

DISSERTATIO

Estadística

COLEGIO DE GRADUADOS EN CIENCIAS ECONÓMICAS DE ROSARIO
CONSEJO PROFESIONAL DE CIENCIAS ECONÓMICAS
DE LA PROVINCIA DE SANTA FE CÁMARA II
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA

TRABAJOS FINALES

RECENSIÓN DE TESIS Y PRÁCTICAS PROFESIONALES
DE LA CARRERA LICENCIATURA EN ESTADÍSTICA



CONSEJO PROFESIONAL
DE CIENCIAS ECONÓMICAS
DE LA PROVINCIA DE SANTA FE
CÁMARA II



Colegio de Graduados
en Ciencias Económicas
de Rosario



Universidad
Nacional
de Rosario

ÍNDICE

CONFORMACIONES 03

CONSTRUYENDO JUNTOS EL FUTURO DE LAS PROFESIONES 04

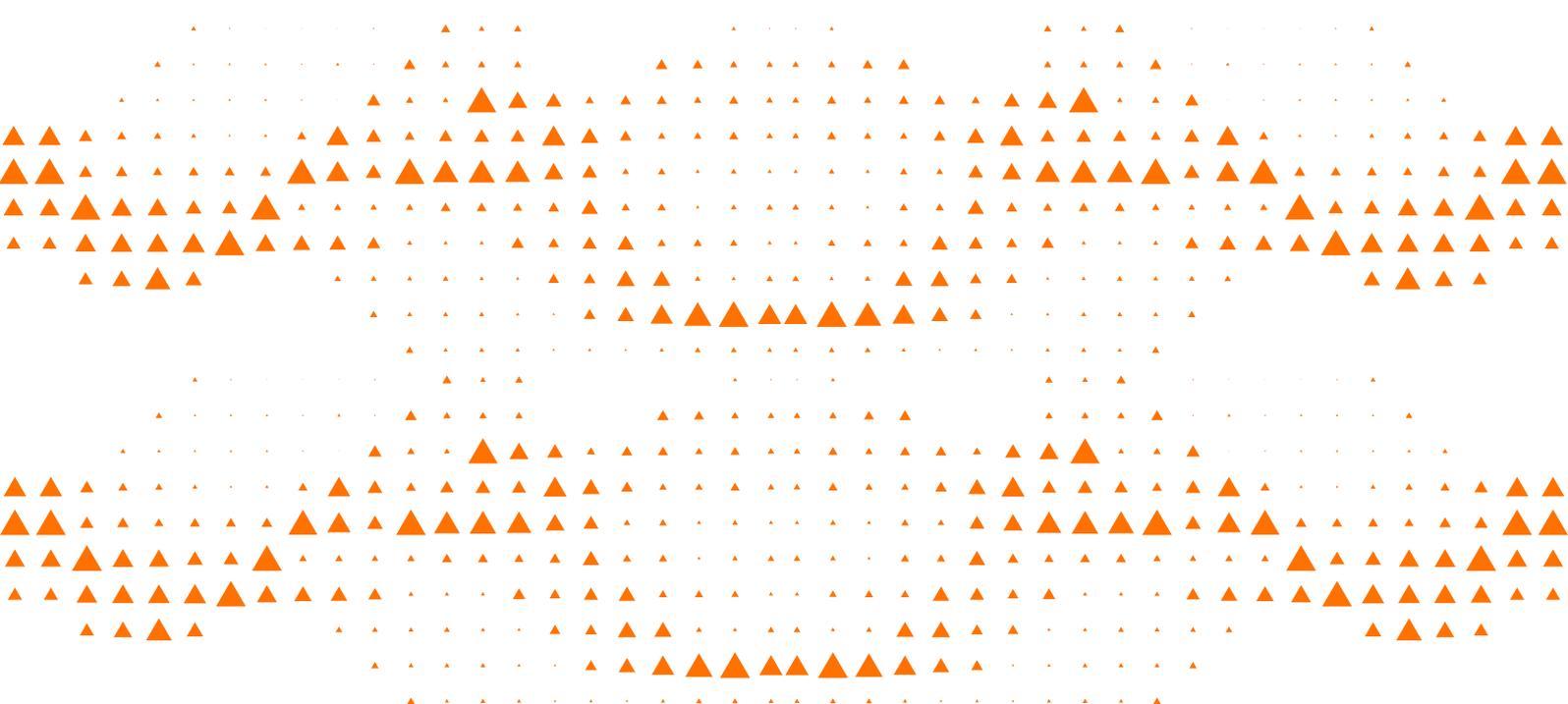
ARTÍCULOS

EFFECTIVIDAD DE UNA INTERVENCIÓN MULTI COMPONENTE MEDIADA
POR EL EQUIPO DE SALUD FAMILIAR Y COMUNITARIA PARA EL
MEJORAMIENTO DE LA ADHERENCIA TERAPÉUTICA EN PERSONAS
CON ENFERMEDADES CRÓNICAS 05
LIC. MARÍA CECILIA CÓCCARO

ADECUACIÓN DEL MODELO DE POISSON CONDICIONAL
AUTORREGRESIVO EN SERIES DE TIEMPO DE DISTINTA MAGNITUD 13
LIC. NOELIA CARLA HEGI

COMPARACIÓN DE ANÁLISIS INTERMEDIOS EN ENSAYOS CLÍNICOS.
ESTUDIO PARA EL TRATAMIENTO DE PACIENTES HOSPITALIZADOS POR COVID19 23
LIC. FRANCO IVÁN SCARAFIA

TÉCNICAS GRÁFICAS PARA LA EVALUACIÓN DE SUPUESTOS EN MODELOS
LINEALES MIXTOS PARA DATOS LONGITUDINALES 34
LIC. MARLÉN ALTAMIRANO



COMITÉ DIRECTIVO

Mg. Javier Ganem (FCEyE)
Dr. Carlos Omegna (CPCE)
Dra. Nanci Eterovich (CGCE)

COMITÉ ACADÉMICO

Mg. Cristina Beatriz Cuesta (FCEyE)
Mg. Virginia Laura Borra (FCEyE)
Dra. Daniela Dianda (FCEyE)
Est. Nora Ventroni (CPCE-CGCE)

COMITÉ EDITORIAL

Lic. Maite San Martin (FCEyE)
Mg. Guillermina Beatriz Harvey (FCEyE)
Mg. Laura Rita Balparda (CPCE-CGCE)
Lic. Florencia Yamila Ruiz (CPCE- CGCE)

Esta revista se pone a disposición de los profesionales matriculados al Consejo Profesional de Ciencias Económicas de la Provincia de Santa Fe Cámara II (CPCE), asociados del Colegio de Graduados en Ciencias Económicas de Rosario (CGCE), estudiantes y docentes de la Facultad de Ciencias Económicas y Estadística (FCEyE) de la Universidad Nacional de Rosario (UNR) y otras Instituciones vinculadas al quehacer profesional y académico.

Su contenido puede ser reproducido en forma parcial o total citando la fuente. En caso de utilización deberá enviar dos ejemplares de la publicación respectiva a **Maipú 1344 – 2000 Rosario Tel. 4772727 email: consejo@cpcesfe2.org.ar**

El contenido de los trabajos finales no necesariamente refleja la opinión de los Comités responsables de esta publicación digital.

Las Instituciones no son responsables por el contenido de las informaciones y opiniones que viertan en esta revista quienes son identificados como autores de dichos trabajos finales, en todos los casos deberán ser cotejadas por los Profesionales y/o las fuentes.





CONSTRUYENDO JUNTOS EL FUTURO DE LAS PROFESIONES

El Consejo Profesional de Ciencias Económicas de la Provincia de Santa Fe Cámara II, el Colegio de Graduados en Ciencias Económicas de Rosario y la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario llevan a cabo muchas actividades conjuntas en pos del desarrollo de los futuros graduados, incentivando la investigación y fomentando la jerarquización profesional a través de la actualización y capacitación permanentes. Entendiendo indispensable la formación para el ejercicio laboral acorde a los tiempos cambiantes y a la transformación digital, y brindando las herramientas para hacerla posible. Esta revista, que compila una selección de tesis de grado y trabajos finales de las licenciaturas que integran la Facultad, es una de

las tantas acciones que las instituciones profesionales y la casa de altos estudios realizan en este sentido y con el objetivo de acompañar a estudiantes avanzados y jóvenes graduados en su paso de la vida universitaria a la vida profesional, ofreciendo oportunidades de nuevos aprendizajes y desafíos. De esta manera, orgullosos del camino recorrido presentamos la novena edición de *Dissertatio Economía*, que fue el origen de este proyecto común en 2015, y la séptima edición de *Dissertatio Estadística* y la sexta de *Dissertatio Administración*, cuyos primeros ejemplares se publicaron en 2017 y 2018 respectivamente.

EFECTIVIDAD DE UNA INTERVENCIÓN MULTI COMPONENTE MEDIADA POR EL EQUIPO DE SALUD FAMILIAR Y COMUNITARIA PARA EL MEJORAMIENTO DE LA ADHERENCIA TERAPÉUTICA EN PERSONAS CON ENFERMEDADES CRÓNICAS

Lic. María Cecilia Cocco

Responsable de la FCEyE: **Mg. Cristina Cuesta**

Responsable del Instituto de Efectividad Clínica y Sanitaria:

Lic. Laura Gutierrez

Un problema de salud que ha cobrado relevancia últimamente es la deficiente adherencia de los pacientes al tratamiento de las enfermedades crónicas. Los motivos son tan variados que cualquier enfoque para abordarlo debe, obligadamente, ser multifactorial.

A los fines de estudiar la adherencia al tratamiento de la hipertensión arterial, se diseñó e implementó una intervención multi-componente en el municipio de Almirante Brown, provincia de Buenos Aires. El objetivo principal es comparar el grado de adherencia antes y después de la intervención. Los componentes de la misma fueron: 1) Enfermero/a: evaluación de la tasa de posesión de medicamentos antihipertensivos y emisión de alarmas, 2) Médico/a: implementación de una lista de cotejo en la consulta para la evaluación, monitoreo y refuerzo de la adherencia terapéutica y 3) Agente sanitario/enfermería: llamados telefónicos al primer y tercer o cuarto mes de iniciado el tratamiento, orientados a monitorear y fortalecer la adherencia terapéutica.

Se utilizó como indicador de adherencia el retiro de medicación del paciente en su centro de salud habitual. Los resultados principales muestran un incremento significativo de la adherencia luego de la intervención. Se detectó, probablemente como consecuencia de lo anterior, una disminución significativa de los niveles medios de presión arterial post intervención.



INTRODUCCIÓN

El tratamiento de enfermedades crónicas comúnmente incluye el uso de farmacoterapia a largo plazo. Aunque estos medicamentos son efectivos para combatir la enfermedad, la totalidad de sus beneficios a menudo no son alcanzados debido al alto porcentaje de los pacientes que no toma la medicación como fue prescrita (Sabaté, 2003). Los factores que contribuyen a la pobre adherencia de la medicación son variados e incluyen aquellos relacionados a los pacientes, a los médicos y al sistema de salud.

La Organización Mundial de la Salud (OMS) ha proclamado a la inadecuada adherencia terapéutica a los tratamientos crónicos como uno de los problemas más serios de la salud pública (Brown & Bussell, 2011; Osterberg & Blaschke, 2005), afectando principalmente a los países en desarrollo e incluso dentro de los mismos países, distinguiendo entre zonas de mayores y menores recursos (Yusuf & Rangarajan, 2014). Aún en los países desarrollados, la adherencia a los tratamientos crónicos en la población general ronda el 50%, tasa que es mucho menor en países en desarrollo (Rohina & Jan, 2008). Esta realidad, está íntimamente vinculada al fracaso terapéutico, mayores tasas de hospitalización, aumento de costos sanitarios y aumento de mortalidad por todas las causas (Rohina & Jan, 2008; Khatib & Schwalm, 2014).

En Argentina la tasa de diagnóstico de la hipertensión arterial (HTA) es del 52.7%, de los pacientes diagnosticados sólo están en tratamiento el 43.0% y con la presión arterial controlada apenas el 15.9% (Tolonen & Keil, 2005).

Por otro lado, una revisión sistemática específica sobre las barreras relacionadas con el control de la HTA describe un amplio número de problemas asociados tanto con el paciente como con el sistema de salud, indicando la necesidad del diseño de intervenciones multi-componentes para el mejor control de este factor de riesgo cardiovascular. En especial se destaca la necesidad de evaluar la implementación de estas intervenciones en países de bajos y medianos recursos con el objetivo de perfeccionar las políticas futuras de salud pública (Yusuf & Rangarajan, 2014).

El especialista en Cardiología y miembro del Instituto de Efectividad Clínica y Sanitaria (IECS), Dr. Pablo Gulayin, se ha interesado sobre la baja adherencia al tratamiento de enfermedades crónicas, en particular la HTA, ya que es uno de los principales factores de riesgo cardiovasculares y la enfermedad cardiovascular (ECV) es la primera causa de enfermedad tanto en nuestro país como en el resto del mundo. A fin de atender esta inquietud, propuso estudiar el efecto de una intervención en pacientes hipertensos no controlados, sobre la adherencia al tratamiento antihipertensivo, en el nivel primario de salud. El diseño de investigación consistió en un estudio cuasi-experimental del tipo antes/después en 5 Centros de Atención Primaria de Salud (CAPS) del municipio de Almirante Brown, donde se reclutaron 125 participantes. Se realizaron mediciones pre-intervención, luego se implementaron los componentes de la intervención durante 6 meses.

La intervención desarrollada en el estudio tuvo 3 componentes:

- a) monitoreo de la posesión de medicamentos antihipertensivos y emisión de alarmas al equipo,
- b) implementación de una lista de cotejo en la consulta (para la evaluación, monitoreo y refuerzo de la adherencia terapéutica),
- c) contactos telefónicos orientados a monitorear y fortalecer la adherencia terapéutica.

Se analizó la adherencia al tratamiento de la hipertensión arterial, antes y después de la intervención.

En este informe se presentan los resultados más relevantes obtenidos del análisis estadístico de los datos recolectados. Se espera que la evidencia generada en este estudio promueva el desarrollo de intervenciones más efectivas para el mejoramiento de la adherencia terapéutica en patologías crónicas.

OBJETIVOS

Objetivo primario: evaluar si la intervención multi-componente mediada por los miembros del equipo nuclear de salud familiar y comunitaria durante 6 meses logra obtener una adherencia adecuada a la medicación anti-hipertensiva en pacientes hipertensos mal controlados.

Objetivo secundario: evaluar si la intervención logra una reducción de los valores de presión arterial en pacientes hipertensos mal controlados.

MATERIALES Y MÉTODOS

Entre enero y marzo de 2021 se reclutaron 125 pacientes con diagnóstico de HTA mal controlada con al menos 1.5 años de diagnóstico, en 5 CAPS del sistema público de la localidad de Almirante Brown (25 participantes en cada CAPS). Estos pacientes tuvieron un seguimiento de 6 meses, finalizando el trabajo de campo en octubre de 2021.

El tipo de diseño elegido para evaluar los objetivos mencionados fue un estudio cuasi-experimental del tipo antes/después no controlado. Este tipo de diseño es especialmente utilizado en el campo de la investigación de intervenciones educativas en el ámbito de la salud pública. Se caracteriza por la realización de una medición basal, seguida de la implementación de una intervención y una nueva evaluación luego de la intervención.

El componente principal de la intervención se focalizó en el monitoreo de los retiros de farmacia, lo que permitió evaluar la Tasa de Posesión de la Medicación (TPM), con el objetivo de identificar tempranamente a los pacientes con retiro de medicación insuficiente (menor al 80%).

Luego de ser enrolados los pacientes al estudio se registraron, de manera retrospectiva, mediciones de la TPM de medicamentos anti-hipertensivos haciendo uso de los retiros históricos de farmacia desde noviembre de 2019 hasta julio de 2020, excepto en marzo de 2020 (debido a ser el primer mes del período de Aislamiento Social Preventivo y Obligatorio). Luego, durante todo el seguimiento, se evaluó el retiro de medicación anti-hipertensiva de la farmacia de los CAPS (Figuras 1 y 2).

Figura 1. Etapas del estudio

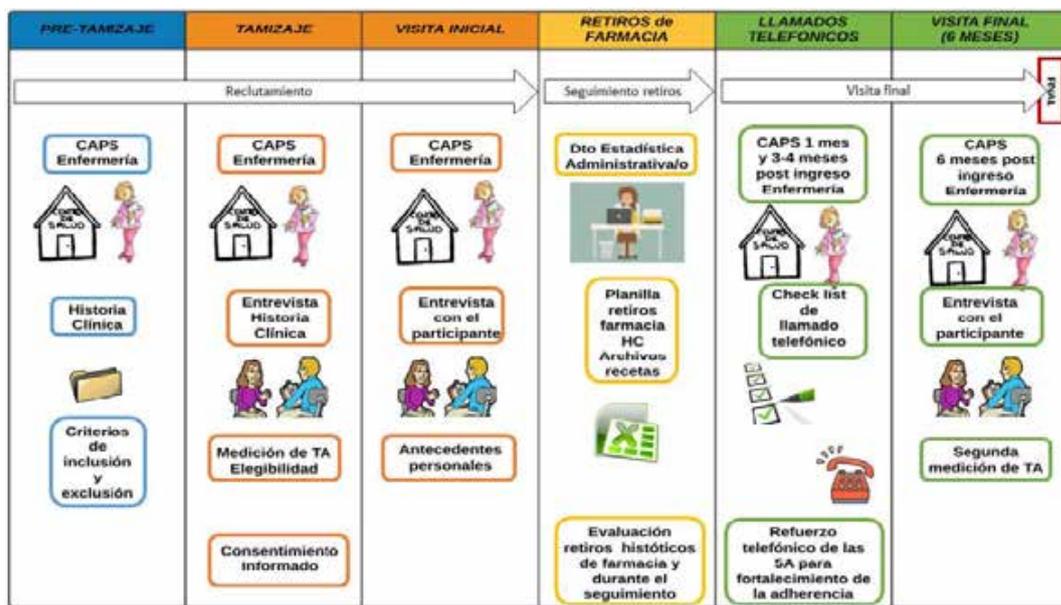
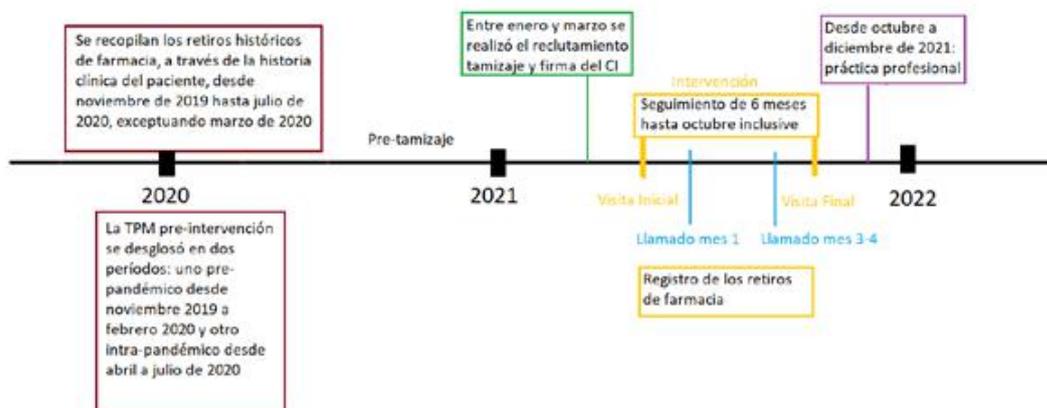


Figura 2. Línea de tiempo de la intervención y práctica profesional



Se calculó la TPM como:

$$TPM = \left(\frac{\text{Cantidad total de comprimidos retirados de la farmacia en el periodo } X}{\text{Cantidad teórica de comprimidos a retirar de la farmacia en el periodo } X} \right) \times 100$$

También se realizó la toma de la presión arterial en dos momentos cruciales: al ingreso del estudio y al finalizar la intervención.

Se resumieron las características basales de la población presentando media y desvío estándar (DE), mediana y rango intercuartil (RI) para las variables continuas y proporciones para las variables categóricas. Para las comparaciones antes/después propuestas por los objetivos del estudio se utilizó el test de Mc Nemar para las variables categóricas y el test de los rangos señalados de Wilcoxon en el caso de las variables continuas. Se consideró significativa una probabilidad asociada al test menor a 0.05.

Para el análisis de los datos se utilizó el paquete estadístico *R Core Team* (Team, 2021) y *R Studio* en la versión 4.1.2.

La recolección de la información del estudio se llevó a cabo con REDCap. Este es un sistema informático que permite cargar datos en dispositivos electrónicos (en este caso una *Tablet*), seleccionando opciones preestablecidas o de forma manual, preguntas de respuesta simple o múltiple. En esta plataforma se encontraban todos los formularios del estudio.

RESULTADOS

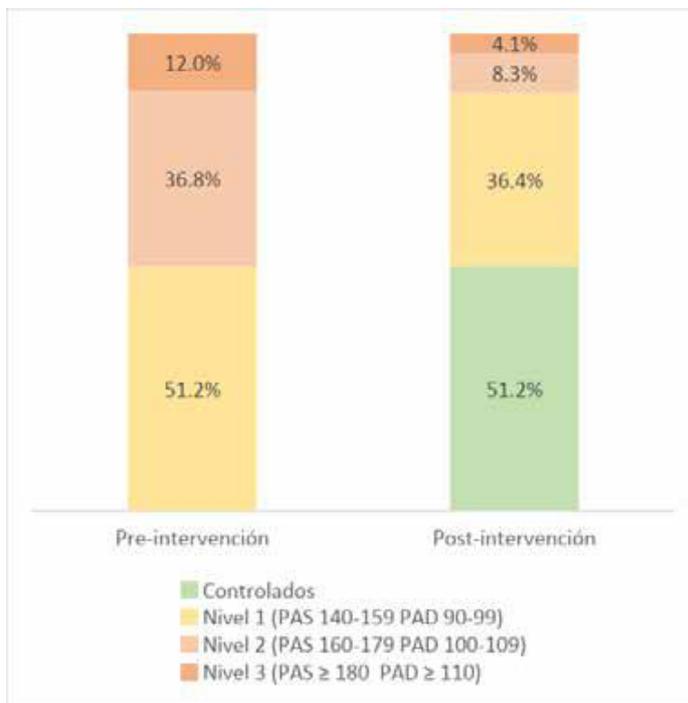
Casi la mitad de la población fue de sexo masculino, promedio de edad de 57 años, el 87.2% de la población tenía un nivel educativo de secundaria incompleta o menor y el 83.5% contaba con cobertura en salud pública exclusiva (16.5% de los participantes reportaron una fuente de cobertura adicional). En relación al tratamiento de los participantes al inicio de la evaluación pre-intervención, la mediana de antigüedad de dichos esquemas terapéuticos de medicación fue de 15 días (mín 6 – máx 43 días). Por otro lado, el 56.8% (71/125) de los participantes se encontraba tratado sólo con una droga, 36.0% (45/125) con dos drogas y 7.2% (9/125) con 3 o más drogas. La droga más utilizada, en el 75.0% de los casos, era Enalapril.

En relación al efecto de la intervención del presente estudio sobre la adherencia terapéutica evaluada por la TPM, se observó una diferencia estadísticamente significativa en la proporción de participantes con TPM adecuada ($TPM \geq 80\%$) previa a la intervención (16.8%) respecto a la observada luego de la implementación de la misma (47.2%); $p < 0.001$.

Tanto en la Presión Arterial Sistólica (PAS) como en la Presión Arterial Diastólica (PAD) se observaron reducciones significativas en los valores de “antes” versus “después” de la intervención ($p < 0.001$ y $p < 0.001$, respectivamente). La reducción promedio en la PAS fue de 16.4 mmHg (DE 18.3) y en la PAD de 12 mmHg (DE 12.1).

Luego de la intervención, más de la mitad de los participantes llegaron a tener un nivel de presión arterial controlado (Figura 3).

Figura 3. Distribución de los niveles de Tensión Arterial según período del estudio



Por protocolo, todos los participantes del estudio debían recibir un llamado al mes 1 del ingreso al estudio y otro al mes 3-4 de seguimiento. Al mes 1 de seguimiento se pudo concretar el 92.0% de los llamados planificados (115/125). Durante su realización se detectó que un 5.2% (6/115) de los participantes no se encontraba tomando la medicación. Entre estas 6 personas, 3 refirieron que no había stock de la medicación en el CAPS y 3 que no fueron a retirar la medicación. Al mes 3-4 se pudo concretar el 89.6% de los llamados planificados (112/125). Durante su realización se detectó que un 8.0% (9/112) de los participantes no se encontraba tomando la medicación. Entre estas 9 personas, 5 refirieron no tomar la medicación por falta de stock en el centro de salud, 1 tuvo dificultades para recordar tomarla, 1 tuvo efectos no deseados de la medicación y en 2 casos no se pudo definir el motivo.

PANDEMIA COVID-19 Y DESARROLLO DEL ESTUDIO

Dado el contexto epidemiológico particular en el cual se desarrolló el estudio, resultó de interés explorar la adherencia al tratamiento en diferentes momentos de la pandemia.

Durante la intervención, el 9.9% de las personas (12/121) refirieron desarrollar enfermedad por COVID-19, de las cuales, 3 requirieron internación.

Paralelamente a la implementación del estudio tuvo lugar la segunda ola de COVID-19 en Argentina, período en el cual se registró el mayor pico de casos diarios. A causa de esto se exploró la adherencia de los pacientes antes y durante la pandemia. En la Tabla 1 se puede visualizar la TPM calculada en cuatro momentos distintos: pre-intervención pre-pandemia (desde noviembre de 2019 hasta febrero de 2020), pre-intervención intra-pandemia (abril de 2020 hasta julio de 2020), primeros 3 meses de intervención (abril a junio de 2021) y últimos 3 meses de intervención (julio a septiembre de 2021). Se puede observar que la TPM disminuye con la pandemia, pero aumenta durante los meses de intervención.

Tabla 1. TPM pre e intra pandemia y primeros y últimos 3 meses de intervención

TPM pre-pandemia	TPM intra-pandemia	TPM primeros 3 meses intervención	TPM últimos 3 meses intervención
62.47	61.97	69.70	69.97

COMENTARIOS FINALES

El presente trabajo encontró un efecto positivo de la intervención multi-componente sobre la TPM y sobre los valores de TA en personas hipertensas no controladas bajo tratamiento médico. Los componentes principales de la intervención se implementaron con éxito a lo largo de todo el estudio.

La investigación se llevó a cabo en personas con comorbilidades crónicas (diabetes, cardiovascular, oncológica o respiratoria), con características de vulnerabilidad social y en el contexto del pico más alto de casos de COVID-19, a la fecha. La población incluida contempló proporciones similares de ambos sexos. Este punto es de relevancia considerando que los hombres suelen realizar menos controles de salud y su representatividad en estudios de implementación suele ser más difícil.

El problema más importante asociado a la insuficiente TPM fue la falta de disponibilidad de medicamentos antihipertensivos en los CAPS. En estas circunstancias, los dos llamados telefónicos por protocolo durante el seguimiento de 6 meses sumado a la intervención telefónica y asignación de turnos en aquellas personas con TPM insuficiente condujo a una reducción significativa de la TA, lo que llevó a que poco más de la mitad de la población tuviera su TA controlada. Es importante destacar que el promedio de drogas antihipertensivas y tipo de drogas antihipertensivas utilizadas al final del seguimiento fue muy similar a los datos basales. Por este motivo, cabe interpretar que el mejor control de la presión arterial estaría mediado mayoritariamente por la mejoría en la adherencia a la toma de la medicación y no a otros factores, como la intensificación del tratamiento.

Este proyecto demostró cómo una intervención gatillada por el monitoreo activo de la TPM puede colaborar al control de la hipertensión arterial en población que se atiende en el primer nivel de atención del sistema público de salud.

Los resultados colaboran en la elaboración y el diseño de intervenciones más complejas con el objetivo de mejorar la adherencia terapéutica en las enfermedades crónicas; especialmente aquellas que faciliten el acceso a la medicación, incluyan indicadores de adherencia terapéutica en los sistemas de información clínica y ayuden al equipo de salud a tomar las mejores decisiones basadas en la evidencia científica.

El estudio tiene la limitación de haber sido realizado en una sola localidad, por lo cual no es posible extrapolar los resultados, por ejemplo, a los demás CAPS de la provincia. Afortunadamente, el proyecto tendrá una nueva implementación en otra localidad y se enriquecerá el análisis con los aportes de otros participantes en un contexto epidemiológico diferente.

Bibliografía

- Brown, M. T., & Bussell, J. K. (2011). Medication Adherence: WHO cares? *Mayo Clinic Proceedings*, 304-314.
- Khatib, R., & Schwalm, J.-D. (2014). Patient and healthcare provider barriers to hypertension awareness, treatment and follow up: a systematic review and meta-analysis of qualitative and quantitative studies. *PLoS One*.
- Osterberg, L., & Blaschke, T. (2005, Agosto 4). Adherence to medication. *New England Journal of Medicine*, 487-497.
- Rohina, J., & Jan, S. (2008). Global inequalities in access to cardiovascular health care: our greatest challenge. *Journal of the American College of Cardiology*, 1817-1825.
- Sabaté, E. (2003). *Adherence to Long-Term Therapies: Evidence for Action*. Geneva, : World Health Organization.
- Team, R. C. (2021). R: A language and environment for statistical computing. Viena, Austria: R Foundation for Statistical Computing.
- Tolonen, H., & Keil, U. (2005). Prevalence, awareness and treatment of hypercholesterolaemia in 32 populations: results from the WHO MONICA Project. *International Journal of Epidemiology*, 181-192.
- www.iecs.org.ar. (n.d.). Retrieved from <https://www.iecs.org.ar/institucional/quienes-somos/>
- Yusuf, S., & Rangarajan, S. (2014). Cardiovascular risk and events in 17 low-, middle-, and high-income countries. *New England Journal of Medicine*, 818-827.

ADECUACIÓN DEL MODELO DE POISSON CONDICIONAL AUTORREGRESIVO EN SERIES DE TIEMPO DE DISTINTA MAGNITUD

Lic. Noelia Carla Hegi

Directora: Mg. María Teresa Blaconá

El modelo Autorregresivo Integrado de Media Móvil (ARIMA) es un modelo estadístico de series de tiempo, que permite encontrar patrones para una predicción hacia el futuro. El modelo Poisson Condicional Autorregresivo (ACP), permite modelar sólo series de tiempo donde la variable es discreta y además presenta sobredispersión. En este trabajo se comparan los modelos ARIMA y ACP, en dos series de tiempo con datos de conteo de distinta magnitud de la variable (consultas mensuales en la guardia del Hospital de Emergencias Clemente Álvarez (HECA) por heridas de arma de fuego y por accidentes en la vía pública). De dicha comparación se deduce que en series de tiempo que presentan datos de conteo, se deben tener en cuenta las componentes de la serie como son la tendencia y la estacionalidad, como así también el cumplimiento de la distribución normal. En general se tiene una mejor aproximación a dicha distribución cuando los datos presentan mayor magnitud. En caso de que la serie presente estacionalidad y los datos no se aproximen a la distribución normal, el método de Holt-Winters (HW) puede ser una alternativa apropiada para realizar pronósticos.



INTRODUCCIÓN

Un propósito del análisis de series de tiempo es modelar el mecanismo que da lugar a la serie observada para pronosticar los valores de su comportamiento futuro, a partir de los cuales, sea posible llevar a cabo una planificación y toma de decisiones. Un modelo muy utilizado en este tipo de series es el modelo Autorregresivo Integrado de Media Móvil, también conocido como ARIMA (Box y Jenkins -1970). Dicho modelo supone que los errores se distribuyen con una distribución de Gauss, con media cero y variancia constante, supuesto que generalmente no se cumple en las series de tiempo para datos de conteo, donde una posible distribución de la variable respuesta es la de Poisson.

Las variables con datos de conteo son aquellas en las que se observa el número de veces que ocurre un evento en un determinado tiempo o espacio. La distribución Poisson tiene la característica de que su variancia es igual a su media (equidispersión), sin embargo la mayoría de los datos de conteo muestran una dispersión excesiva.

El modelo Poisson Condicional Autorregresivo (Andréas Heinen, 2003), también conocido como ACP, permite modelar series de tiempo donde la variable es discreta, presenta sobredispersión (variancia mayor que la media) y correlación serial. Se adopta un enfoque completamente paramétrico y se especifica una distribución marginal para los conteos, donde la media es autorregresiva, condicionada a observaciones pasadas. Esto permite obtener una inferencia mejorada sobre los coeficientes de regresores exógenos en relación con la regresión de Poisson estática, al tiempo que modela la correlación serial de una manera flexible.

El modelo ACP exhibe sobredispersión a pesar de que utiliza una distribución condicional equidispersa, de todos modos, las estimaciones de los parámetros de la media condicional seguirán siendo consistentes.

OBJETIVO

Comparar los modelos ARIMA y ACP cuando se trabaja con series de tiempo de datos de conteo de distinta magnitud, mediante el análisis de su comportamiento, tanto en su bondad de ajuste como al utilizarlos para realizar pronósticos.

METODOLOGÍA

A continuación, se describen brevemente ambos modelos y se reseñan las propiedades que deben cumplir.

Modelo ARIMA

En series temporales, un modelo ARIMA (*Autoregressive Integrated Moving Average*) es el que utiliza la estructura de autocorrelación de los datos para identificarlo, luego estimarlo, comprobar su bondad de ajuste y en caso de satisfacer todas las condiciones, utilizarlo para realizar pronósticos. Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes.

Las características generales que debe tener un modelo ARIMA óptimo son:

Parsimonia: Box y Jenkins ponen énfasis en que la clave de un buen modelo ARIMA es que cumpla con el principio de la parsimonia (sencillez). Se busca encontrar un modelo que se aproxime correctamente, tanto práctica como estadísticamente, al comportamiento de la serie temporal en estudio.

Residuos ruido blanco y normalmente distribuidos: Esta proposición es muy importante a la hora de verificar un modelo ARIMA, una vez realizadas las etapas de identificación y estimación del modelo. También, deben tener variancia constante y ausencia de correlación serial.

Debe proporcionar buenas predicciones: se requiere que realice predicciones satisfactorias fuera de la muestra, es decir hacia el futuro. El error porcentual medio absoluto (MAPE) mide el tamaño del error (absoluto) en términos porcentuales. Esto lo hace un indicador frecuentemente utilizado para elaborar pronósticos debido a su fácil interpretación.

Modelo Poisson Condicional Autorregresivo

El modelo ACP hace posible tratar los problemas de variables discretas con sobredispersión (variancia mayor que la media) y correlación serial. Además, dado que la autocorrelación y la densidad están modeladas explícitamente, el modelo es adecuado para pronósticos puntuales y de densidad, que pueden ser de interés en muchas aplicaciones.

El modelo ACP tiene conteos que siguen una distribución de Poisson con una media autorregresiva. La distribución de Poisson es el punto de partida natural para los datos de conteo, tiene la característica de que la media es igual a la variancia. Sin embargo, la mayoría de los datos de conteo muestran una dispersión excesiva. Modelar la media como un proceso autorregresivo genera sobredispersión incluso en el caso simple de Poisson.

ANÁLISIS EMPÍRICO

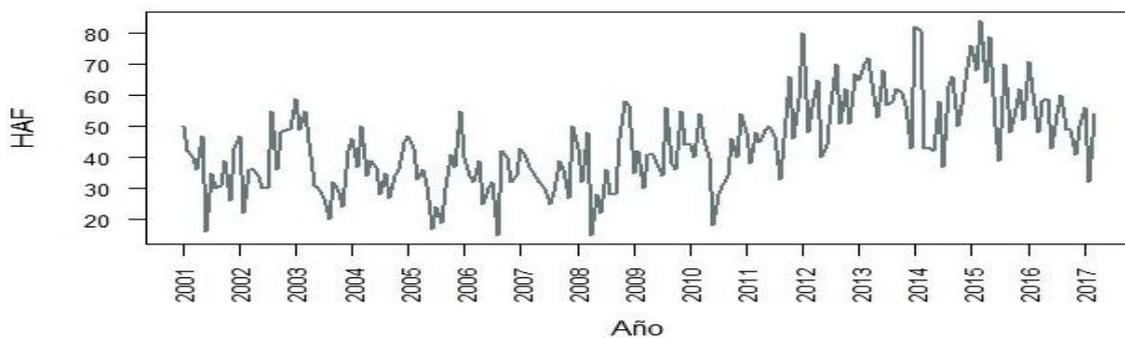
Las series por analizar son: consultas mensuales en la guardia del HECA por heridas de arma de fuego (HAF) y consultas mensuales en la guardia del HECA por accidentes en la vía pública (AVP), en el período comprendido entre enero de 2001 y marzo de 2017.

Se presenta un análisis descriptivo de las series y luego se ajustan los modelos ARIMA y ACP, como así también se utiliza el método de HW para realizar pronósticos.

Consultas mensuales en la guardia del HECA por heridas de arma de fuego

Esta serie de datos de conteo, que fluctúa aproximadamente entre las 20 y 80 consultas mensuales no presenta una tendencia marcada pero sí estacionalidad (Gráfico 1).

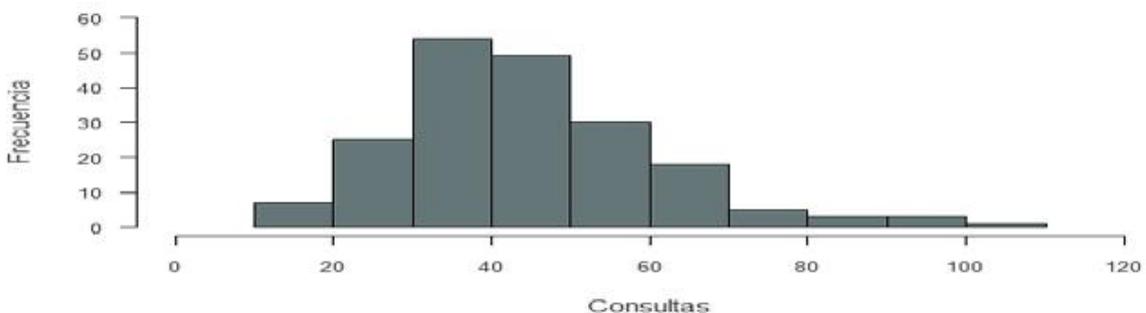
Gráfico 1. Serie de las consultas mensuales por HAF, en la guardia del HECA, ciudad de Rosario, 01/2001 – 03/2017



Fuente: Elaboración en base a datos de consultas mensuales en la guardia del HECA por HAF.

La variable analizada es una variable discreta con valores positivos y mayores a cero. Tiene una leve asimetría a la derecha, con valores extremos en dicha cola, por lo que puede considerarse una distribución Poisson (Gráfico 2). Además, la variable refiere al número de casos en un intervalo de tiempo.

Gráfico 2. Distribución de las consultas mensuales por HAF, en la guardia del HECA, ciudad de Rosario, 01/2001 – 03/2017



En primer lugar, se ajusta un modelo tradicional ARIMA y luego el modelo ACP. Se comparan los mismos y se agregan los pronósticos de un método “naive”, como es el suavizado de HW debido a que este método no requiere ningún supuesto.

Modelo ARIMA

Para esta serie se elige el modelo ARIMA (1,1,1), que cumple tener los errores no correlacionados ($Q=23.6$, $p=0.2616$) y el $AIC=1423.4$ es el más pequeño de todos los modelos probados. No se cumple la normalidad de los errores ($J-B=17.9$, $p=0.0001$), lo que era de esperar por ser justamente una variable de conteo.

Cuadro 1. Pronósticos del modelo ARIMA (1,1,1) de la serie de consultas por HAF

MES	VALOR REAL	PRONÓSTICO	MAPE	IC _{95%}	
Abr-17	37	49	24.5	27.2	70.5
May-17	44	43	13.4	21.4	65.4
Jun-17	41	38	11.6	15.3	60.0
Jul-17	38	38	8.7	15.6	61.1
ago-17	27	44	14.7	20.6	66.8
sep-17	38	42	13.8	19.1	66.0
oct-17	46	47	12.1	23.3	70.9

El Cuadro 1 muestra los MAPE un paso hacia adelante para los meses de abril a octubre de 2017. El hecho de que no se cumpla la normalidad de los errores por ser una variable de conteo, podría ser la causa de que los MAPE varíen mucho según el horizonte de pronóstico.

Modelo ACP

Para modelar los datos, se elige el modelo ACP(1,1), que si bien no es el que tiene el menor valor de AIC (8.8), tiene todos sus parámetros significativos.

Cuadro 2. Pronósticos del modelo ACP(1,1) de la serie de consultas por HAF

MES	VALOR REAL	PRONÓSTICO	MAPE
abr-17	37	50	26.0
may-17	44	50	19.0
jun-17	41	50	18.7
jul-17	38	49	19.6
ago-17	27	49	24.7
sep-17	38	49	24.3
oct-17	46	49	21.7

Los MAPE para la serie HAF son altos, hecho que se puede deber a que este modelo no contempla la estacionalidad que presenta la serie. Aun así, se trabaja con el modelo ACP, porque uno de los motivos de esta tesina es probar en qué situaciones es bueno o no utilizar estos modelos.

Estimación de los pronósticos de Holt-Winters

El método de HW, es un método automático que se utiliza para realizar pronósticos teniendo en cuenta tendencia y estacionalidad, sin realizar ningún supuesto.

Cuadro 3. Pronósticos por Holt-Winters de la serie HAF

MES	VALOR REAL	PRONÓSTICO	MAPE
abr-17	37	47	21.3
may-17	44	42	13.0
jun-17	41	36	13.3
jul-17	38	36	11.4
ago-17	27	40	15.6
sep-17	38	38	13.0
oct-17	46	41	12.9

Los valores del MAPE de HW se asemejan a los del modelo ARIMA, siendo levemente inferiores los de este último (Cuadro 1 y Cuadro 3).

Los valores del Cuadro 4 se obtienen promediando los MAPE un paso hacia adelante para los meses de abril a octubre de 2017, en el modelo ARIMA, en el modelo ACP y en HW.

Cuadro 4. Promedios de los MAPE de 7 meses para HW y los modelos ARIMA y ACP de la serie HAF

	ARIMA	ACP	HW
MAPE	14.1	22.0	14.3

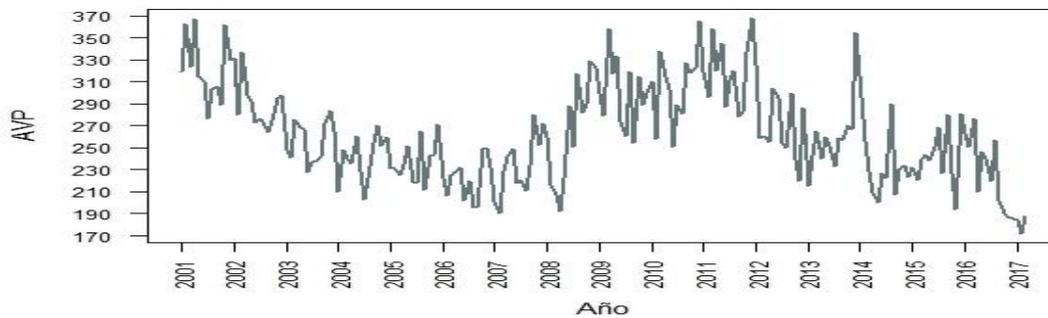
Conclusión de la serie HAF

El modelo ARIMA tiene en cuenta la estacionalidad, pero no la normalidad de la variable. Los modelos ACP no tienen en cuenta la estacionalidad y el método de HW contempla tendencia, estacionalidad y no realiza ningún supuesto. El hecho de comparar el modelo ARIMA que contempla estacionalidad, con el modelo ACP que no la contempla, es una manera de demostrar que en series con estacionalidad, los modelos ACP no son adecuados. En base a los valores del MAPE, los mejores resultados se presentan en el modelo ARIMA (MAPE=14.1), a pesar de la falta de normalidad; el modelo ACP por no tener en cuenta la estacionalidad es el que presenta los valores más grandes (MAPE=22.0) y por último el suavizado de HW a pesar de su sencillez presenta un MAPE similar al modelo ARIMA (MAPE=14.3) (Cuadro 4).

Consultas mensuales en la guardia del HECA por accidentes en la vía pública

Esta serie presenta valores que fluctúan aproximadamente entre 170 y 370 consultas mensuales. A través del tiempo, la serie presenta una baja en las consultas entre los años 2002 y 2008, luego comienzan a aumentar nuevamente marcando picos en 2009, 2011 y 2014, año en que empiezan a bajar nuevamente (Gráfico 3).

Gráfico 3. Serie de las consultas mensuales por AVP, en la guardia del HECA, ciudad de Rosario, 01/2001- 03/2017



Fuente: Elaboración en base a datos de consultas mensuales en la guardia del HECA por AVP.

El test de bondad de ajuste para una distribución normal presenta un valor de $p=0.24$. Al igual que en la serie HAF, se ajusta un modelo tradicional ARIMA y luego el modelo ACP, para luego poder realizar la comparación de los mismos, sumando al análisis los pronósticos con el método de suavizado de HW.

Modelo ARIMA

Se elige el modelo ARIMA (1,0,1) (2,0,0), que tiene un $AIC=1873.9$. A partir de este modelo se obtienen los MAPE un paso hacia adelante para los meses de abril a octubre de 2017 (Cuadro 5).

Cuadro 5. MAPE del modelo ARIMA (1,0,1) (2,0,0) de la serie AVP

MES	VALOR REAL	PRONÓSTICO	MAPE
abr-17	194	190	1.9
may-17	204	196	3.1
jun-17	199	201	2.4
jul-17	172	206	5.9
ago-17	242	202	8.7
sep-17	226	211	8.4
oct-17	273	199	12.5

Modelo ACP

Se elige trabajar con el modelo ACP(1,1) por ser el que tiene buen ajuste, menor AIC y menor cantidad de parámetros a estimar. El Cuadro 6 muestra los pronósticos a partir de dicho modelo.

Cuadro 6. Pronósticos del modelo ACP(1,1) de la serie AVP

MES	VALOR REAL	PRONÓSTICO	MAPE
abr-17	194	193	0.6
may-17	204	196	2.3
jun-17	199	199	1.6
jul-17	172	203	5.0
ago-17	242	205	7.5
sep-17	226	208	7.7
oct-17	273	211	10.

Estimación de los pronósticos de Holt-Winters

De igual manera que en la serie anterior, se realizan los pronósticos por el método de HW, mostrados en el Cuadro 7.

Cuadro 7. Pronósticos por HW de la serie AVP

MES	VALOR REAL	PRONÓSTICO	MAPE
abr-17	194	170	13.9
may-17	204	182	13.1
jun-17	199	178	12.6
jul-17	172	172	9.5
ago-17	242	183	14.0
sep-17	226	171	17.0
oct-17	273	168	23.6

Los valores del Cuadro 8 se obtienen promediando los MAPE un paso hacia adelante para los meses de abril a octubre de 2017, en el modelo ARIMA, en el modelo ACP y en HW, presentados en los Cuadros 5, 6 y 7.

Cuadro 8. Promedios de los MAPE de 7 meses para HW y los modelos ARIMA y ACP de la serie AVP

	ARIMA	ACP	HW
MAPE	6.1	5.1	14.8

Conclusión de la serie AVP

Considerando el MAPE, los mejores resultados se obtienen con el modelo ACP (MAPE=5.1), aunque no existe mucha diferencia con el modelo ARIMA (MAPE=6.1). Para esta serie el método de HW no obtiene buenos resultados (MAPE=14.8), lo cual se puede deber a que no se realiza ningún supuesto (Cuadro 8).

CONCLUSIONES

Este trabajo es la continuación del realizado en la tesina de la Licenciatura en Estadística (UNR) de Lucía Angeloni (2016). En la misma se realizó el desarrollo teórico de los modelos ACP, dejando planteado considerar ese trabajo como punto de partida para el análisis de distintas bases de datos a fin de evaluar el comportamiento de los modelos en distintos contextos.

Para continuar con esa línea, se utilizan las series de las consultas mensuales en la guardia del HECA por HAF y por AVP, para realizar la comparación entre el modelo ARIMA y el modelo ACP. Las series antes mencionadas tienen distinta magnitud en cuanto a la variable de datos de conteo y por lo tanto distinta aproximación a la normal, siendo HAF la serie de menor magnitud.

El modelo ARIMA permite describir un valor como una función lineal de datos anteriores y errores debidos al azar, además, puede incluir un componente estacional. Una vez que se ha obtenido un modelo adecuado se realizan predicciones con el mismo.

El modelo ACP hace posible tratar los problemas de variables con sobredisperción y correlación serial. Además, dado que la autocorrelación y la densidad están modeladas explícitamente, el modelo es adecuado para pronósticos puntuales y de densidad, que pueden ser de interés en muchas aplicaciones. Cuenta con la desventaja de que puede ser sólo aplicado a series de tiempo cuyas observaciones se encuentren correlacionadas positivamente.

Para poder obtener una conclusión en base a cuándo conviene utilizar cada uno de estos modelos, se realiza en cada serie la comparación de los modelos ARIMA y ACP.

En la serie HAF, la de menor magnitud, el modelo ACP presenta un mejor ajuste en cuanto al análisis de residuos, aunque teniendo en cuenta el MAPE, el modelo ARIMA es el que arroja mejores resultados para realizar pronósticos. Esto puede deberse a que el modelo ACP no tiene en cuenta la estacionalidad. En este caso, también se aplicó el método de HW, método automático que contempla estacionalidad, presentando pronósticos con un MAPE a 7 meses muy similar al modelo ARIMA, por lo que en estos casos en que los modelos ACP no contemplan la estacionalidad y en los modelos ARIMA no se cumple la normalidad, puede ser una alternativa sencilla para realizar pronósticos.

Por el contrario, en la serie AVP (la de mayor magnitud) el modelo ARIMA es el que presenta mejor ajuste en cuanto a su análisis de residuos, verificándose la normalidad de los mismos. En este caso no

hay mucha diferencia en el MAPE entre los modelos ARIMA y ACP, siendo el modelo ACP el que presenta menor valor. El método de HW tiene resultados muy pobres en cuanto a pronósticos, por lo que se pueden utilizar tanto el modelo ARIMA como el ACP.

De este análisis empírico se puede recomendar que, en series de tiempo que presentan datos de conteo, se deben tener en cuenta las componentes de la serie como son la tendencia y la estacionalidad, como así también el cumplimiento de la distribución normal, que tiene una mejor aproximación a dicha distribución cuando los datos tienen una mayor magnitud. En caso de que la serie presente estacionalidad y por la magnitud de los datos no se aproximen a la distribución normal, el método de HW puede ser una alternativa apropiada para realizar pronósticos.

Por lo expuesto se puede concluir que, para cada serie de conteo, es conveniente tener en cuenta las componentes presentes en la serie, como tendencia y estacionalidad, realizar una revisión del cumplimiento de supuestos y no descartar utilizar el método de HW para series de magnitud pequeña de la variable.

Bibliografía

Angeloni, Lucía Soledad (2016). "Modelo Poisson Condicional Autorregresivo para series de tiempo con datos de conteo". Tesina de la Licenciatura en Estadística, UNR.

Box y Jenkins (1970). "*Time Series Analysis, Forecasting and Control*". Wiley.

Heinen, Andrés (2003). "Modelling Time Series Count Data: An Autoregressive Conditional Poisson Model", Online <http://mpa.ub.uni-muenchen.de/8113/> MPRA Paper No. 8113, posted 7. April 2008 00:28 UTC

Orange Matter (2019). Holt-Winters Forecasting and Exponential Smoothing Simplified. [https://orangematter.solarwinds.com/2019/12/15/holt-winters-forecasting-simplified/#:~:text=Holt%2DWinters%20is%20a%20model,cyclical%20repeating%20pattern%20\(seasonality\).](https://orangematter.solarwinds.com/2019/12/15/holt-winters-forecasting-simplified/#:~:text=Holt%2DWinters%20is%20a%20model,cyclical%20repeating%20pattern%20(seasonality).)

COMPARACIÓN DE ANÁLISIS INTERMEDIOS EN ENSAYOS CLÍNICOS. ESTUDIO PARA EL TRATAMIENTO DE PACIENTES HOSPITALIZADOS POR COVID19

Lic. Franco Iván Scarafia

Responsable de la FCEyE: **Lic. Mara Catalano**

Responsable de Estudios Clínicos Latinoamérica: **Lic. Noelia Castellana**

Durante el desarrollo de un ensayo clínico es común realizar un monitoreo sobre los datos a través de los denominados análisis intermedios, que realizan evaluaciones parciales de la información obtenida. Una decisión no menor recae en elecciones referidas a estos análisis intermedios, tales como la cantidad a realizar y los métodos que se utilizarán.

En este informe se busca estudiar y comparar distintos procedimientos de análisis intermedios para evaluar tempranamente los eventos de un ensayo clínico. Dicha comparación se lleva a cabo bajo escenarios hipotéticos del ensayo COLCOVID conducido por una organización de investigaciones clínicas, que buscaba evaluar los efectos de la Colchicina con el objetivo de reducir el requerimiento de ventilación mecánica y/o la mortalidad.

Para la comparación de los métodos de evaluación de análisis intermedios se seleccionaron cinco escenarios según la cantidad de pacientes enrolados o el tiempo transcurrido entre análisis.

En lo que respecta a los resultados de los análisis intermedios hipotéticos de COLCOVID se pudo observar cómo algunos procedimientos, bajo ciertos escenarios, hubieran llevado a pensar en una posible interrupción temprana, causada por un potencial efecto benéfico de la Colchicina respecto a la reducción del tiempo hasta el evento compuesto "Intubación o Muerte".



INTRODUCCIÓN

Este informe se corresponde con la labor realizada durante una Práctica Profesional de la Licenciatura en Estadística de la Universidad Nacional de Rosario realizada durante el año 2021 en la organización Estudios Clínicos Latinoamérica (ECLA), una entidad de Investigación Académica dedicada a brindar apoyo y experiencia para ensayos clínicos multicéntricos internacionales. ECLA otorga soporte en las etapas de planificación, coordinación y gestión con orientación académica y facilita la puesta en marcha y operación de ensayos clínicos para así obtener investigaciones clínicas de alta calidad.

La conducción de un ensayo clínico conlleva un sinnúmero de decisiones, procesos y responsabilidades para poder arribar a resultados fiables y válidos para los tratamientos que se desean evaluar. Durante el desarrollo del estudio es común realizar un monitoreo sobre los datos a través de los denominados análisis intermedios, que realizan evaluaciones parciales de la información obtenida. Una decisión no menor a la hora de llevar adelante un ensayo clínico recae en elecciones referidas a estos análisis intermedios, tales como la cantidad a realizar y los métodos que se utilizarán para su evaluación.

En este informe se busca estudiar y comparar distintos procedimientos de análisis intermedios para evaluar tempranamente los eventos de un ensayo clínico. Dicha comparación se lleva a cabo bajo escenarios hipotéticos del ensayo clínico COLCOVID conducido por ECLA y el Population Health Research Institute (PHRI) de Hamilton, Canadá. Este estudio, realizado en plena pandemia de COVID19, buscaba evaluar si la Colchicina (agente inmunomodulador y antiinflamatorio de bajo costo y efectos secundarios leves) producía una posible reducción de los requerimientos de nuevas intubaciones y/o de la mortalidad en pacientes hospitalizados con COVID19, comparada con el tratamiento que se estaba utilizando para estos pacientes.

OBJETIVOS

Comparar diferentes metodologías de análisis intermedios en ensayos clínicos, utilizando el estudio COLCOVID para simular escenarios de acuerdo a variaciones en los posibles tamaños muestrales y en los tiempos de espaciamentos entre análisis intermedios.

METODOLOGÍA

Un ensayo clínico es un estudio de investigación en el que uno o más sujetos humanos se asignan prospectivamente a una o más intervenciones (que pueden incluir un placebo u otro control) para

evaluar los efectos de las mismas en resultados biomédicos o conductuales relacionados con la salud.

La sensibilidad de la información que se maneja en ciertos ensayos clínicos, así como la de las consecuencias que pueden traer sus resultados, hace necesaria la participación de un “comité de monitoreo” interdisciplinario e independiente que asesore al grupo de investigadores que lleva a cabo el ensayo. Dentro de sus principales responsabilidades se encuentran: (1) revisar y evaluar periódicamente los datos acumulados del estudio en relación a la seguridad y eficacia y (2) hacer recomendaciones al grupo de investigadores con respecto a la continuación, modificación o terminación del ensayo. Para ello se basan en los resultados que arrojan los análisis intermedios de los datos acumulados en forma periódica.

La necesidad de tener que realizar análisis intermedios en un ensayo clínico genera un problema fundamental a la hora de realizar las correspondientes pruebas de hipótesis sobre distintos aspectos de la variable respuesta, ya que el nivel de significación aumenta al realizar pruebas repetidas sobre datos que se acumulan en el tiempo. Es decir, por el solo hecho de evaluar la misma hipótesis reiteradas veces, eventualmente se podría rechazar la misma por consecuencia del azar, lo que equivale a decir que la probabilidad de cometer un error de tipo I (α) será mayor a la establecida previo al inicio del ensayo.

Esto, sumado a razones tanto éticas como administrativas y económicas para realizar análisis intermedios, ha llevado a que se desarrollen métodos específicos para su evaluación y que dichos métodos se hayan ido mejorando a través del tiempo.

Una de las principales dificultades a la hora de realizar un ensayo clínico es la decisión de qué método o enfoque se utilizará para llevar adelante los análisis intermedios ya que éste debe ser definido previamente al inicio del reclutamiento, cuando se tiene poca o nula información sobre el comportamiento de la variable respuesta de interés.

Métodos de grupos secuenciales

Los métodos de grupos secuenciales surgen a partir de procedimientos más primitivos y, si bien salvan algunas de las restricciones de estos, igualmente presentan dos grandes limitaciones que condicionan su utilización: la necesidad de definir de forma anticipada el número de análisis intermedios que se van a realizar y que la cantidad de individuos que se incorporan entre cada análisis sea constante (o se aproxime a serlo). Se comparan tres de estos métodos: Haybittle-Peto, Pocock, y O'Brien-Fleming.

El método de Haybittle-Peto, quizás el más sencillo entre los métodos de grupos secuenciales, consiste simplemente en recurrir a un valor crítico normalizado de las pruebas de hipótesis lo suficientemente grande, como ± 3 por ejemplo, para que la región de rechazo de los análisis intermedios sea pequeña y, por ende, la probabilidad de error de tipo 1 sea baja. De esta forma, el ajuste necesario por testeos repetidos para el análisis final puede ser tomado como despreciable y se utiliza el valor crítico convencional (por ejemplo, 1.96 si se fija una cota de probabilidad de error de tipo 1, α , de 0.05) sin temor a que aumente la probabilidad de error de tipo 1. Una característica es que no requiere que la cantidad de individuos que se incorpora entre cada análisis sea constante.

El método de Pocock define un valor K de análisis a realizar (intermedios y final) y divide a los participantes esperados en K grupos de tamaño n (con $n/2$ para cada uno de los dos grupos, intervención y control). A medida que se van obteniendo los datos de cada grupo de n individuos, se computa Z_i ($i = 1, \dots, K$), la estadística normalizada asociada al test sobre la hipótesis nula del ensayo realizada en el i -ésimo análisis, y se la compara contra $\pm Z_{N_i}$, límites de rechazo para dicha estadística normalizada, constantes para todos los análisis y definidos de forma tal que la probabilidad de error de tipo 1 total sea el valor preespecificado α .

Si en el i -ésimo test la estadística observada cae por fuera de dichos límites, se procede a rechazar la hipótesis nula y se podría llegar a interrumpir el estudio. Por el contrario, si esto no ocurriese, se continúa de igual forma hasta el K -ésimo test y si dicha decisión se mantiene, no se rechaza H_0 .

El método de O'Brien-Fleming compara la estadística Z_i con el valor crítico $Z^* * \sqrt{K/i}$ con $i = 1, \dots, K$ donde Z^* se determina de forma tal que se logre el nivel de significación total deseado. Una característica a destacar de este método es que el valor crítico utilizado en el test final resultará cercano al que se debería utilizar si no se hicieran análisis intermedios y se realizara un solo test. No obstante, la restricción de incorporar igual cantidad de participantes en cada análisis sigue vigente y nuevamente se deberá dividir a los participantes en K grupos de tamaño n , llegando a un tamaño muestral total esperado de $N = Kn$.

Tanto Pocock como O'Brien-Fleming tienen sus valores críticos ya tabulados para ciertos valores frecuentes del nivel de significación total y cantidad de análisis intermedios a realizar.

Funciones de gasto de alfa

El método denominado "función de gasto de alfa", propuesto por Lan y DeMets en 1983, permite a los investigadores elegir "cómo gastar" la probabilidad de error de tipo 1 durante el desarrollo del ensayo. Este procedimiento garantiza un α final total igual al valor preestablecido y, además, permite evadir las restricciones propias de los métodos de grupos secuenciales. Existen funciones

de gasto de alfa que se aproximan a métodos de grupos secuenciales tales como O'Brien-Fleming o Pocock.

Este procedimiento introduce el concepto de "fracción de información" (t^*), que no debe confundirse con "tiempo de calendario" (t), y que se refiere a la "proporción de información" disponible en el tiempo t . Se calcula generalmente como la proporción entre el número de pacientes n ya aleatorizados respecto al total N esperado o bien, como la fracción entre el número de eventos observados d , en relación al total D de eventos esperados.

Las "funciones de gasto de alfa", $\alpha(t^*)$, entonces serán todas aquellas funciones que cumplan: (1) Al inicio del estudio valen cero [$t^* = 0 \rightarrow \alpha(t^*) = 0$], (2) Al final del estudio valen alfa [$t^* = 1 \rightarrow \alpha(t^*) = \alpha$] y (3) Son siempre crecientes [$\alpha(t_1^*) \leq \alpha(t_2^*) \forall t_1^* \leq t_2^*$].

Por lo tanto, infinitas funciones de gasto de alfa pueden ser definidas siempre y cuando cumplan con dichas condiciones. Incluso se pueden establecer funciones de gasto de alfa que se aproximen a los métodos ya mencionados, como Pocock y O'Brien-Fleming, u otras más utilizadas en la práctica como la función de gasto de alfa uniforme, de la forma $\alpha(t^*) = \alpha \times t^*$.

Enfoque de potencia condicional

Un enfoque diferente a lo planteado en los otros métodos es el de potencia condicional o reducción estocástica. Mientras que los anteriores consideran solo la información ya recolectada para tomar decisiones, este método tiene en cuenta no solo los datos que ya se obtuvieron sino también los aún no recolectados, para evaluar la terminación temprana de un estudio.

Es decir, lo que se busca es resolver la pregunta "dados los resultados actuales del ensayo, ¿qué tan probable es que las conclusiones actuales cambien si este sigue su curso hasta su finalización programada?". Y, por lo tanto, si fuera "muy probable" que los resultados a recolectar no cambien el curso de las decisiones, entonces el enfoque recomendaría finalizar el estudio de forma temprana.

Para poder explicar la reducción estocástica se deben definir primero la región de rechazo R , la región de aceptación A y retomar el concepto de fracción de información t^* ya utilizado en la función de gasto de alfa.

Además, siendo $Z(t^*)$ la estadística normalizada asociada al test sobre la hipótesis nula del ensayo, Z , aplicada en un análisis donde se cuenta con la fracción de información t^* , se definen dos probabilidades asociadas:

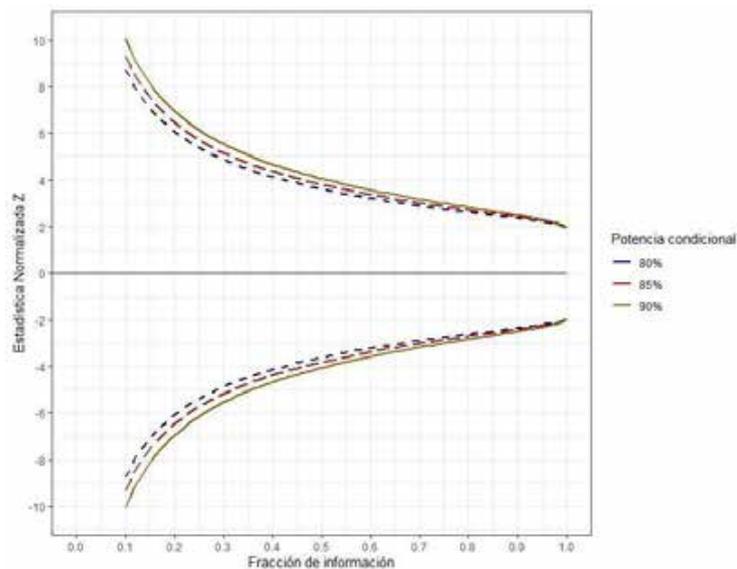
- $P(Z(1) \in R / Z(t^*), H_0) = \gamma$ que es la probabilidad de rechazar H_0 al finalizar el ensayo dados la fracción de información t^* y asumiendo cierta la hipótesis nula H_0
- $P(Z(1) \in A / Z(t^*), H_1) = \gamma'$ que es la probabilidad de no rechazar H_0 al finalizar el ensayo dada la fracción de información t^* y asumiendo cierta la hipótesis alternativa H_1

La elección de estas probabilidades dependerá del criterio de los investigadores. Usualmente γ varía entre 50% y 90% mientras que γ' se ubicará entre 0% y 20% para la mayoría de los casos.

La curva de potencial condicional que define los límites de los valores críticos se calcula a través de una fórmula que depende de $Z(t^*)$, de α y de β (error tipo II). Además, dependerá también de γ , si se quiere definir la curva bajo H_0 , o de γ' , si se desea en cambio la curva bajo H_1 . En la práctica, se acostumbra a utilizar solo la curva bajo H_0 o combinada con una curva bajo H_1 .

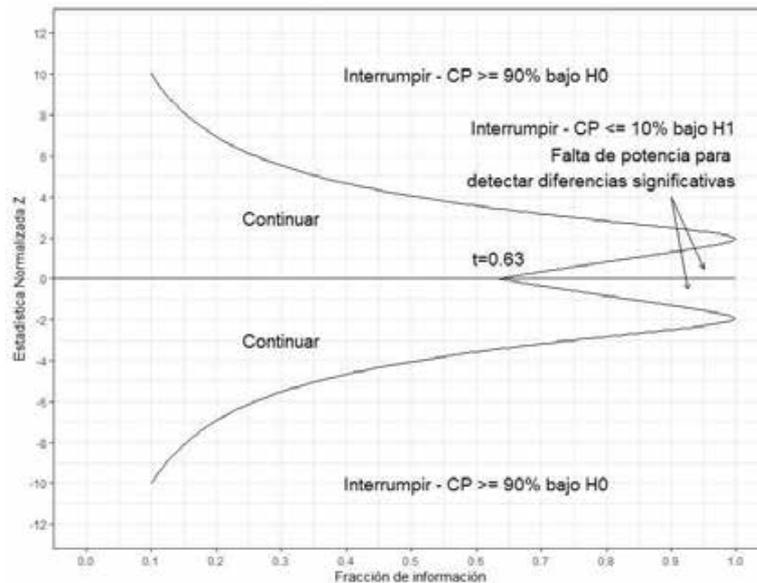
Si se utiliza solo la curva bajo H_0 y se piensa en una hipótesis alternativa bilateral, entonces se obtienen límites como los de la Figura 1 para valores de γ en 80%, 85% o 90%. A la hora de tomar una decisión durante la realización de un análisis intermedio, se calcularía la $Z_{\text{obs}}(t^*)$ y se la compararía con los límites correspondientes. Si fuera mayor al límite superior se debería finalizar el estudio por efecto benéfico, mientras que si fuera menor al límite inferior se debería finalizar por efecto nocivo. Finalmente, si se encontrara entre ambos límites la recomendación sería continuar el ensayo.

Figura 1: Límites de la estadística normalizada según fracción de información para potencias condicionales de 80%, 85% y 90%



Por otro lado, si se utiliza una curva de potencia condicional bajo H_0 (por ejemplo, de 90%) combinada con una curva bajo H_1 (supongamos de 10%) y, si además se definen los parámetros $\alpha = 0.05$ y $\beta = 0.10$, se obtienen límites como los de la Figura 2.

Figura 2: Estadística normalizada según fracción de información para potencia condicional bajo H_0 de 90% y bajo H_1 de 10%



Si bien los “sectores” definidos por las curvas son muy similares a los del caso anterior, se suma una nueva “área” que marca la falta de potencia para detectar diferencias significativas, también conocido como “futilidad”. Sin embargo, esta decisión solo aparece como una posibilidad a partir de cierto momento del estudio (en el ejemplo, a partir de $t^* = 0.63$).

ESTUDIO COLCOVID

El estudio COLCOVID (NCT04328480, 2020) es un ensayo clínico controlado, abierto, aleatorizado, simple y pragmático para evaluar los efectos de la Colchicina con el objetivo de reducir el requerimiento de ventilación mecánica y/o la mortalidad.

Se comparan dos tratamientos: tratamiento usual o estándar (indicado como “rama tratamiento usual”) contra el tratamiento estándar más un agregado de Colchicina (indicado como “rama Colchicina”). Además, previo a la aleatorización de los dos tratamientos se estratificó a los pacientes en intubados o no, al inicio del estudio.

El ensayo COLCOVID enroló a un total de 1279 pacientes hospitalizados por COVID19 en diferentes sanatorios de Argentina. El tiempo de enrolamiento del estudio comprendió el período entre el 17

de abril de 2020 y el 28 de marzo de 2021, es decir un total de 345 días. El reclutamiento de pacientes de este estudio presentó un crecimiento leve durante los primeros cinco meses, habiéndose incorporado menos de 200 pacientes. Luego, se produjo un incremento en la tasa de enrolamiento que se sostendría hasta el final del ensayo.

El evento primario de interés se definió como el tiempo hasta que el paciente presentara requerimiento de intubación o muerte durante los 28 días de seguimiento. En caso de que el participante ya estuviera intubado al momento de ingresar al ensayo, el evento solo podía ocurrir si fallecía. Por lo tanto, se podría considerar que el evento refleja un “empeoramiento” del paciente ya sea a través de una intubación, cuando no la tuviera ya, o con la muerte.

La aleatorización ubicó a 640 pacientes (50.04%) en la “rama Colchicina” mientras que a otros 639 (49.96%) los asignó a la “rama tratamiento usual”.

El primer análisis que se realizó fue el de características basales que comparó variables referidas al sexo, edad, antecedentes médicos varios y uso de medicamentos, entre otras, a fin de evaluar el balanceo de los grupos respecto a ellas.

A fin de estudiar el comportamiento de la variable respuesta de interés a partir del tratamiento asignado, se ajustó un modelo de regresión de Cox de riesgos proporcionales estratificando por el estado de intubación al momento de la aleatorización. A partir de dicho modelo se estimó la razón de *hazards* (relativa de que ocurra un evento en la “rama Colchicina” del ensayo en comparación a la “rama tratamiento usual”).

Los 1279 pacientes reclutados en 42 hospitales y sanatorios de Argentina presentaban una media etaria de 61.8 años (DE = 14.6 años) y estaban conformados por 449 mujeres (35.1%) y 830 hombres (64.9%). Se observó que las características basales estaban bien balanceadas entre ambos grupos.

En lo que respecta al evento de interés de intubación o muerte durante los 28 días que duró el seguimiento, éste ocurrió en 184 pacientes del grupo control (28.8%) y en 160 del grupo intervención (25.0%). La razón de *hazards* resultó de 0.83 (0.67 – 1.02) y su valor p asociado fue de 0.08 por lo que se puede decir que la intervención de Colchicina redujo en 17% el riesgo de presentar intubación o muerte, aunque dicha reducción no es estadísticamente significativa al 5%.

APLICACIÓN

Se utilizaron para este estudio los 1279 pacientes reclutados en el ensayo COLCOVID y se determinaron cinco escenarios para evaluar los procedimientos de análisis anteriormente presentados. De estos escenarios, cuatro son hipotéticos y uno corresponde al caso real utilizado en el estudio.

Los escenarios simulados fueron creados utilizando el tamaño muestral total esperado N y dividiéndolo en K partes iguales según la cantidad de pacientes n que se quiere reclutar entre cada análisis intermedio, o bien definiendo una cierta cantidad prefijada de tiempo entre análisis.

De esta forma, los cinco escenarios que se evaluaron fueron: (1) análisis intermedios cada 400 pacientes, (2) análisis intermedios cada 300 pacientes, (3) análisis intermedios cada 4 meses, (4) análisis intermedios cada 3 meses, (5) escenario real, con un análisis con al menos 50% de los pacientes y otro con al menos 75%.

Para cada uno de estos escenarios, se compararon los resultados que se hubieran obtenido en caso de aplicar 6 diferentes procedimientos: (1) método de Haybittle-Peto, (2) método de Pocock, (3) método de O'Brien-Fleming, (4) función de gasto de alfa uniforme, (5) curva de potencia condicional bajo H_0 de 80%, y (6) curva de potencia condicional bajo H_0 de 90% junto con curva de potencia condicional bajo H_1 de 10%.

Cabe aclarar, no obstante, que los primeros tres procedimientos mencionados caen dentro de los llamados métodos de grupos secuenciales y que, como tales, no pudieron ser aplicados en los escenarios (3) y (4) pues no cumplen con la condición necesaria de incrementos similares de muestra o eventos entre análisis por haber sido definidos en base a cantidades prefijadas de tiempo.

Luego de aplicar los métodos a los escenarios planteados se resumen los resultados en la Tabla 1. Se puede ver cómo en pocos casos alguno de los métodos hubiera sugerido una interrupción temprana, siendo la función de gasto de alfa uniforme la más propensa a hacerlo. Incluso para algunos escenarios, como el análisis cada 400 pacientes o cada 3 meses, no hubo ningún procedimiento para el cual el Z observado cayera por fuera de sus límites.

Tabla 1. Resultados obtenidos con cada método según escenario

Método/Escenario	Análisis cada 400 pacientes	Análisis cada 300 pacientes	Análisis cada 4 meses	Análisis cada 3 meses	Escenario real
Haybittle-Peto	NI	NI	NI	NI	NI
Pocock	NI	I2	NA	NA	NA
O'Brien-Fleming	NI	NI	NA	NA	NA
Func. de gasto de alfa Uniforme	NI	I2	I2	NI	I1
Pot. Cond. de 80%	NI	NI	NI	NI	NI
Pot. Cond. bajo H_0 de 90% y bajo H_1 de 10%	NI	NI	NI	NI	NI

I1: Interrumpe en el 1er análisis intermedio

I2: Interrumpe en el 2do análisis intermedio

NA: No aplicable

NI: No interrumpe

COMENTARIOS FINALES

Si bien la experimentación en general y los ensayos clínicos en particular tienen considerables años de existencia y desarrollo, los procedimientos para la realización de análisis intermedios son relativamente nuevos si se los compara con dichos saberes. Por ello es natural que, aunque existan métodos ya desarrollados y documentados, el área siga bajo avance continuo.

Uno de los objetivos tácitos de esta Práctica Profesional era poder crear un compilado que sirva de utilidad para un investigador que se enfrente a una situación donde deba escoger cuántos análisis intermedios realizar en su ensayo clínico y qué método utilizar en los mismos.

En lo que respecta a los resultados de los análisis intermedios hipotéticos se pudo observar cómo algunos procedimientos, bajo ciertos escenarios, hubieran llevado al Comité a pensar en una posible interrupción temprana, causada por un potencial efecto benéfico de la Colchicina respecto a la reducción del tiempo hasta el evento compuesto "Intubación o Muerte". En todos esos casos, la función de gasto de alfa uniforme y/o el método de Pocock eran los métodos cuyos límites no

incluían al valor Z observado. Esto se debe a que ambos no solo presentan límites siempre muy similares entre sí, sino que en general son los menos conservadores de todos los procedimientos presentados y, por ende, se dejan “engañar” por la tendencia emergente que se dio para este ensayo durante sus primeras etapas de reclutamiento.

Quizás el punto más importante para enfatizar es que en el escenario real un simple cambio de elección respecto al método a utilizar hubiera puesto al Comité de Monitoreo en una situación de decisión propensa a la interrupción temprana por beneficio en un ensayo que resultó demostrar que no había diferencias significativas entre los tratamientos.

Bibliografía

Davis, B. R., & Hardy, R. J. (1994). *Data monitoring in clinical trials: The case for stochastic curtailment*. *Journal of Clinical Epidemiology*. [https://doi.org/10.1016/0895-4356\(94\)90119-8](https://doi.org/10.1016/0895-4356(94)90119-8)

Ellenberg, S., Fleming, T. & DeMets, D. (2002). *Data monitoring Committees in Clinical Trials. A Practical Perspective*. John Wiley and Sons

Friedman, L. M., Furberg, C., & Demets, D. L. (2010). *Fundamentals of clinical trials*. Springer.

Gernot Wassmer, & Brannath, W. (2016). *Group sequential and confirmatory adaptive designs in clinical trials*. Springer.

Jennison, C., & Turnbull, B. W. (2000). *Group sequential methods with applications to clinical trials*. Chapman & Hall/Crc.

NCT04328480. U.S. National Library of Medicine. (2020) *Anti-Coronavirus Therapies to Prevent Progression of Coronavirus Disease 2019 (COVID-19) Trial (ACTCOVID19)* Disponible en: <https://clinicaltrials.gov/ct2/show/NCT04324463>. Fecha de consulta: 07-06-2022

O'Brien, P. C., & Fleming, T. R. (1979). *A Multiple Testing Procedure for Clinical Trials*. *Biometrics*. doi:10.2307/2530245

Pocock, S. J. (1977). *Group sequential methods in the design and analysis of clinical trials*. *Biometrika*. doi:10.1093/biomet/64.2.191

TÉCNICAS GRÁFICAS PARA LA EVALUACIÓN DE SUPUESTOS EN MODELOS LINEALES MIXTOS PARA DATOS LONGITUDINALES

Lic. Marlén Altamirano

Directora: **Lic. Cecilia Rapelli**

Codirectora: **Lic. Mara Catalano**

La etapa de diagnóstico es un proceso importante que permite evaluar el ajuste del modelo propuesto a un conjunto de datos y verificar el cumplimiento de los supuestos postulados sobre el mismo. El análisis de residuos se utiliza con frecuencia como herramienta de diagnóstico, mediante la realización de gráficas que permitan comprobar el cumplimiento de los supuestos del modelo.

Los métodos de diagnóstico para los modelos lineales mixtos para datos longitudinales, a diferencia de los modelos lineales clásicos, son más difíciles de realizar e interpretar dado que el modelo es más complejo. Esa complejidad se debe a la presencia de efectos aleatorios y a las diferentes estructuras de covariancias, que permiten modelar las distintas fuentes de variabilidad presentes en los datos longitudinales. Debido a estas fuentes de variabilidad, el ajuste de modelos lineales mixtos lleva a definir tres tipos de residuos.

En esta tesina se presenta un compendio de técnicas diagnósticas gráficas basadas en los distintos tipos de residuos que se desprenden de los modelos lineales mixtos y se describe su uso en el proceso de selección del modelo, en la evaluación del cumplimiento de los supuestos y en la identificación de posibles valores atípicos.



INTRODUCCIÓN

Los estudios basados en datos longitudinales son frecuentemente utilizados en una amplia variedad de disciplinas. Éstos consisten en medir repetidamente la variable respuesta de interés a cada uno de los individuos o unidades a través del tiempo, permitiendo caracterizar el cambio en la respuesta y los factores que influyen dicho cambio.

Una característica distintiva de los estudios longitudinales es que las mediciones realizadas sobre una misma unidad están correlacionadas positivamente, lo cual debe ser tenido en cuenta en el análisis de los datos para obtener inferencias válidas.

Para describir esas respuestas longitudinales son convenientes los modelos lineales mixtos, ya que permiten considerar las múltiples fuentes de correlación presentes en este tipo de datos.

El proceso de selección del modelo lineal mixto apropiado está compuesto por reiterados ajustes y evaluaciones de modelos con el objetivo de identificar una forma funcional que describa adecuadamente la evolución de la respuesta en el tiempo, su relación con las covariables y la correlación presente en los datos. Durante este proceso, se pueden utilizar diversas herramientas diagnósticas basadas en los residuos.

Debido a las diferentes fuentes de variabilidad presentes en los datos longitudinales se pueden definir distintos tipos de residuos. Cada uno de ellos es de utilidad para evaluar algún supuesto del modelo.

MODELO LINEAL MIXTO

Los modelos lineales mixtos son útiles para analizar datos con mediciones repetidas, en particular, datos longitudinales debido a su capacidad para explicar las múltiples fuentes de variación y correlación presentes en este tipo de datos de manera parsimoniosa a través de la introducción de efectos aleatorios.

Una característica de los modelos mixtos es que permiten modelar la respuesta media como una combinación de características poblacionales, que son compartidas por todas las unidades, y un conjunto de efectos específicos que son únicos para cada unidad. Los primeros se denominan efectos fijos, mientras que los últimos reciben el nombre de efectos aleatorios. El término mixto se utiliza para mostrar que el modelo contiene una combinación de ambos efectos.

Por otra parte, la distinción entre efectos fijos y aleatorios en estos modelos no solo permite estimar los parámetros que describen el cambio de la respuesta media en el conjunto de datos bajo estudio, sino que es posible predecir cómo cambian las trayectorias individuales en el tiempo.

DIAGNÓSTICOS PARA MODELOS LINEALES MIXTOS

Al igual que en los modelos lineales clásicos, en los modelos lineales mixtos se realiza un análisis de residuos para evaluar el cumplimiento de los supuestos. Este proceso, conocido como etapa de diagnóstico, consiste en evaluar el ajuste del modelo propuesto a un conjunto de datos y verificar el cumplimiento de los supuestos postulados sobre el mismo. Sin embargo, este análisis es más difícil de realizar e interpretar debido a la complejidad de los modelos mixtos, causada por la presencia de efectos aleatorios y las diferentes estructuras de covariancia.

La presencia de diferentes fuentes de variación en los modelos lineales mixtos (variabilidad intra y entre unidades), lleva a definir tres tipos de residuos: marginales, condicionales y de efectos aleatorios. Cada uno de ellos es de utilidad para evaluar algún supuesto del modelo. Los marginales se usan para estudiar la linealidad de la respuesta con respecto a las variables explicativas y para evaluar si la estructura de covariancias intra unidad es adecuada; los condicionales para comprobar la homocedasticidad y normalidad del error condicional y los residuos de efectos aleatorios para verificar la normalidad de los efectos aleatorios.

El uso de estas herramientas diagnósticas basadas en los residuos en el proceso de construcción del modelo es fundamental debido a que permite conocer mejor el problema bajo investigación y así sugerir modelos adecuados.

Estructura de covariancia

La correcta especificación de la estructura de covariancia intra unidad permite obtener estimaciones e inferencias válidas. Verificar si esta estructura fue adecuadamente seleccionada es un paso no trivial en el análisis. Puesto que todas las herramientas de diagnóstico dependen de su especificación, el proceso de diagnóstico debe comenzar por la evaluación de este supuesto.

En este sentido, Lesaffre y Verbeke (1998) propusieron el uso de los residuos de covariancia (índice de Lesaffre y Verbeke). De esta manera se obtiene un conjunto de residuos con variancia unitaria, aunque algo correlacionados, debido a que están transformados por la inversa de la raíz cuadrada de la matriz de covariancia de los datos.

Para subsanar este inconveniente, Singer *et al.* (2016) proponen una modificación en estos últimos, reemplazando los residuos marginales por los residuos marginales transformados, que se obtienen dividiendo al residuo por la raíz cuadrada de su matriz de covariancia.

Si el modelo asumido para la media es correcto, el gráfico de los residuos de covariancias modificados versus el índice de la unidad permite detectar si la estructura de covariancia asumida es apropiada o debe modificarse. Si la estructura de covariancia es adecuada, los valores de los residuos de covariancia modificados serán cercanos a cero.

Por otra parte, Fitzmaurice *et al.* (2004) y Pinheiro y Bates (2000) propusieron el uso del variograma, utilizando en su construcción los residuos marginales transformados mediante la descomposición de Cholesky. Estos residuos son aproximadamente no correlacionados y con variancia unitaria.

En un modelo cuya matriz de covariancias está correctamente especificada, el gráfico del variograma estimado de los residuos Cholesky versus el tiempo transcurrido entre las mediciones debería presentar un comportamiento aleatorio alrededor de una línea horizontal centrada en uno sin mostrar ningún patrón sistemático. Esta herramienta no sólo permite identificar si la estructura seleccionada no es adecuada, sino que también brinda información acerca del motivo de la falta de ajuste, permitiendo proponer modificaciones en el modelo.

Homocedasticidad del error condicional

La especificación del modelo para la covariancia del error condicional se compone de un modelo para la correlación y otro para la variancia. El modelo para la variancia se evalúa mediante los residuos condicionales. Dado que estos residuos pueden tener diferentes variancias, algunos autores sugieren la utilización de los residuos condicionales estudentizados, que se obtienen dividiendo el residuo ordinario por una estimación de su desvío estándar.

Nobre y Singer (2007) sugieren estandarizar los residuos tomando una estimación de la diagonal principal de la variancia del residuo condicional. Los elementos de dicha diagonal son funciones del *leverage* conjunto de los efectos fijos y aleatorios.

Si el modelo es adecuado, se espera que en un gráfico de los residuos condicionales estandarizados versus los valores ajustados los residuos se distribuyan aleatoriamente alrededor de cero (media del residuo condicional) con un rango de variación constante sin presentar ningún patrón sistemático.

Linealidad

El supuesto de linealidad en los modelos lineales mixtos puede evaluarse a través de los residuos marginales, de la misma forma que en los modelos lineales clásicos. Singer *et al.* (2016) proponen utilizar los residuos marginales estandarizados para probar este supuesto, de manera de eliminar el efecto de variancias heterogéneas presentes en los residuos marginales ordinarios.

Para evaluar este supuesto se realiza un gráfico de los residuos marginales estandarizados versus las covariables o los valores ajustados. Si el modelo está correctamente especificado, el gráfico no mostrará ningún patrón sistemático y se espera que los residuos varíen aleatoriamente alrededor de un valor constante igual a cero. Por otro lado, con el mismo objetivo, Fitzmaurice *et al.* (2004) proponen utilizar los residuos marginales Cholesky en lugar de los residuos marginales estandarizados.

Normalidad del error condicional

La normalidad del error condicional se comprueba utilizando un gráfico probabilístico normal de los residuos condicionales. Hilden-Minton (1995) argumenta que los residuos condicionales están confundidos por otras fuentes de error ajenas y esto debilita su habilidad para detectar alejamientos de la normalidad.

Hilden-Minton (1995) define los residuos con mínima confusión para comprobar este supuesto. De esta forma se elimina la influencia de la distribución de los efectos aleatorios sobre la distribución del error condicional.

Para comprobar el supuesto de normalidad sobre los errores condicionales se sugiere realizar un gráfico probabilístico normal con bandas de confianza simuladas de los residuos con confusión mínima estandarizados.

Normalidad de los efectos aleatorios

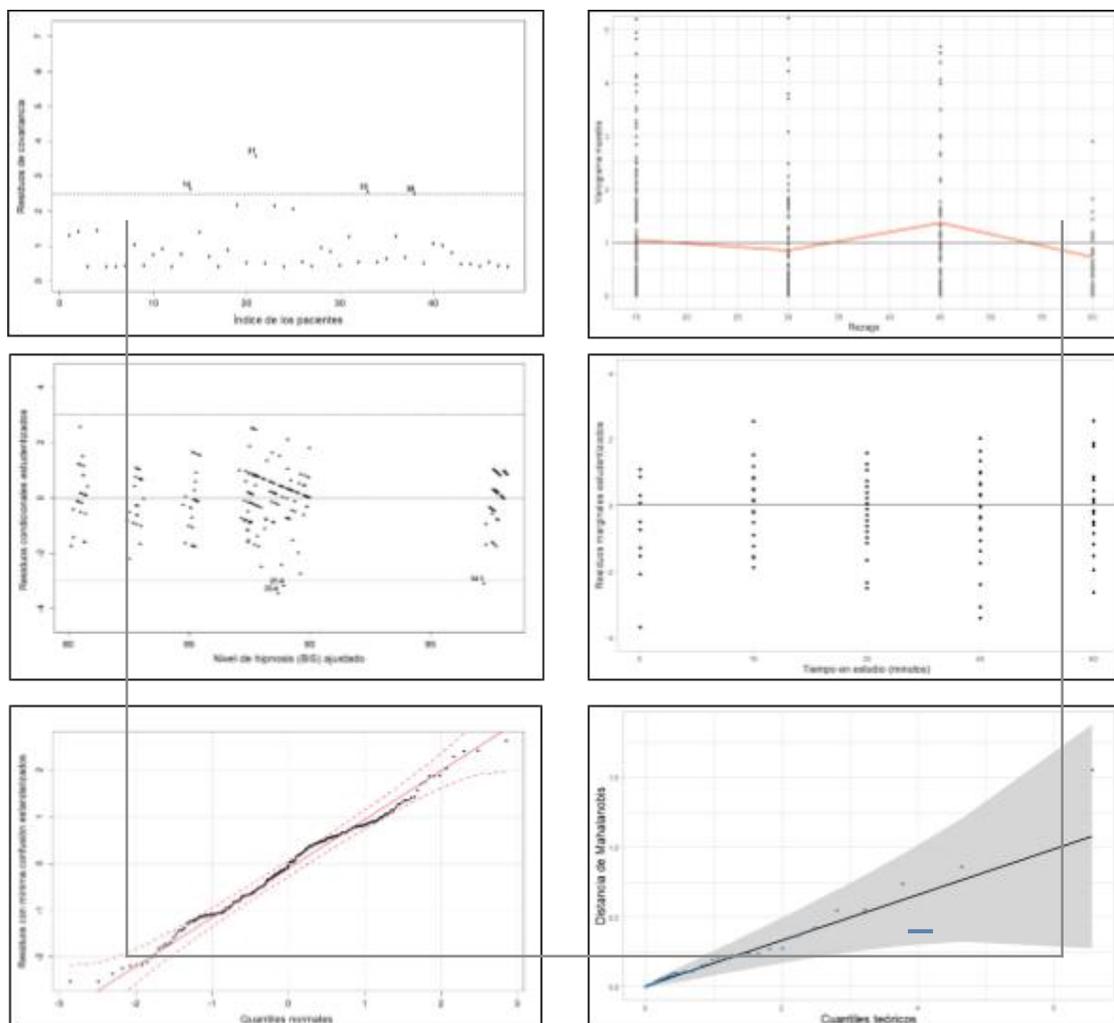
Para evaluar la normalidad de los efectos aleatorios se propusieron diferentes gráficos. Pinheiro y Bates (2000) utilizan como herramienta de diagnóstico el gráfico probabilístico normal de los efectos aleatorios predichos. Este gráfico presenta una desventaja: evalúa la normalidad de los efectos aleatorios por separado (normalidad marginal), siendo necesario evaluar la normalidad multivariada. Por otro lado, Singer *et al.* (2016) proponen evaluar el supuesto de normalidad utilizando la distancia de Mahalanobis entre los efectos aleatorios predichos y su esperanza. Cuando no hay confusión y los efectos aleatorios tienen distribución normal multivariada, la distancia de Mahalanobis se distribuye Chi-cuadrado. En consecuencia, verificar el supuesto de normalidad de los efectos aleatorios es equivalente a comprobar que las distancias se distribuyen Chi-cuadrado con q grados de libertad (siendo q la cantidad de efectos aleatorios en el modelo), y esto se puede realizar mediante el gráfico probabilístico chi-cuadrado de la distancia de Mahalanobis entre los efectos aleatorios predichos y su esperanza con bandas de confianza simuladas. Si en este gráfico se observa que hay puntos que caen por fuera de las bandas de confianza, se concluye que el supuesto de normalidad no se cumple.

RESULTADOS

Las herramientas presentadas se ilustran utilizando un conjunto de datos provenientes de un estudio clínico realizado en un efector de salud de la ciudad de Rosario. El objetivo es comparar el efecto de dos drogas sobre la profundidad de la hipnosis lograda durante la cirugía en pacientes que tienen una alta probabilidad de padecer el síndrome de apnea obstructiva del sueño (SAOS). El

estudio cuenta con 48 pacientes, a 25 de los cuales se les suministró dexmedetomidina, y a los 23 restantes se les suministró midazolam. La profundidad de la hipnosis se midió mediante el índice Biespectral (BIS). Este índice proporciona una medida del nivel de conciencia del paciente y toma valores entre 0 y 100, donde 0 representa la ausencia de función cerebral y 100 un nivel de alerta completa. A todos los pacientes se les midió el índice BIS en las mismas 5 ocasiones, al inicio del estudio (minuto 0) y luego cada 15 minutos hasta completar una hora de seguimiento. Este trabajo se enfoca en la utilización de técnicas gráficas basadas en los diferentes tipos de residuos, tanto en la construcción del modelo como en su diagnóstico. La Figura 1 muestra los gráficos de residuos correspondientes al modelo seleccionado.

Figura 1. Gráficos de residuos del modelo seleccionado para describir la evolución del índice BIS



CONSIDERACIONES FINALES

A partir del análisis se realizan las siguientes consideraciones:

- El uso de los gráficos de residuos marginales de covariancia y el variograma de los residuos de Cholesky en el proceso de construcción del modelo se complementan al momento de seleccionar la estructura de covariancias. El primer gráfico permitió detectar individuos para los cuales el modelo de covariancia no era adecuado y el segundo brindó información acerca de qué aspecto de la covariancia no estaba bien representado, sugiriendo que era necesario considerar una estructura de correlación serial para el error condicional.
- El gráfico de los residuos condicionales estudentizados versus los valores ajustados del modelo permitió comprobar la adecuación del modelo para la variancia, tanto en la etapa de construcción del modelo para la covariancia como en la etapa diagnóstico del modelo final.
- El gráfico de los residuos marginales estandarizados versus los valores ajustados permitió confirmar que la función propuesta para la parte media era correcta.
- El gráfico probabilístico normal para evaluar el supuesto de normalidad de los errores condicionales, mostró un leve alejamiento de la normalidad en las colas. Este gráfico presenta la limitación de utilizar residuos confundidos. Contrario a esto, el uso de residuos condicionales con mínima confusión sostuvo el supuesto de normalidad.
- El uso de ambos gráficos presentados para comprobar el supuesto de normalidad de los efectos aleatorios resulta indistinto en este estudio, debido a la presencia de un único efecto aleatorio. En ambos se demostró que el supuesto de normalidad del efecto aleatorio se verifica. Sin embargo, cuando se incluyen más efectos aleatorios en el modelo, los gráficos probabilísticos normales de los efectos aleatorios predichos no evalúan la normalidad multivariada como supone el modelo, sino la normalidad marginal. Se prefiere en ese caso utilizar el gráfico probabilístico Chi-cuadrado de la distancia de Mahalanobis entre el vector de efectos aleatorios predichos y su esperanza.

Bibliografía

- Diggle, P. J., Heagerty, P., Liang, K., & Zeger, S. L. (2001). *Analysis of Longitudinal Data*. Oxford Statistical Science Series, second edition.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied Longitudinal Analysis*. John Wiley & Sons.
- García, M. C., & Rapelli, C. (2012). Estudio de la validez de los supuestos en los modelos lineales mixtos mediante un análisis de residuos. *Revista FABICIB*, 16, 11-20.
- García, M. C. & Rapelli, C. (2018). Una comparación de métodos de diagnósticos para los modelos lineales mixtos. Instituto de Investigaciones Teóricas y Aplicadas, de la Escuela de Estadística.
- Hilden-Minton, J. A. (1995). *Multilevel Diagnostics for Mixed and Hierarchical Linear Models*. University of California.
- Lesaffre, E., & Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, 54, 570-582.
- Nobre, J. S., & Singer, J. M. (2007). Residual Analysis for Lineal Mixed Model. *Biometrical Journal*, 49(6), 863-875.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Model in S and S-plus*. Springer-Verlag.
- Schützenmeister, A., & Piepho, H. P. (2012). Residual analysis of linear mixed models using a simulation approach. *Comput. Stat. Data Anal.*, 56, 1405-1416.
- Singer, J. M., Rocha, F. M. M., & Nobre, J. S. (2016). Graphical Tools for Detecting Departures from Linear Mixed Model Assumptions and Some Remedial Measures. *International Statistical Review*, 1-35.
- Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.



CONSEJO PROFESIONAL
DE CIENCIAS ECONÓMICAS
DE LA PROVINCIA DE SANTA FE
CAMARA II



Colegio de Graduados
en Ciencias Económicas
de Rosario



Universidad
Nacional
de Rosario