538*** (0.107)
2.dayg	-0.0852*** (0.00626)	0.918*** (0.00574)
3.dayg	-0.0570*** (0.00657)	0.945*** (0.00621)
1.lag1	0.252*** (0.0177)	1.287*** (0.0228)
1.lag2	0.0192 (0.0164)	1.019 (0.0167)
1.lag1#1.lag2	-0.238*** (0.0195)	0.788*** (0.0153)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Model III (Muestra)</i>	<i>Logit Model III (Expansión de data)</i>
1.lag3	0.391*** (0.0146)	1.478*** (0.0216)
1.lag1#1.lag3	-0.217*** (0.0187)	0.805*** (0.0151)
1.lag2#1.lag3	-0.192*** (0.0189)	0.826*** (0.0156)
1.lag1#1.lag2#1.lag3	0.533*** (0.0245)	1.705*** (0.0418)
1.lag1#c.tc	-0.000755*** (3.32e-05)	0.999*** (3.32e-05)
tc	0.000597*** (2.89e-05)	1.001*** (2.89e-05)
d	0.00107*** (0.000350)	1.001*** (0.000351)
d2	-1.23e-06*** (4.05e-07)	1.000*** (4.05e-07)
t2	2.17e-08*** (5.60e-09)	1.000*** (5.60e-09)
1.lag_prof	-0.420*** (0.0493)	0.657*** (0.0324)
1.lag1#1.lag_prof	0.570*** (0.0546)	1.768*** (0.0965)
1.lag_espe	-0.129*** (0.0125)	0.879*** (0.0110)
1.lag1#1.lag_espe	0.282*** (0.0147)	1.326*** (0.0195)
1.lag_prof#1.lag_espe	-0.213*** (0.0505)	0.808*** (0.0408)
1.lag1#1.lag_prof#1.lag_espe	0.539*** (0.0562)	1.714*** (0.0963)
1.lag_espe2	-0.141*** (0.0125)	0.869*** (0.0108)
1.lag2#1.lag_espe2	0.385*** (0.0147)	1.470*** (0.0217)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Model III (Muestra)</i>	<i>Logit Model III (Expansión de data)</i>
1.lag_prof2	-0.499*** (0.0461)	0.607*** (0.0280)
1.lag2#1.lag_prof2	0.815*** (0.0534)	2.260*** (0.121)
1.lag_espe2#1.lag_prof2	-0.112** (0.0478)	0.894** (0.0427)
1.lag2#1.lag_espe2#1.lag_prof2	0.447*** (0.0552)	1.563*** (0.0862)
lagfiabilidad	1.559*** (0.0571)	4.754*** (0.271)
lagpor	2.124*** (0.0250)	8.364*** (0.209)
c.lagfiabilidad#c.lagpor	-2.261*** (0.0739)	0.104*** (0.00770)
1.futbol	-0.0301*** (0.0117)	0.970*** (0.0113)
1.preferiado	-0.0367*** (0.0127)	0.964*** (0.0122)
1.feriado	0.0783*** (0.0212)	1.081*** (0.0229)
1.postferiado	-0.0495*** (0.0111)	0.952*** (0.0105)
edad	-0.0144*** (0.000830)	0.986*** (0.000819)
3.gender	0.0121 (0.0144)	1.012 (0.0146)
3.gender#c.edad	0.00200*** (0.000338)	1.002*** (0.000338)
1.email#c.edad	0.000663** (0.000334)	1.001** (0.000334)
edad2	0.000238*** (9.08e-06)	1.000*** (9.08e-06)
1.lluviag	-0.0100 (0.0150)	0.990 (0.0149)

(Continúa en la página siguiente)

(Continúa de la página anterior)

<i>Variable</i>	<i>Logit Model III (Muestra)</i>	<i>Logit Model III (Expansión de data)</i>
2.lluviag	-0.0223* (0.0123)	0.978* (0.0120)
3.lluviag	-0.202*** (0.0267)	0.817*** (0.0218)
4.lluviag	-0.991*** (0.0776)	0.371*** (0.0288)
1.gescaec	0.463*** (0.0497)	1.588*** (0.0790)
1.post	0.752*** (0.0414)	2.121*** (0.0877)
laten	-0.0249*** (0.000309)	0.975*** (0.000302)
lag_laten	0.00140*** (0.000417)	1.001*** (0.000418)
1.lag1#c.lag_laten	0.0131*** (0.000518)	1.013*** (0.000525)
ipc	-0.0544*** (0.0104)	0.947*** (0.00981)
antigüedad	6.00e-05*** (5.68e-06)	1.000*** (5.68e-06)
Constant	-1.462*** (0.145)	0.232*** (0.0336)
Observations	300,000	1,585,098
Prob $<\chi^2$	0	0
McFadden's R2	0.328	0.322
LL	-567249	-567249.086
Clusters	77,141	111,998

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1