

Curso de posgrado: “INTRODUCCIÓN A LOS MODELOS LINEALES GENERALIZADOS Y SU IMPLEMENTACIÓN CON LENGUAJE R”.

Temáticas: estadística aplicada, modelos lineales generalizados y lenguaje de programación R.

Equipo docente

Docente disertante y coordinador: Dr. Figueroa, Marcelo Isidro

Docente expositora: Dra. Mazzini, Flavia

Colaborador: Lic. Andrade, Lautaro Daniel

Fecha: 16/09/2024 al 20/09/2024

Fundamento

Los modelos lineales generalizados proporcionan un enfoque unificado para muchos de los procedimientos estadísticos asociativos más comúnmente utilizados tales como el análisis de la varianza, regresión lineal, gamma, beta, logística o de Poisson tanto uni como multivariadas, entre otras. Su campo de aplicación es diverso tal como agricultura, demografía, ecología, economía, educación, ingeniería, estudios ambientales y contaminación, geografía, geología, historia, medicina, ciencias políticas, psicología y sociología, por mencionar algunas disciplinas. Desde que el término “Modelo Lineal Generalizado - MLG” fue introducido por Nelder y Wedderburn en 1972, se han vuelto lentamente conocidos y, actualmente, ampliamente utilizados. No obstante, los libros de estadística introductoria y los cursos aún suelen centrarse en el modelo lineal general como si nada hubiera ocurrido en estadística en los últimos 70 años. Esto resulta en estudiantes con una visión muy limitada de la utilidad posible de la estadística en su campo de trabajo elegido.

La introducción de la idea de los modelos lineales generalizados a principios de la década de 1970 tuvo un gran impacto en la forma en que se lleva a cabo la estadística aplicada. Al principio, su uso estaba principalmente restringido a personal estadístico con amplia experticia dado que el material explicativo y el software disponibles estaban dirigidos a ellos. Fue necesario esperar aproximadamente veinte años para que los procedimientos de modelización lineal generalizada se implementen en software o paquetes informáticos como Genstat, Lisp-Stat, R, S-Plus o SAS con interfaces amigables para el usuario. Hoy en día los MLG forman parte de la “caja de herramientas” de aquellos investigadores que buscan realizar un análisis exhaustivo de sus datos.

Tanto en ciencias básicas, aplicadas, y otros ámbitos no científicos, se requiere la integración analítica de conjuntos de datos cada vez más complejos con el fin de dar una descripción más acabada de la realidad observada. A su vez, la generación de conjuntos de datos masivos, producto de la implementación de nuevas tecnologías (secuenciaciones masivas, análisis de imágenes satelitales, herramientas de monitoreo, análisis de redes sociales, etc...), posibilita este enfoque integrador y genera nichos de oportunidad para

futuros analistas. Dada la naturaleza versátil de los MLG es que podemos generar interpretaciones, predicciones y potenciales aplicaciones de los conocimientos derivados de estos datos.

Objetivos

- Comprender los fundamentos teóricos de los Modelos Lineales Generalizados (MLG).
- Implementar correctamente MLG en proyectos de investigación científica con software estadístico R.
- Interpretar y comunicar resultados de MLG.

Programa analítico

Unidad 1 – Introducción a las aplicaciones estadísticas con software estadístico R

Estadística descriptiva: medidas de tendencia central y dispersión. Representación gráfica de datos. Probabilidad. Distribuciones de probabilidad. Teorema del límite central. Inferencia estadística: muestreo, estimación puntual y por intervalos, pruebas de hipótesis. Análisis de asociación. Coeficiente de correlación de Pearson. Regresión lineal simple, interpretación de coeficientes y bondad de ajuste del modelo. Introducción al muestreo. Tipos de muestreo. Métodos de muestreo probabilístico. Métodos de muestreo no probabilístico. Tamaño de la muestra y técnicas de cálculo. Diseño muestral y estratificación. Muestreo sistemático. Muestreo por conglomerados. Muestreo aleatorio simple. Diseños analíticos: experimental, cuasi-experimental y observacional. Validez interna y externa en diseños experimentales. Amenazas a la validez interna y externa. Diseños longitudinales y transversales: ecológico, corte transversal, cohorte, casos y controles. Introducción a R y RStudio. Manejo de datos en R: lectura, escritura y manipulación de conjuntos de datos. Estructuras de datos en R: vectores, matrices, data frames y listas. Sintaxis en R.

Unidad 2 - Introducción a los modelos lineales generalizados

Naturaleza de variables. Variable de respuesta y variable explicativa. Definición, supuestos y elementos de un MLG: estimador lineal, función de enlace y distribución de probabilidades. Significado, diagnóstico y vulneraciones. Familias de distribuciones: gaussiana, gamma, beta, binomial, Poisson y binomial negativa. Gráficos y pruebas para el diagnóstico de MLG.

Unidad 3 – Modelado bajo distribuciones de probabilidad continuas y discretas

Distribución normal, gamma y beta. Aplicaciones prácticas, diagnóstico y reporte. Test de hipótesis y criterio de información. Variables respuesta de tipo conteo (distribución Poisson y Binomial Negativa). Sobre dispersión. Tasas en variables de conteo. Variables respuesta del tipo dicotómica o de proporciones (distribución binomial).

Unidad 4 – Inferencia sobre múltiples modelos

Introducción a la selección de modelos. Múltiples hipótesis y complejidad óptima del modelo: sobre ajuste y subajuste. Inferencia estadística utilizando prueba de hipótesis y teoría de la información. Similitudes y diferencias. Fortalezas y debilidades.

Unidad 5 - Reporte de resultados de MLG e introducción a sus extensiones.

Aspectos a tener en cuenta a la hora de reportar resultados de un MLG: posibles destinatarios (revistas científicas, tesis e informes estadísticos), información a reportar (flujos de trabajo, modelos planteados, resultados y predicciones) y modo de reporte (fórmulas, tablas y gráficos). Extensiones de los MLG: modelos lineales generalizados con efectos mixtos (GLMM), modelos no lineales (NLM), modelos aditivos generalizados (GAM), modelos aditivos generalizados para ubicación, escala y forma (GAMLSS) y modelos lineales generalizados de variables latentes (GLLVM).

Cursado y metodología de enseñanza

La modalidad de cursado será intensiva, 8 horas diarias durante 5 días. La carga horaria estimada de tutorías más la correspondiente al proyecto de análisis de datos alcanzará las 20 horas. Por lo tanto, la carga horaria total del curso será de 60 horas (40 presenciales y 20 virtuales).

La modalidad de enseñanza será bajo la figura de “Curso de posgrado teórico-práctico” establecida por resolución C.S. 0302/16. Las jornadas se dividirán en dos partes:

- **Matutina:** 5 (cinco) clases teóricas expositivas con ejemplos prácticos ejecutados por el docente.
- **Vespertina:** 5 (cinco) clases prácticas con ejemplos brindados por el docente (ejecutados por el estudiante), propios y puesta en común de resultados obtenidos.

Bibliografía

- Burnham, K. P., & D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Segunda edición. Springer, New York.
- Crawley, M. J. 2013. The R book. Segunda edición. John Wiley and Sons, Ltd., Chichester.
- Garamszegi, L. Z. 2011. Information-theoretic approaches to statistical analysis in behavioural ecology: an introduction. Behavioral Ecology and Sociobiology 65: 1-11.
- Hilbe, J. M. 2011. Negative binomial regression. Segunda edición. Cambridge University Press, New York.
- Logan, M. 2010. Biostatistical design and analysis using R: a practical guide. John Wiley and Sons, Ltd., Chichester.
- McCullagh, P., & J. A. Nelder. 1989. Generalized linear models. Segunda edición. Chapman & Hall, New York.
- R Development Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.